1, TITLE: A New Representation of Successor Features for Transfer across Dissimilar Environments
http://proceedings.mlr.press/v139/abdolshah21a.html
AUTHORS: Majid Abdolshah, Hung Le, Thommen Karimpanal George, Sunil Gupta, Santu Rana, Svetha Venkatesh
HIGHLIGHT: To address this problem, we propose an approach based on successor features in which we model successor feature functions with Gaussian Processes permitting the source successor features to be treated as noisy measurements of the target successor feature function.

2, TITLE: Massively Parallel and Asynchronous Tsetlin Machine Architecture Supporting Almost Constant-Time Scaling
http://proceedings.mlr.press/v139/abeyrathna21a.html
AUTHORS: Kuruge Darshana Abeyrathna, Bimal Bhattarai, Morten Goodwin, Saeed Rahimi Gorji, Ole-Christoffer Granmo, Lei Jiao, Rupsa Saha, Rohan K Yadav
HIGHLIGHT: In this paper, we propose a novel scheme for desynchronizing the evaluation of clauses, eliminating the voting bottleneck.

3, TITLE: Debiasing Model Updates for Improving Personalized Federated Training
http://proceedings.mlr.press/v139/acar21a.html
AUTHORS: Durmus Alp Emre Acar, Yue Zhao, Ruizhao Zhu, Ramon Matas, Matthew Mattina, Paul Whatmough, Venkatesh Saligrama
HIGHLIGHT: We propose a novel method for federated learning that is customized specifically to the objective of a given edge device.

4, TITLE: Memory Efficient Online Meta Learning
http://proceedings.mlr.press/v139/acar21b.html
AUTHORS: Durmus Alp Emre Acar, Ruizhao Zhu, Venkatesh Saligrama
HIGHLIGHT: We propose a novel algorithm for online meta learning where task instances are sequentially revealed with limited supervision and a learner is expected to meta learn them in each round, so as to allow the learner to customize a task-specific model rapidly with little task-level supervision.

5, TITLE: Robust Testing and Estimation under Manipulation Attacks
http://proceedings.mlr.press/v139/acharya21a.html
AUTHORS: Jayadev Acharya, Ziteng Sun, Huanyu Zhang
HIGHLIGHT: We study robust testing and estimation of discrete distributions in the strong contamination model.

6, TITLE: GP-Tree: A Gaussian Process Classifier for Few-Shot Incremental Learning
http://proceedings.mlr.press/v139/achituve21a.html
AUTHORS: Idan Achituve, Aviv Navon, Yochai Yemini, Gal Chechik, Ethan Fetaya
HIGHLIGHT: Here, we propose GP-Tree, a novel method for multi-class classification with Gaussian processes and DKL.

7, TITLE: f-Domain Adversarial Learning: Theory and Algorithms
http://proceedings.mlr.press/v139/acuna21a.html
AUTHORS: David Acuna, Guojun Zhang, Marc T. Law, Sanja Fidler
HIGHLIGHT: In this paper, we introduce a novel and general domain-adversarial framework.

8, TITLE: Towards Rigorous Interpretations: a Formalisation of Feature Attribution
http://proceedings.mlr.press/v139/afchar21a.html
AUTHORS: Darius Afchar, Vincent Guigue, Romain Hennequin
HIGHLIGHT: In this paper we propose to formalise feature selection/attribution based on the concept of relaxed functional dependence.

9, TITLE: Acceleration via Fractal Learning Rate Schedules
http://proceedings.mlr.press/v139/agarwal21a.html
AUTHORS: Naman Agarwal, Surbhi Goel, Cyril Zhang
HIGHLIGHT: We provide some experiments to challenge conventional beliefs about stable learning rates in deep learning: the fractal schedule enables training to converge with locally unstable updates which make negative progress on the objective.

10, TITLE: A Regret Minimization Approach to Iterative Learning Control
http://proceedings.mlr.press/v139/agarwal21b.html
AUTHORS: Naman Agarwal, Elad Hazan, Anirudha Majumdar, Karan Singh
HIGHLIGHT: In this setting, we propose a new performance metric, planning regret, which replaces the standard stochastic uncertainty assumptions with worst case regret.

11, TITLE: Towards the Unification and Robustness of Perturbation and Gradient Based Explanations
http://proceedings.mlr.press/v139/agarwal21c.html
AUTHORS: Sushant Agarwal, Shahin Jabbari, Chirag Agarwal, Sohini Upadhyay, Steven Wu, Himabindu Lakkaraju
HIGHLIGHT: In this work, we analyze two popular post hoc interpretation techniques: SmoothGrad which is a gradient based method, and a variant of LIME which is a perturbation based method.

12, TITLE: Label Inference Attacks from Log-loss Scores
http://proceedings.mlr.press/v139/aggarwal21a.html
AUTHORS: Abhinav Aggarwal, Shiva Kasiviswanathan, Zekun Xu, Oluwaseyi Feyisetan, Nathanael Teissier
HIGHLIGHT: In this paper, we investigate the problem of inferring the labels of a dataset from single (or multiple) log-loss score(s), without any other access to the dataset.

13, TITLE: Deep kernel processes
http://proceedings.mlr.press/v139/aitchison21a.html
AUTHORS: Laurence Aitchison, Adam Yang, Sebastian W. Ober
HIGHLIGHT: We define deep kernel processes in which positive definite Gram matrices are progressively transformed by nonlinear kernel functions and by sampling from (inverse) Wishart distributions.

14, TITLE: How Does Loss Function Affect Generalization Performance of Deep Learning? Application to Human Age Estimation
http://proceedings.mlr.press/v139/akbari21a.html
AUTHORS: Ali Akbari, Muhammad Awais, Manijeh Bashar, Josef Kittler
HIGHLIGHT: In summary, our main statement in this paper is: choose a stable loss function, generalize better.

15, TITLE: On Learnability via Gradient Method for Two-Layer ReLU Neural Networks in Teacher-Student Setting
http://proceedings.mlr.press/v139/akiyama21a.html
AUTHORS: Shunta Akiyama, Taiji Suzuki
HIGHLIGHT: In this paper, we explore theoretical analysis on training two-layer ReLU neural networks in a teacher-student regression model, in which a student network learns an unknown teacher network through its outputs.

16, TITLE: Slot Machines: Discovering Winning Combinations of Random Weights in Neural Networks
http://proceedings.mlr.press/v139/aladago21a.html
AUTHORS: Maxwell M Aladago, Lorenzo Torresani
HIGHLIGHT: In contrast to traditional weight optimization in a continuous space, we demonstrate the existence of effective random networks whose weights are never updated.

17, TITLE: A large-scale benchmark for few-shot program induction and synthesis
http://proceedings.mlr.press/v139/alet21a.html
AUTHORS: Ferran Alet, Javier Lopez-Contreras, James Koppel, Maxwell Nye, Armando Solar-Lezama, Tomas Lozano-Perez, Leslie Kaelbling, Joshua Tenenbaum
HIGHLIGHT: In this work, we propose a new way of leveraging unit tests and natural inputs for small programs as meaningful input-output examples for each sub-program of the overall program.

18, TITLE: Robust Pure Exploration in Linear Bandits with Limited Budget
http://proceedings.mlr.press/v139/alieva21a.html
AUTHORS: Ayya Alieva, Ashok Cutkosky, Abhimanyu Das
HIGHLIGHT: We consider the pure exploration problem in the fixed-budget linear bandit setting.

19, TITLE: Communication-Efficient Distributed Optimization with Quantized Preconditioners
http://proceedings.mlr.press/v139/alimisis21a.html
AUTHORS: Foivos Alimisis, Peter Davies, Dan Alistarh
HIGHLIGHT: We investigate fast and communication-efficient algorithms for the classic problem of minimizing a sum of strongly convex and smooth functions that are distributed among n different nodes, which can communicate using a limited number of bits.

20, TITLE: Non-Exponentially Weighted Aggregation: Regret Bounds for Unbounded Loss Functions
http://proceedings.mlr.press/v139/alquier21a.html
AUTHORS: Pierre Alquier
HIGHLIGHT: In this paper, we study a generalized aggregation strategy, where the weights no longer depend exponentially on the losses.

21, TITLE: Dataset Dynamics via Gradient Flows in Probability Space
http://proceedings.mlr.press/v139/alvarez-melis21a.html
AUTHORS: David Alvarez-Melis, Nicol? Fusi
HIGHLIGHT: In this work, we propose a novel framework for dataset transformation, which we cast as optimization over data-generating joint probability distributions.

22, TITLE: Submodular Maximization subject to a Knapsack Constraint: Combinatorial Algorithms with Near-optimal Adaptive Complexity
http://proceedings.mlr.press/v139/amanatidis21a.html
AUTHORS: Georgios Amanatidis, Federico Fusco, Philip Lazos, Stefano Leonardi, Alberto Marchetti-Spaccamela, Rebecca Reiffenh?user
HIGHLIGHT: In this work we obtain the first \emph{constant factor} approximation algorithm for non-monotone submodular maximization subject to a knapsack constraint with \emph{near-optimal} $O(\log n)$ adaptive complexity.

23, TITLE: Safe Reinforcement Learning with Linear Function Approximation
http://proceedings.mlr.press/v139/amani21a.html
AUTHORS: Sanae Amani, Christos Thrampoulidis, Lin Yang
HIGHLIGHT: In this paper, we address both problems by first modeling safety as an unknown linear cost function of states and actions, which must always fall below a certain threshold.

24, TITLE: Automatic variational inference with cascading flows
http://proceedings.mlr.press/v139/ambrogioni21a.html
AUTHORS: Luca Ambrogioni, Gianluigi Silvestri, Marcel Van Gerven
HIGHLIGHT: Here, we combine the flexibility of normalizing flows and the prior-embedding property of ASVI in a new family of variational programs, which we named cascading flows.

25, TITLE: Sparse Bayesian Learning via Stepwise Regression
http://proceedings.mlr.press/v139/ament21a.html
AUTHORS: Sebastian E. Ament, Carla P. Gomes
HIGHLIGHT: Herein, we propose a coordinate ascent algorithm for SBL termed Relevance Matching Pursuit (RMP) and show that, as its noise variance parameter goes to zero, RMP exhibits a surprising connection to Stepwise Regression.

26, TITLE: Locally Persistent Exploration in Continuous Control Tasks with Sparse Rewards
http://proceedings.mlr.press/v139/amin21a.html
AUTHORS: Susan Amin, Maziar Gomrokchi, Hossein Aboutalebi, Harsh Satija, Doina Precup
HIGHLIGHT: We propose a new exploration method, based on two intuitions: (1) the choice of the next exploratory action should depend not only on the (Markovian) state of the environment, but also on the agent's trajectory so far, and (2) the agent should utilize a measure of spread in the state space to avoid getting stuck in a small region.

27, TITLE: Preferential Temporal Difference Learning
http://proceedings.mlr.press/v139/anand21a.html
AUTHORS: Nishanth V. Anand, Doina Precup
HIGHLIGHT: We propose an approach to re-weighting states used in TD updates, both when they are the input and when they provide the target for the update.

28, TITLE: Unitary Branching Programs: Learnability and Lower Bounds
http://proceedings.mlr.press/v139/andino21a.html
AUTHORS: Fidel Ernesto Diaz Andino, Maria Kokkou, Mateus De Oliveira Oliveira, Farhad Vadiee
HIGHLIGHT: In this work, we study a generalized version of bounded width branching programs where instructions are defined by unitary matrices of bounded dimension.

29, TITLE: The Logical Options Framework
http://proceedings.mlr.press/v139/araki21a.html
AUTHORS: Brandon Araki, Xiao Li, Kiran Vodrahalli, Jonathan Decastro, Micah Fry, Daniela Rus
HIGHLIGHT: We introduce a hierarchical reinforcement learning framework called the Logical Options Framework (LOF) that learns policies that are satisfying, optimal, and composable.

30, TITLE: Annealed Flow Transport Monte Carlo
http://proceedings.mlr.press/v139/arbel21a.html

AUTHORS: Michael Arbel, Alex Matthews, Arnaud Doucet
HIGHLIGHT: We propose here a novel Monte Carlo algorithm, Annealed Flow Transport (AFT), that builds upon AIS and SMC and combines them with normalizing flows (NFs) for improved performance.

31, TITLE: Permutation Weighting
http://proceedings.mlr.press/v139/arbour21a.html
AUTHORS: David Arbour, Drew Dimmery, Arjun Sondhi
HIGHLIGHT: In this work we introduce permutation weighting, a method for estimating balancing weights using a standard binary classifier (regardless of cardinality of treatment).

32, TITLE: Analyzing the tree-layer structure of Deep Forests
http://proceedings.mlr.press/v139/arnould21a.html
AUTHORS: Ludovic Arnould, Claire Boyer, Erwan Scornet
HIGHLIGHT: In this paper, our aim is not to benchmark DF performances but to investigate instead their underlying mechanisms.

33, TITLE: Dropout: Explicit Forms and Capacity Control
http://proceedings.mlr.press/v139/arora21a.html
AUTHORS: Raman Arora, Peter Bartlett, Poorya Mianjy, Nathan Srebro
HIGHLIGHT: We investigate the capacity control provided by dropout in various machine learning problems.

34, TITLE: Tighter Bounds on the Log Marginal Likelihood of Gaussian Process Regression Using Conjugate Gradients
http://proceedings.mlr.press/v139/artemev21a.html
AUTHORS: Artem Artemev, David R Burt, Mark Van Der Wilk
HIGHLIGHT: We propose a lower bound on the log marginal likelihood of Gaussian process regression models that can be computed without matrix factorisation of the full kernel matrix.

35, TITLE: Deciding What to Learn: A Rate-Distortion Approach
http://proceedings.mlr.press/v139/arumugam21a.html
AUTHORS: Dilip Arumugam, Benjamin Van Roy
HIGHLIGHT: In this work, leveraging rate-distortion theory, we automate this process such that the designer need only express their preferences via a single hyperparameter and the agent is endowed with the ability to compute its own learning targets that best achieve the desired trade-off.

36, TITLE: Private Adaptive Gradient Methods for Convex Optimization
http://proceedings.mlr.press/v139/asi21a.html
AUTHORS: Hilal Asi, John Duchi, Alireza Fallah, Omid Javidbakht, Kunal Talwar
HIGHLIGHT: We study adaptive methods for differentially private convex optimization, proposing and analyzing differentially private variants of a Stochastic Gradient Descent (SGD) algorithm with adaptive stepsizes, as well as the AdaGrad algorithm.

37, TITLE: Private Stochastic Convex Optimization: Optimal Rates in L1 Geometry
http://proceedings.mlr.press/v139/asi21b.html
AUTHORS: Hilal Asi, Vitaly Feldman, Tomer Koren, Kunal Talwar
HIGHLIGHT: We show that, up to logarithmic factors the optimal excess population loss of any (ϵ,δ)-differentially private optimizer is $\sqrt{\log(d)/n} + \sqrt{d}/\epsilon n.$ The upper bound is based on a new algorithm that combines the iterative localization approach of Feldman et al. (2020) with a new analysis of private regularized mirror descent.

38, TITLE: Combinatorial Blocking Bandits with Stochastic Delays
http://proceedings.mlr.press/v139/atsidakou21a.html
AUTHORS: Alexia Atsidakou, Orestis Papadigenopoulos, Soumya Basu, Constantine Caramanis, Sanjay Shakkottai
HIGHLIGHT: In this work, we extend the above model in two directions: (i) We consider the general combinatorial setting where more than one arms can be played at each round, subject to feasibility constraints. (ii) We allow the blocking time of each arm to be stochastic.

39, TITLE: Dichotomous Optimistic Search to Quantify Human Perception
http://proceedings.mlr.press/v139/audiffren21a.html
AUTHORS: Julien Audiffren
HIGHLIGHT: In this paper we address a variant of the continuous multi-armed bandits problem, called the threshold estimation problem, which is at the heart of many psychometric experiments.

40, TITLE: Federated Learning under Arbitrary Communication Patterns
http://proceedings.mlr.press/v139/avdiukhin21a.html
AUTHORS: Dmitrii Avdiukhin, Shiva Kasiviswanathan
HIGHLIGHT: In this paper, we investigate the performance of an asynchronous version of local SGD wherein the clients can communicate with the server at arbitrary time intervals.

41, TITLE: Asynchronous Distributed Learning : Adapting to Gradient Delays without Prior Knowledge
http://proceedings.mlr.press/v139/aviv21a.html
AUTHORS: Rotem Zamir Aviv, Ido Hakimi, Assaf Schuster, Kfir Yehuda Levy
HIGHLIGHT: We propose a robust training method for the constrained setting and derive non asymptotic convergence guarantees that do not depend on prior knowledge of update delays, objective smoothness, and gradient variance.

42, TITLE: Decomposable Submodular Function Minimization via Maximum Flow
http://proceedings.mlr.press/v139/axiotis21a.html
AUTHORS: Kyriakos Axiotis, Adam Karczmarz, Anish Mukherjee, Piotr Sankowski, Adrian Vladu
HIGHLIGHT: We solve this minimization problem by lifting the solutions of a parametric cut problem, which we obtain via a new efficient combinatorial reduction to maximum flow.

43, TITLE: Differentially Private Query Release Through Adaptive Projection
http://proceedings.mlr.press/v139/aydore21a.html
AUTHORS: Sergul Aydore, William Brown, Michael Kearns, Krishnaram Kenthapadi, Luca Melis, Aaron Roth, Ankit A Siva
HIGHLIGHT: We propose, implement, and evaluate a new algo-rithm for releasing answers to very large numbersof statistical queries likek-way marginals, sub-ject to differential privacy.

44, TITLE: On the Implicit Bias of Initialization Shape: Beyond Infinitesimal Mirror Descent
http://proceedings.mlr.press/v139/azulay21a.html
AUTHORS: Shahar Azulay, Edward Moroshko, Mor Shpigel Nacson, Blake E Woodworth, Nathan Srebro, Amir Globerson, Daniel Soudry
HIGHLIGHT: We develop a novel technique for deriving the inductive bias of gradient-flow and use it to obtain closed-form implicit regularizers for multiple cases of interest.

45, TITLE: On-Off Center-Surround Receptive Fields for Accurate and Robust Image Classification
http://proceedings.mlr.press/v139/babaiee21a.html
AUTHORS: Zahra Babaiee, Ramin Hasani, Mathias Lechner, Daniela Rus, Radu Grosu
HIGHLIGHT: To this end, our paper extends the receptive field of convolutional neural networks with two residual components, ubiquitous in the visual processing system of vertebrates: On-center and off-center pathways, with an excitatory center and inhibitory surround; OOCS for short.

46, TITLE: Uniform Convergence, Adversarial Spheres and a Simple Remedy
http://proceedings.mlr.press/v139/bachmann21a.html
AUTHORS: Gregor Bachmann, Seyed-Mohsen Moosavi-Dezfooli, Thomas Hofmann
HIGHLIGHT: We provide an extensive theoretical investigation of the previously studied data setting through the lens of infinitely-wide models.

47, TITLE: Faster Kernel Matrix Algebra via Density Estimation
http://proceedings.mlr.press/v139/backurs21a.html
AUTHORS: Arturs Backurs, Piotr Indyk, Cameron Musco, Tal Wagner
HIGHLIGHT: We study fast algorithms for computing basic properties of an n x n positive semidefinite kernel matrix K corresponding to n points x_1,...,x_n in R^d.

48, TITLE: Robust Reinforcement Learning using Least Squares Policy Iteration with Provable Performance Guarantees
http://proceedings.mlr.press/v139/badrinath21a.html
AUTHORS: Kishan Panaganti Badrinath, Dileep Kalathil
HIGHLIGHT: This paper addresses the problem of model-free reinforcement learning for Robust Markov Decision Process (RMDP) with large state spaces.

49, TITLE: Skill Discovery for Exploration and Planning using Deep Skill Graphs
http://proceedings.mlr.press/v139/bagaria21a.html
AUTHORS: Akhil Bagaria, Jason K Senthil, George Konidaris

HIGHLIGHT: We introduce a new skill-discovery algorithm that builds a discrete graph representation of large continuous MDPs, where nodes correspond to skill subgoals and the edges to skill policies.

50, TITLE: Locally Adaptive Label Smoothing Improves Predictive Churn
http://proceedings.mlr.press/v139/bahri21a.html
AUTHORS: Dara Bahri, Heinrich Jiang
HIGHLIGHT: In this paper, we present several baselines for reducing churn and show that training on soft labels obtained by adaptively smoothing each example's label based on the example's neighboring labels often outperforms the baselines on churn while improving accuracy on a variety of benchmark classification tasks and model architectures.

51, TITLE: How Important is the Train-Validation Split in Meta-Learning?
http://proceedings.mlr.press/v139/bai21a.html
AUTHORS: Yu Bai, Minshuo Chen, Pan Zhou, Tuo Zhao, Jason Lee, Sham Kakade, Huan Wang, Caiming Xiong
HIGHLIGHT: We provide a detailed theoretical study on whether and when the train-validation split is helpful in the linear centroid meta-learning problem.

52, TITLE: Stabilizing Equilibrium Models by Jacobian Regularization
http://proceedings.mlr.press/v139/bai21b.html
AUTHORS: Shaojie Bai, Vladlen Koltun, Zico Kolter
HIGHLIGHT: In this paper, we propose a regularization scheme for DEQ models that explicitly regularizes the Jacobian of the fixed-point update equations to stabilize the learning of equilibrium models.

53, TITLE: Don't Just Blame Over-parametrization for Over-confidence: Theoretical Analysis of Calibration in Binary Classification
http://proceedings.mlr.press/v139/bai21c.html
AUTHORS: Yu Bai, Song Mei, Huan Wang, Caiming Xiong
HIGHLIGHT: In this paper, we show theoretically that over-parametrization is not the only reason for over-confidence.

54, TITLE: Principled Exploration via Optimistic Bootstrapping and Backward Induction
http://proceedings.mlr.press/v139/bai21d.html
AUTHORS: Chenjia Bai, Lingxiao Wang, Lei Han, Jianye Hao, Animesh Garg, Peng Liu, Zhaoran Wang
HIGHLIGHT: In this paper, we propose a principled exploration method for DRL through Optimistic Bootstrapping and Backward Induction (OB2I).

55, TITLE: GLSearch: Maximum Common Subgraph Detection via Learning to Search
http://proceedings.mlr.press/v139/bai21e.html
AUTHORS: Yunsheng Bai, Derek Xu, Yizhou Sun, Wei Wang
HIGHLIGHT: We propose GLSearch, a Graph Neural Network (GNN) based learning to search model.

56, TITLE: Breaking the Limits of Message Passing Graph Neural Networks
http://proceedings.mlr.press/v139/balcilar21a.html
AUTHORS: Muhammet Balcilar, Pierre Heroux, Benoit Gauzere, Pascal Vasseur, Sebastien Adam, Paul Honeine
HIGHLIGHT: In this paper, we show that if the graph convolution supports are designed in spectral-domain by a non-linear custom function of eigenvalues and masked with an arbitrary large receptive field, the MPNN is theoretically more powerful than the 1-WL test and experimentally as powerful as a 3-WL existing models, while remaining spatially localized.

57, TITLE: Instance Specific Approximations for Submodular Maximization
http://proceedings.mlr.press/v139/balkanski21a.html
AUTHORS: Eric Balkanski, Sharon Qian, Yaron Singer
HIGHLIGHT: We develop an algorithm that gives an instance-specific approximation for any solution of an instance of monotone submodular maximization under a cardinality constraint.

58, TITLE: Augmented World Models Facilitate Zero-Shot Dynamics Generalization From a Single Offline Environment
http://proceedings.mlr.press/v139/ball21a.html
AUTHORS: Philip J Ball, Cong Lu, Jack Parker-Holder, Stephen Roberts
HIGHLIGHT: However, little attention has been paid to potentially changing dynamics when transferring a policy to the online setting, where performance can be up to 90% reduced for existing methods. In this paper we address this problem with Augmented World Models (AugWM).

59, TITLE: Regularized Online Allocation Problems: Fairness and Beyond

http://proceedings.mlr.press/v139/balseiro21a.html
AUTHORS: Santiago Balseiro, Haihao Lu, Vahab Mirrokni
HIGHLIGHT: In this paper, we introduce the regularized online allocation problem, a variant that includes a non-linear regularizer acting on the total resource consumption.

60, TITLE: Predict then Interpolate: A Simple Algorithm to Learn Stable Classifiers
http://proceedings.mlr.press/v139/bao21a.html
AUTHORS: Yujia Bao, Shiyu Chang, Dr.Regina Barzilay
HIGHLIGHT: We propose Predict then Interpolate (PI), a simple algorithm for learning correlations that are stable across environments.

61, TITLE: Variational (Gradient) Estimate of the Score Function in Energy-based Latent Variable Models
http://proceedings.mlr.press/v139/bao21b.html
AUTHORS: Fan Bao, Kun Xu, Chongxuan Li, Lanqing Hong, Jun Zhu, Bo Zhang
HIGHLIGHT: This paper presents new estimates of the score function and its gradient with respect to the model parameters in a general energy-based latent variable model (EBLVM).

62, TITLE: Compositional Video Synthesis with Action Graphs
http://proceedings.mlr.press/v139/bar21a.html
AUTHORS: Amir Bar, Roei Herzig, Xiaolong Wang, Anna Rohrbach, Gal Chechik, Trevor Darrell, Amir Globerson
HIGHLIGHT: To address this challenge, we propose to represent the actions in a graph structure called Action Graph and present the new "Action Graph To Video" synthesis task.

63, TITLE: Approximating a Distribution Using Weight Queries
http://proceedings.mlr.press/v139/barak21a.html
AUTHORS: Nadav Barak, Sivan Sabato
HIGHLIGHT: We propose an interactive algorithm that iteratively selects data set examples and performs corresponding weight queries.

64, TITLE: Graph Convolution for Semi-Supervised Classification: Improved Linear Separability and Out-of-Distribution Generalization
http://proceedings.mlr.press/v139/baranwal21a.html
AUTHORS: Aseem Baranwal, Kimon Fountoulakis, Aukosh Jagannath
HIGHLIGHT: To understand the merits of this approach, we study the classification of a mixture of Gaussians, where the data corresponds to the node attributes of a stochastic block model.

65, TITLE: Training Quantized Neural Networks to Global Optimality via Semidefinite Programming
http://proceedings.mlr.press/v139/bartan21a.html
AUTHORS: Burak Bartan, Mert Pilanci
HIGHLIGHT: In this work, we introduce a convex optimization strategy to train quantized NNs with polynomial activations.

66, TITLE: Beyond $log^2(T)$ regret for decentralized bandits in matching markets
http://proceedings.mlr.press/v139/basu21a.html
AUTHORS: Soumya Basu, Karthik Abinav Sankararaman, Abishek Sankararaman
HIGHLIGHT: We propose a phase based algorithm, where in each phase, besides deleting the globally communicated dominated arms the agents locally delete arms with which they collide often.

67, TITLE: Optimal Thompson Sampling strategies for support-aware CVaR bandits
http://proceedings.mlr.press/v139/baudry21a.html
AUTHORS: Dorian Baudry, Romain Gautron, Emilie Kaufmann, Odalric Maillard
HIGHLIGHT: In this paper we study a multi-arm bandit problem in which the quality of each arm is measured by the Conditional Value at Risk (CVaR) at some level alpha of the reward distribution.

68, TITLE: On Limited-Memory Subsampling Strategies for Bandits
http://proceedings.mlr.press/v139/baudry21b.html
AUTHORS: Dorian Baudry, Yoan Russac, Olivier Capp?
HIGHLIGHT: Our first contribution is to show that a simple deterministic subsampling rule, proposed in the recent work of \citet{baudry2020sub} under the name of “last-block subsampling”, is asymptotically optimal in one-parameter exponential families.

69, TITLE: Generalized Doubly Reparameterized Gradient Estimators

http://proceedings.mlr.press/v139/bauer21a.html
AUTHORS: Matthias Bauer, Andriy Mnih
HIGHLIGHT: Here, we develop two generalizations of the DReGs estimator and show that they can be used to train conditional and hierarchical VAEs on image modelling tasks more effectively.

70, TITLE: Directional Graph Networks
http://proceedings.mlr.press/v139/beani21a.html
AUTHORS: Dominique Beani, Saro Passaro, Vincent L?tourneau, Will Hamilton, Gabriele Corso, Pietro Li?
HIGHLIGHT: To overcome this limitation, we propose the first globally consistent anisotropic kernels for GNNs, allowing for graph convolutions that are defined according to topologicaly-derived directional flows.

71, TITLE: Policy Analysis using Synthetic Controls in Continuous-Time
http://proceedings.mlr.press/v139/bellot21a.html
AUTHORS: Alexis Bellot, Mihaela Van Der Schaar
HIGHLIGHT: We propose a continuous-time alternative that models the latent counterfactual path explicitly using the formalism of controlled differential equations.

72, TITLE: Loss Surface Simplexes for Mode Connecting Volumes and Fast Ensembling
http://proceedings.mlr.press/v139/benton21a.html
AUTHORS: Gregory Benton, Wesley Maddox, Sanae Lotfi, Andrew Gordon Gordon Wilson
HIGHLIGHT: In this paper, we in fact demonstrate the existence of mode-connecting simplicial complexes that form multi-dimensional manifolds of low loss, connecting many independently trained models.

73, TITLE: TFix: Learning to Fix Coding Errors with a Text-to-Text Transformer
http://proceedings.mlr.press/v139/berabi21a.html
AUTHORS: Berkay Berabi, Jingxuan He, Veselin Raychev, Martin Vechev
HIGHLIGHT: In this paper, we address this challenge and present a new learning-based system, called TFix.

74, TITLE: Learning Queueing Policies for Organ Transplantation Allocation using Interpretable Counterfactual Survival Analysis
http://proceedings.mlr.press/v139/berrevoets21a.html
AUTHORS: Jeroen Berrevoets, Ahmed Alaa, Zhaozhi Qian, James Jordon, Alexander E.S. Gimson, Mihaela Van Der Schaar
HIGHLIGHT: In this paper, we develop a data-driven model for (real-time) organ allocation using observational data for transplant outcomes.
Furthermore, we introduce a novel organ-allocation simulator to accurately test new policies.

75, TITLE: Learning from Biased Data: A Semi-Parametric Approach
http://proceedings.mlr.press/v139/bertail21a.html
AUTHORS: Patrice Bertail, Stephan Cl?men?on, Yannick Guyonvarch, Nathan Noiry
HIGHLIGHT: We consider risk minimization problems where the (source) distribution P_S of the training observations $Z_1, \ldots, Z_n$ differs from the (target) distribution P_T involved in the risk that one seeks to minimize.

76, TITLE: Is Space-Time Attention All You Need for Video Understanding?
http://proceedings.mlr.press/v139/bertasius21a.html
AUTHORS: Gedas Bertasius, Heng Wang, Lorenzo Torresani
HIGHLIGHT: We present a convolution-free approach to video classification built exclusively on self-attention over space and time.

77, TITLE: Confidence Scores Make Instance-dependent Label-noise Learning Possible
http://proceedings.mlr.press/v139/berthon21a.html
AUTHORS: Antonin Berthon, Bo Han, Gang Niu, Tongliang Liu, Masashi Sugiyama
HIGHLIGHT: To alleviate this issue, we introduce confidence-scored instance-dependent noise (CSIDN), where each instance-label pair is equipped with a confidence score.

78, TITLE: Size-Invariant Graph Representations for Graph Classification Extrapolations
http://proceedings.mlr.press/v139/bevilacqua21a.html
AUTHORS: Beatrice Bevilacqua, Yangze Zhou, Bruno Ribeiro
HIGHLIGHT: In this work we consider an underexplored area of an otherwise rapidly developing field of graph representation learning: The task of out-of-distribution (OOD) graph classification, where train and test data have different distributions, with test data unavailable during training.

79, TITLE: Principal Bit Analysis: Autoencoding with Schur-Concave Loss
http://proceedings.mlr.press/v139/bhadane21a.html
AUTHORS: Sourbh Bhadane, Aaron B Wagner, Jayadev Acharya
HIGHLIGHT: We consider a linear autoencoder in which the latent variables are quantized, or corrupted by noise, and the constraint is Schur-concave in the set of latent variances.

80, TITLE: Lower Bounds on Cross-Entropy Loss in the Presence of Test-time Adversaries
http://proceedings.mlr.press/v139/bhagoji21a.html
AUTHORS: Arjun Nitin Bhagoji, Daniel Cullina, Vikash Sehwag, Prateek Mittal
HIGHLIGHT: In this paper, we determine optimal lower bounds on the cross-entropy loss in the presence of test-time adversaries, along with the corresponding optimal classification outputs.

81, TITLE: Additive Error Guarantees for Weighted Low Rank Approximation
http://proceedings.mlr.press/v139/bhaskara21a.html
AUTHORS: Aditya Bhaskara, Aravinda Kanchana Ruwanpathirana, Maheshakya Wijewardena
HIGHLIGHT: We study a natural greedy algorithm for weighted low rank approximation and develop a simple condition under which it yields bi-criteria approximation up to a small additive factor in the error.

82, TITLE: Sample Complexity of Robust Linear Classification on Separated Data
http://proceedings.mlr.press/v139/bhattacharjee21a.html
AUTHORS: Robi Bhattacharjee, Somesh Jha, Kamalika Chaudhuri
HIGHLIGHT: We consider the sample complexity of learning with adversarial robustness.

83, TITLE: Finding k in Latent $k-$ polytope
http://proceedings.mlr.press/v139/bhattacharyya21a.html
AUTHORS: Chiranjib Bhattacharyya, Ravindran Kannan, Amit Kumar
HIGHLIGHT: The first important contribution of this paper is to show that under \emph{standard assumptions} k equals the \INR of a \emph{subset smoothed data matrix} defined from Data generated from an $\LkP$.

84, TITLE: Non-Autoregressive Electron Redistribution Modeling for Reaction Prediction
http://proceedings.mlr.press/v139/bi21a.html
AUTHORS: Hangrui Bi, Hengyi Wang, Chence Shi, Connor Coley, Jian Tang, Hongyu Guo
HIGHLIGHT: To address these issues, we devise a non-autoregressive learning paradigm that predicts reaction in one shot.

85, TITLE: TempoRL: Learning When to Act
http://proceedings.mlr.press/v139/biedenkapp21a.html
AUTHORS: Andr? Biedenkapp, Raghu Rajan, Frank Hutter, Marius Lindauer
HIGHLIGHT: To address this, we propose a proactive setting in which the agent not only selects an action in a state but also for how long to commit to that action.

86, TITLE: Follow-the-Regularized-Leader Routes to Chaos in Routing Games
http://proceedings.mlr.press/v139/bielawski21a.html
AUTHORS: Jakub Bielawski, Thiparat Chotibut, Fryderyk Falniowski, Grzegorz Kosiorowski, Michal Misiurewicz, Georgios Piliouras
HIGHLIGHT: We study the emergence of chaotic behavior of Follow-the-Regularized Leader (FoReL) dynamics in games.

87, TITLE: Neural Symbolic Regression that scales
http://proceedings.mlr.press/v139/biggio21a.html
AUTHORS: Luca Biggio, Tommaso Bendinelli, Alexander Neitz, Aurelien Lucchi, Giambattista Parascandolo
HIGHLIGHT: In this paper, we introduce the first symbolic regression method that leverages large scale pre-training. We procedurally generate an unbounded set of equations, and simultaneously pre-train a Transformer to predict the symbolic equation from a corresponding set of input-output-pairs.

88, TITLE: Model Distillation for Revenue Optimization: Interpretable Personalized Pricing
http://proceedings.mlr.press/v139/biggs21a.html
AUTHORS: Max Biggs, Wei Sun, Markus Ettl
HIGHLIGHT: We present a novel, customized, prescriptive tree-based algorithm that distills knowledge from a complex black-box machine learning algorithm, segments customers with similar valuations and prescribes prices in such a way that maximizes revenue while maintaining interpretability.

89, TITLE: Scalable Normalizing Flows for Permutation Invariant Densities
http://proceedings.mlr.press/v139/bilos21a.html
AUTHORS: Marin Bilo?, Stephan G?nnemann
HIGHLIGHT: In this work, we demonstrate how calculating the trace, a crucial step in this method, raises issues that occur both during training and inference, limiting its practicality.

90, TITLE: Online Learning for Load Balancing of Unknown Monotone Resource Allocation Games
http://proceedings.mlr.press/v139/bistritz21a.html
AUTHORS: Ilai Bistritz, Nicholas Bambos
HIGHLIGHT: To overcome this, we propose a simple algorithm that learns to shift the NE of the game to meet the total load constraints by adjusting the pricing coefficients in an online manner.

91, TITLE: Low-Precision Reinforcement Learning: Running Soft Actor-Critic in Half Precision
http://proceedings.mlr.press/v139/bjorck21a.html
AUTHORS: Johan Bj?rck, Xiangyu Chen, Christopher De Sa, Carla P Gomes, Kilian Weinberger
HIGHLIGHT: In this paper we consider continuous control with the state-of-the-art SAC agent and demonstrate that a naïve adaptation of low-precision methods from supervised learning fails.

92, TITLE: Multiplying Matrices Without Multiplying
http://proceedings.mlr.press/v139/blalock21a.html
AUTHORS: Davis Blalock, John Guttag
HIGHLIGHT: Consequently, the task of efficiently approximating matrix products has received significant attention. We introduce a learning-based algorithm for this task that greatly outperforms existing methods.

93, TITLE: One for One, or All for All: Equilibria and Optimality of Collaboration in Federated Learning
http://proceedings.mlr.press/v139/blum21a.html
AUTHORS: Avrim Blum, Nika Haghtalab, Richard Lanas Phillips, Han Shao
HIGHLIGHT: Inspired by game theoretic notions, this paper introduces a framework for incentive-aware learning and data sharing in federated learning.

94, TITLE: Black-box density function estimation using recursive partitioning
http://proceedings.mlr.press/v139/bodin21a.html
AUTHORS: Erik Bodin, Zhenwen Dai, Neill Campbell, Carl Henrik Ek
HIGHLIGHT: We present a novel approach to Bayesian inference and general Bayesian computation that is defined through a sequential decision loop.

95, TITLE: Weisfeiler and Lehman Go Topological: Message Passing Simplicial Networks
http://proceedings.mlr.press/v139/bodnar21a.html
AUTHORS: Cristian Bodnar, Fabrizio Frasca, Yuguang Wang, Nina Otter, Guido F Montufar, Pietro Li?, Michael Bronstein
HIGHLIGHT: To overcome these limitations, we propose Message Passing Simplicial Networks (MPSNs), a class of models that perform message passing on simplicial complexes (SCs).

96, TITLE: The Hintons in your Neural Network: a Quantum Field Theory View of Deep Learning
http://proceedings.mlr.press/v139/bondesan21a.html
AUTHORS: Roberto Bondesan, Max Welling
HIGHLIGHT: In this work we develop a quantum field theory formalism for deep learning, where input signals are encoded in Gaussian states, a generalization of Gaussian processes which encode the agent’s uncertainty about the input signal.

97, TITLE: Offline Contextual Bandits with Overparameterized Models
http://proceedings.mlr.press/v139/brandfonbrener21a.html
AUTHORS: David Brandfonbrener, William Whitney, Rajesh Ranganath, Joan Bruna
HIGHLIGHT: We formally prove upper bounds on the regret of overparameterized value-based learning and lower bounds on the regret for policy-based algorithms.

98, TITLE: High-Performance Large-Scale Image Recognition Without Normalization
http://proceedings.mlr.press/v139/brock21a.html
AUTHORS: Andy Brock, Soham De, Samuel L Smith, Karen Simonyan
HIGHLIGHT: In this work, we develop an adaptive gradient clipping technique which overcomes these instabilities, and design a significantly improved class of Normalizer-Free ResNets.

99, TITLE: Evaluating the Implicit Midpoint Integrator for Riemannian Hamiltonian Monte Carlo
http://proceedings.mlr.press/v139/brofos21a.html
AUTHORS: James Brofos, Roy R Lederman
HIGHLIGHT: In this work, we examine the implicit midpoint integrator as an alternative to the generalized leapfrog integrator.

100, TITLE: Reinforcement Learning of Implicit and Explicit Control Flow Instructions
http://proceedings.mlr.press/v139/brooks21a.html
AUTHORS: Ethan Brooks, Janarthanan Rajendran, Richard L Lewis, Satinder Singh
HIGHLIGHT: We focus here on the problem of learning control flow that deviates from a strict step-by-step execution of instructions{—}that is, control flow that may skip forward over parts of the instructions or return backward to previously completed or skipped steps.

101, TITLE: Machine Unlearning for Random Forests
http://proceedings.mlr.press/v139/brophy21a.html
AUTHORS: Jonathan Brophy, Daniel Lowd
HIGHLIGHT: In this paper, we introduce data removal-enabled (DaRE) forests, a variant of random forests that enables the removal of training data with minimal retraining.

102, TITLE: Value Alignment Verification
http://proceedings.mlr.press/v139/brown21a.html
AUTHORS: Daniel S Brown, Jordan Schneider, Anca Dragan, Scott Niekum
HIGHLIGHT: In this paper we formalize and theoretically analyze the problem of efficient value alignment verification: how to efficiently test whether the behavior of another agent is aligned with a human’s values?

103, TITLE: Model-Free and Model-Based Policy Evaluation when Causality is Uncertain
http://proceedings.mlr.press/v139/bruns-smith21a.html
AUTHORS: David A Bruns-Smith
HIGHLIGHT: We develop worst-case bounds to assess sensitivity to these unobserved confounders in finite horizons when confounders are drawn iid each period.

104, TITLE: Narrow Margins: Classification, Margins and Fat Tails
http://proceedings.mlr.press/v139/buet-golfouse21a.html
AUTHORS: Francois Buet-Golfouse
HIGHLIGHT: We investigate the case where this convergence property is not guaranteed to hold and show that it can be fully characterised by the distribution of error terms in the latent variable interpretation of linear classifiers.

105, TITLE: Differentially Private Correlation Clustering
http://proceedings.mlr.press/v139/bun21a.html
AUTHORS: Mark Bun, Marek Elias, Janardhan Kulkarni
HIGHLIGHT: We propose an algorithm that achieves subquadratic additive error compared to the optimal cost.

106, TITLE: Disambiguation of Weak Supervision leading to Exponential Convergence rates
http://proceedings.mlr.press/v139/cabannnes21a.html
AUTHORS: Vivien A Cabannnes, Francis Bach, Alessandro Rudi
HIGHLIGHT: In this paper, we focus on partial labelling, an instance of weak supervision where, from a given input, we are given a set of potential targets.

107, TITLE: Finite mixture models do not reliably learn the number of components
http://proceedings.mlr.press/v139/cai21a.html
AUTHORS: Diana Cai, Trevor Campbell, Tamara Broderick
HIGHLIGHT: In this paper, we add rigor to data-analysis folk wisdom by proving that under even the slightest model misspecification, the FMM component-count posterior diverges: the posterior probability of any particular finite number of components converges to 0 in the limit of infinite data.

108, TITLE: A Theory of Label Propagation for Subpopulation Shift
http://proceedings.mlr.press/v139/cai21b.html
AUTHORS: Tianle Cai, Ruiqi Gao, Jason Lee, Qi Lei

HIGHLIGHT: In this work, we propose a provably effective framework based on label propagation by using an input consistency loss.

109, TITLE: Lenient Regret and Good-Action Identification in Gaussian Process Bandits
http://proceedings.mlr.press/v139/cai21c.html
AUTHORS: Xu Cai, Selwyn Gomes, Jonathan Scarlett
HIGHLIGHT: In this paper, we study the problem of Gaussian process (GP) bandits under relaxed optimization criteria stating that any function value above a certain threshold is “good enough”.

110, TITLE: A Zeroth-Order Block Coordinate Descent Algorithm for Huge-Scale Black-Box Optimization
http://proceedings.mlr.press/v139/cai21d.html
AUTHORS: Hanqin Cai, Yuchen Lou, Daniel Mckenzie, Wotao Yin
HIGHLIGHT: In this paper, we propose a novel algorithm, coined ZO-BCD, that exhibits favorable overall query complexity and has a much smaller per-iteration computational complexity.

111, TITLE: GraphNorm: A Principled Approach to Accelerating Graph Neural Network Training
http://proceedings.mlr.press/v139/cai21e.html
AUTHORS: Tianle Cai, Shengjie Luo, Keyulu Xu, Di He, Tie-Yan Liu, Liwei Wang
HIGHLIGHT: In this paper, we study what normalization is effective for Graph Neural Networks (GNNs).

112, TITLE: On Lower Bounds for Standard and Robust Gaussian Process Bandit Optimization
http://proceedings.mlr.press/v139/cai21f.html
AUTHORS: Xu Cai, Jonathan Scarlett
HIGHLIGHT: In this paper, we consider algorithm independent lower bounds for the problem of black-box optimization of functions having a bounded norm is some Reproducing Kernel Hilbert Space (RKHS), which can be viewed as a non-Bayesian Gaussian process bandit problem.

113, TITLE: High-dimensional Experimental Design and Kernel Bandits
http://proceedings.mlr.press/v139/camilleri21a.html
AUTHORS: Romain Camilleri, Kevin Jamieson, Julian Katz-Samuels
HIGHLIGHT: In this work, we propose a rounding procedure that frees N of any dependence on the dimension d, while achieving nearly the same performance guarantees of existing rounding procedures.

114, TITLE: A Gradient Based Strategy for Hamiltonian Monte Carlo Hyperparameter Optimization
http://proceedings.mlr.press/v139/campbell21a.html
AUTHORS: Andrew Campbell, Wenlong Chen, Vincent Stimper, Jose Miguel Hernandez-Lobato, Yichuan Zhang
HIGHLIGHT: Instead, we propose to optimize an objective that quantifies directly the speed of convergence to the target distribution.

115, TITLE: Asymmetric Heavy Tails and Implicit Bias in Gaussian Noise Injections
http://proceedings.mlr.press/v139/camuto21a.html
AUTHORS: Alexander Camuto, Xiaoyu Wang, Lingjiong Zhu, Chris Holmes, Mert Gurbuzbalaban, Umut Simsekli
HIGHLIGHT: In this paper, we focus on the so-called ‘implicit effect’ of GNIs, which is the effect of the injected noise on the dynamics of SGD.

116, TITLE: Fold2Seq: A Joint Sequence(1D)-Fold(3D) Embedding-based Generative Model for Protein Design
http://proceedings.mlr.press/v139/cao21a.html
AUTHORS: Yue Cao, Payel Das, Vijil Chenthamarakshan, Pin-Yu Chen, Igor Melnyk, Yang Shen
HIGHLIGHT: To overcome these challenges, we propose Fold2Seq, a novel transformer-based generative framework for designing protein sequences conditioned on a specific target fold.

117, TITLE: Learning from Similarity-Confidence Data
http://proceedings.mlr.press/v139/cao21b.html
AUTHORS: Yuzhou Cao, Lei Feng, Yitian Xu, Bo An, Gang Niu, Masashi Sugiyama
HIGHLIGHT: In this paper, we investigate a novel weakly supervised learning problem of learning from similarity-confidence (Sconf) data, where only unlabeled data pairs equipped with confidence that illustrates their degree of similarity (two examples are similar if they belong to the same class) are needed for training a discriminative binary classifier.

118, TITLE: Parameter-free Locally Accelerated Conditional Gradients
http://proceedings.mlr.press/v139/carderera21a.html

AUTHORS: Alejandro Carderera, Jelena Diakonikolas, Cheuk Yin Lin, Sebastian Pokutta
HIGHLIGHT: We remove this limitation by introducing a novel, Parameter-Free Locally accelerated CG (PF-LaCG) algorithm, for which we provide rigorous convergence guarantees.

119, TITLE: Optimizing persistent homology based functions
http://proceedings.mlr.press/v139/carriere21a.html
AUTHORS: Mathieu Carriere, Frederic Chazal, Marc Glisse, Yuichi Ike, Hariprasad Kannan, Yuhei Umeda
HIGHLIGHT: Building on real analytic geometry arguments, we propose a general framework that allows us to define and compute gradients for persistence-based functions in a very simple way.

120, TITLE: Online Policy Gradient for Model Free Learning of Linear Quadratic Regulators with $\sqrt$T Regret
http://proceedings.mlr.press/v139/cassel21a.html
AUTHORS: Asaf B Cassel, Tomer Koren
HIGHLIGHT: We present the first model-free algorithm that achieves similar regret guarantees.

121, TITLE: Multi-Receiver Online Bayesian Persuasion
http://proceedings.mlr.press/v139/castiglioni21a.html
AUTHORS: Matteo Castiglioni, Alberto Marchesi, Andrea Celli, Nicola Gatti
HIGHLIGHT: We study, for the first time, an online Bayesian persuasion setting with multiple receivers.

122, TITLE: Marginal Contribution Feature Importance - an Axiomatic Approach for Explaining Data
http://proceedings.mlr.press/v139/catav21a.html
AUTHORS: Amnon Catav, Boyang Fu, Yazeed Zoabi, Ahuva Libi Weiss Meilik, Noam Shomron, Jason Ernst, Sriram Sankararaman, Ran Gilad-Bachrach
HIGHLIGHT: Therefore, we develop a set of axioms to capture properties expected from a feature importance score when explaining data and prove that there exists only one score that satisfies all of them, the Marginal Contribution Feature Importance (MCI).

123, TITLE: Disentangling syntax and semantics in the brain with deep networks
http://proceedings.mlr.press/v139/caucheteux21a.html
AUTHORS: Charlotte Caucheteux, Alexandre Gramfort, Jean-Remi King
HIGHLIGHT: Overall, this study introduces a versatile framework to isolate, in the brain activity, the distributed representations of linguistic constructs.

124, TITLE: Fair Classification with Noisy Protected Attributes: A Framework with Provable Guarantees
http://proceedings.mlr.press/v139/celis21a.html
AUTHORS: L. Elisa Celis, Lingxiao Huang, Vijay Keswani, Nisheeth K. Vishnoi
HIGHLIGHT: We present an optimization framework for learning a fair classifier in the presence of noisy perturbations in the protected attributes.

125, TITLE: Best Model Identification: A Rested Bandit Formulation
http://proceedings.mlr.press/v139/cella21a.html
AUTHORS: Leonardo Cella, Massimiliano Pontil, Claudio Gentile
HIGHLIGHT: We introduce and analyze a best arm identification problem in the rested bandit setting, wherein arms are themselves learning algorithms whose expected losses decrease with the number of times the arm has been played.

126, TITLE: Revisiting Rainbow: Promoting more insightful and inclusive deep reinforcement learning research
http://proceedings.mlr.press/v139/ceron21a.html
AUTHORS: Johan Samir Obando Ceron, Pablo Samuel Castro
HIGHLIGHT: In this work we argue that, despite the community's emphasis on large-scale environments, the traditional small-scale environments can still yield valuable scientific insights and can help reduce the barriers to entry for underprivileged communities.

127, TITLE: Learning Routines for Effective Off-Policy Reinforcement Learning
http://proceedings.mlr.press/v139/cetin21a.html
AUTHORS: Edoardo Cetin, Oya Celiktutan
HIGHLIGHT: We propose a novel framework for reinforcement learning that effectively lifts such constraints.

128, TITLE: Learning Node Representations Using Stationary Flow Prediction on Large Payment and Cash Transaction Networks

http://proceedings.mlr.press/v139/ceylan21a.html
AUTHORS: Ciwan Ceylan, Salla Franz?n, Florian T. Pokorny
HIGHLIGHT: In this work, the gradient model is extended to a gated version and we prove that it, unlike the gradient model, is a universal approximator for flows on graphs.

129, TITLE: GRAND: Graph Neural Diffusion
http://proceedings.mlr.press/v139/chamberlain21a.html
AUTHORS: Ben Chamberlain, James Rowbottom, Maria I Gorinova, Michael Bronstein, Stefan Webb, Emanuele Rossi
HIGHLIGHT: We present Graph Neural Diffusion (GRAND) that approaches deep learning on graphs as a continuous diffusion process and treats Graph Neural Networks (GNNs) as discretisations of an underlying PDE.

130, TITLE: HoroPCA: Hyperbolic Dimensionality Reduction via Horospherical Projections
http://proceedings.mlr.press/v139/chami21a.html
AUTHORS: Ines Chami, Albert Gu, Dat P Nguyen, Christopher Re
HIGHLIGHT: We generalize each of these concepts to the hyperbolic space and propose HoroPCA, a method for hyperbolic dimensionality reduction.

131, TITLE: Goal-Conditioned Reinforcement Learning with Imagined Subgoals
http://proceedings.mlr.press/v139/chane-sane21a.html
AUTHORS: Elliot Chane-Sane, Cordelia Schmid, Ivan Laptev
HIGHLIGHT: In this work, we propose to incorporate imagined subgoals into policy learning to facilitate learning of complex tasks.

132, TITLE: Locally Private k-Means in One Round
http://proceedings.mlr.press/v139/chang21a.html
AUTHORS: Alisa Chang, Badih Ghazi, Ravi Kumar, Pasin Manurangsi
HIGHLIGHT: We provide an approximation algorithm for k-means clustering in the \emph{one-round} (aka \emph{non-interactive}) local model of differential privacy (DP).

133, TITLE: Modularity in Reinforcement Learning via Algorithmic Independence in Credit Assignment
http://proceedings.mlr.press/v139/chang21b.html
AUTHORS: Michael Chang, Sid Kaushik, Sergey Levine, Tom Griffiths
HIGHLIGHT: We introduce what we call the modularity criterion for testing whether a learning algorithm satisfies this constraint by performing causal analysis on the algorithm itself.

134, TITLE: Image-Level or Object-Level? A Tale of Two Resampling Strategies for Long-Tailed Detection
http://proceedings.mlr.press/v139/chang21c.html
AUTHORS: Nadine Chang, Zhiding Yu, Yu-Xiong Wang, Animashree Anandkumar, Sanja Fidler, Jose M Alvarez
HIGHLIGHT: We address object-level resampling by introducing an object-centric sampling strategy based on a dynamic, episodic memory bank.

135, TITLE: DeepWalking Backwards: From Embeddings Back to Graphs
http://proceedings.mlr.press/v139/chanpuriya21a.html
AUTHORS: Sudhanshu Chanpuriya, Cameron Musco, Konstantinos Sotiropoulos, Charalampos Tsourakakis
HIGHLIGHT: Focusing on a variant of the popular DeepWalk method \cite{PerozziAl-RfouSkiena:2014, QiuDongMa:2018}, we present algorithms for accurate embedding inversion – i.e., from the low-dimensional embedding of a graph G, we can find a graph $\tilde G$ with a very similar embedding.

136, TITLE: Differentiable Spatial Planning using Transformers
http://proceedings.mlr.press/v139/chaplot21a.html
AUTHORS: Devendra Singh Chaplot, Deepak Pathak, Jitendra Malik
HIGHLIGHT: We propose Spatial Planning Transformers (SPT), which given an obstacle map learns to generate actions by planning over long-range spatial dependencies, unlike prior data-driven planners that propagate information locally via convolutional structure in an iterative manner.

137, TITLE: Solving Challenging Dexterous Manipulation Tasks With Trajectory Optimisation and Reinforcement Learning
http://proceedings.mlr.press/v139/charlesworth21a.html
AUTHORS: Henry J Charlesworth, Giovanni Montana
HIGHLIGHT: In this work, we first introduce a suite of challenging simulated manipulation tasks where current reinforcement learning and trajectory optimisation techniques perform poorly.

138, TITLE: Classification with Rejection Based on Cost-sensitive Classification
http://proceedings.mlr.press/v139/charoenphakdee21a.html
AUTHORS: Nontawat Charoenphakdee, Zhenghang Cui, Yivan Zhang, Masashi Sugiyama
HIGHLIGHT: In this paper, based on the relationship between classification with rejection and cost-sensitive classification, we propose a novel method of classification with rejection by learning an ensemble of cost-sensitive classifiers, which satisfies all the following properties: (i) it can avoid estimating class-posterior probabilities, resulting in improved classification accuracy.

139, TITLE: Actionable Models: Unsupervised Offline Reinforcement Learning of Robotic Skills
http://proceedings.mlr.press/v139/chebotar21a.html
AUTHORS: Yevgen Chebotar, Karol Hausman, Yao Lu, Ted Xiao, Dmitry Kalashnikov, Jacob Varley, Alex Irpan, Benjamin Eysenbach, Ryan C Julian, Chelsea Finn, Sergey Levine
HIGHLIGHT: In particular, we propose the objective of learning a functional understanding of the environment by learning to reach any goal state in a given dataset.

140, TITLE: Unified Robust Semi-Supervised Variational Autoencoder
http://proceedings.mlr.press/v139/chen21a.html
AUTHORS: Xu Chen
HIGHLIGHT: In this paper, we propose a novel noise-robust semi-supervised deep generative model by jointly tackling noisy labels and outliers simultaneously in a unified robust semi-supervised variational autoencoder (URSVAE).

141, TITLE: Unsupervised Learning of Visual 3D Keypoints for Control
http://proceedings.mlr.press/v139/chen21b.html
AUTHORS: Boyuan Chen, Pieter Abbeel, Deepak Pathak
HIGHLIGHT: In this work, we propose a framework to learn such a 3D geometric structure directly from images in an end-to-end unsupervised manner.

142, TITLE: Integer Programming for Causal Structure Learning in the Presence of Latent Variables
http://proceedings.mlr.press/v139/chen21c.html
AUTHORS: Rui Chen, Sanjeeb Dash, Tian Gao
HIGHLIGHT: We propose a novel exact score-based method that solves an integer programming (IP) formulation and returns a score-maximizing ancestral ADMG for a set of continuous variables that follow a multivariate Gaussian distribution.

143, TITLE: Improved Corruption Robust Algorithms for Episodic Reinforcement Learning
http://proceedings.mlr.press/v139/chen21d.html
AUTHORS: Yifang Chen, Simon Du, Kevin Jamieson
HIGHLIGHT: We propose new algorithms which, compared to the existing results in \cite{lykouris2020corruption}, achieve strictly better regret bounds in terms of total corruptions for the tabular setting.

144, TITLE: Scalable Computations of Wasserstein Barycenter via Input Convex Neural Networks
http://proceedings.mlr.press/v139/chen21e.html
AUTHORS: Yongxin Chen, Jiaojiao Fan, Amirhossein Taghvaei
HIGHLIGHT: In this work, we present a novel scalable algorithm to approximate the Wasserstein Barycenters aiming at high-dimensional applications in machine learning.

145, TITLE: Neural Feature Matching in Implicit 3D Representations
http://proceedings.mlr.press/v139/chen21f.html
AUTHORS: Yunlu Chen, Basura Fernando, Hakan Bilen, Thomas Mensink, Efstratios Gavves
HIGHLIGHT: While the benefits from the global latent space do not correspond to explicit points at local level, we propose to track the continuous point trajectory by matching implicit features with the latent code interpolating between shapes, from which we corroborate the hierarchical functionality of the deep implicit functions, where early layers map the latent code to fitting the coarse shape structure, and deeper layers further refine the shape details.

146, TITLE: Decentralized Riemannian Gradient Descent on the Stiefel Manifold
http://proceedings.mlr.press/v139/chen21g.html
AUTHORS: Shixiang Chen, Alfredo Garcia, Mingyi Hong, Shahin Shahrampour
HIGHLIGHT: We present a decentralized Riemannian stochastic gradient method (DRSGD) with the convergence rate of $\mathcal{O}(1/\sqrt{K})$ to a stationary point.

147, TITLE: Learning Self-Modulating Attention in Continuous Time Space with Applications to Sequential Recommendation

http://proceedings.mlr.press/v139/chen21h.html
AUTHORS: Chao Chen, Haoyu Geng, Nianzu Yang, Junchi Yan, Daiyue Xue, Jianping Yu, Xiaokang Yang
HIGHLIGHT: In this paper, we propose a novel attention network, named \textit{self-modulating attention}, that models the complex and non-linearly evolving dynamic user preferences.

148, TITLE: Mandoline: Model Evaluation under Distribution Shift
http://proceedings.mlr.press/v139/chen21i.html
AUTHORS: Mayee Chen, Karan Goel, Nimit S Sohoni, Fait Poms, Kayvon Fatahalian, Christopher Re
HIGHLIGHT: Our key insight is that practitioners may have prior knowledge about the ways in which the distribution shifts, which we can use to better guide the importance weighting procedure.

149, TITLE: Order Matters: Probabilistic Modeling of Node Sequence for Graph Generation
http://proceedings.mlr.press/v139/chen21j.html
AUTHORS: Xiaohui Chen, Xu Han, Jiajing Hu, Francisco Ruiz, Liping Liu
HIGHLIGHT: In this work, we provide an expression for the likelihood of a graph generative model and show that its calculation is closely related to the problem of graph automorphism.

150, TITLE: CARTL: Cooperative Adversarially-Robust Transfer Learning
http://proceedings.mlr.press/v139/chen21k.html
AUTHORS: Dian Chen, Hongxin Hu, Qian Wang, Li Yinli, Cong Wang, Chao Shen, Qi Li
HIGHLIGHT: To address such a problem, we propose a novel cooperative adversarially-robust transfer learning (CARTL) by pre-training the model via feature distance minimization and fine-tuning the pre-trained model with non-expansive fine-tuning for target domain tasks.

151, TITLE: Finding the Stochastic Shortest Path with Low Regret: the Adversarial Cost and Unknown Transition Case
http://proceedings.mlr.press/v139/chen21l.html
AUTHORS: Liyu Chen, Haipeng Luo
HIGHLIGHT: Specifically, we develop algorithms that achieve $O(\sqrt{S^2ADT_\star K})$ regret for the full-information setting and $O(\sqrt{S^3A^2DT_\star K})$ regret for the bandit feedback setting, where D is the diameter, $T_\star$ is the expected hitting time of the optimal policy, S is the number of states, A is the number of actions, and K is the number of episodes.

152, TITLE: SpreadsheetCoder: Formula Prediction from Semi-structured Context
http://proceedings.mlr.press/v139/chen21m.html
AUTHORS: Xinyun Chen, Petros Maniatis, Rishabh Singh, Charles Sutton, Hanjun Dai, Max Lin, Denny Zhou
HIGHLIGHT: In this work, we present the first approach for synthesizing spreadsheet formulas from tabular context, which includes both headers and semi-structured tabular data.

153, TITLE: Large-Margin Contrastive Learning with Distance Polarization Regularizer
http://proceedings.mlr.press/v139/chen21n.html
AUTHORS: Shuo Chen, Gang Niu, Chen Gong, Jun Li, Jian Yang, Masashi Sugiyama
HIGHLIGHT: To this end, we propose \emph{large-margin contrastive learning} (LMCL) with \emph{distance polarization regularizer}, motivated by the distribution characteristic of pairwise distances in \emph{metric learning}.

154, TITLE: Z-GCNETs: Time Zigzags at Graph Convolutional Networks for Time Series Forecasting
http://proceedings.mlr.press/v139/chen21o.html
AUTHORS: Yuzhou Chen, Ignacio Segovia, Yulia R. Gel
HIGHLIGHT: As convergence of these two emerging ideas, we propose to enhance DL architectures with the most salient time-conditioned topological information of the data and introduce the concept of zigzag persistence into time-aware graph convolutional networks (GCNs).

155, TITLE: A Unified Lottery Ticket Hypothesis for Graph Neural Networks
http://proceedings.mlr.press/v139/chen21p.html
AUTHORS: Tianlong Chen, Yongduo Sui, Xuxi Chen, Aston Zhang, Zhangyang Wang
HIGHLIGHT: To this end, this paper first presents a unified GNN sparsification (UGS) framework that simultaneously prunes the graph adjacency matrix and the model weights, for effectively accelerating GNN inference on large-scale graphs. Leveraging this new tool, we further generalize the recently popular lottery ticket hypothesis to GNNs for the first time, by defining a graph lottery ticket (GLT) as a pair of core sub-dataset and sparse sub-network, which can be jointly identified from the original GNN and the full dense graph by iteratively applying UGS.

156, TITLE: Network Inference and Influence Maximization from Samples

http://proceedings.mlr.press/v139/chen21q.html
AUTHORS: Wei Chen, Xiaoming Sun, Jialin Zhang, Zhijie Zhang
HIGHLIGHT: In this paper, we consider the more realistic sampling setting where the network is unknown and we only have a set of passively observed cascades that record the set of activated nodes at each diffusion step.

157, TITLE: Data-driven Prediction of General Hamiltonian Dynamics via Learning Exactly-Symplectic Maps
http://proceedings.mlr.press/v139/chen21r.html
AUTHORS: Renyi Chen, Molei Tao
HIGHLIGHT: We consider the learning and prediction of nonlinear time series generated by a latent symplectic map.

158, TITLE: Analysis of stochastic Lanczos quadrature for spectrum approximation
http://proceedings.mlr.press/v139/chen21s.html
AUTHORS: Tyler Chen, Thomas Trogdon, Shashanka Ubaru
HIGHLIGHT: We present an error analysis for stochastic Lanczos quadrature (SLQ).

159, TITLE: Large-Scale Multi-Agent Deep FBSDEs
http://proceedings.mlr.press/v139/chen21t.html
AUTHORS: Tianrong Chen, Ziyi O Wang, Ioannis Exarchos, Evangelos Theodorou
HIGHLIGHT: In this paper we present a scalable deep learning framework for finding Markovian Nash Equilibria in multi-agent stochastic games using fictitious play.

160, TITLE: Representation Subspace Distance for Domain Adaptation Regression
http://proceedings.mlr.press/v139/chen21u.html
AUTHORS: Xinyang Chen, Sinan Wang, Jianmin Wang, Mingsheng Long
HIGHLIGHT: Based on this finding, we propose to close the domain gap through orthogonal bases of the representation spaces, which are free from feature scaling.

161, TITLE: Overcoming Catastrophic Forgetting by Bayesian Generative Regularization
http://proceedings.mlr.press/v139/chen21v.html
AUTHORS: Pei-Hung Chen, Wei Wei, Cho-Jui Hsieh, Bo Dai
HIGHLIGHT: In this paper, we propose a new method to over-come catastrophic forgetting by adding generative regularization to Bayesian inference frame-work.

162, TITLE: Cyclically Equivariant Neural Decoders for Cyclic Codes
http://proceedings.mlr.press/v139/chen21w.html
AUTHORS: Xiangyu Chen, Min Ye
HIGHLIGHT: In this work, we propose a novel neural decoder for cyclic codes by exploiting their cyclically invariant property.

163, TITLE: A Receptor Skeleton for Capsule Neural Networks
http://proceedings.mlr.press/v139/chen21x.html
AUTHORS: Jintai Chen, Hongyun Yu, Chengde Qian, Danny Z Chen, Jian Wu
HIGHLIGHT: This paper presents a new capsule structure, which contains a set of optimizable receptors and a transmitter is devised on the capsule's representation.

164, TITLE: Accelerating Gossip SGD with Periodic Global Averaging
http://proceedings.mlr.press/v139/chen21y.html
AUTHORS: Yiming Chen, Kun Yuan, Yingya Zhang, Pan Pan, Yinghui Xu, Wotao Yin
HIGHLIGHT: This paper introduces Gossip-PGA, which adds Periodic Global Averaging to accelerate Gossip SGD.

165, TITLE: ActNN: Reducing Training Memory Footprint via 2-Bit Activation Compressed Training
http://proceedings.mlr.press/v139/chen21z.html
AUTHORS: Jianfei Chen, Lianmin Zheng, Zhewei Yao, Dequan Wang, Ion Stoica, Michael Mahoney, Joseph Gonzalez
HIGHLIGHT: In this work, we propose ActNN, a memory-efficient training framework that stores randomly quantized activations for back propagation.

166, TITLE: SPADE: A Spectral Method for Black-Box Adversarial Robustness Evaluation
http://proceedings.mlr.press/v139/cheng21a.html
AUTHORS: Wuxinlin Cheng, Chenhui Deng, Zhiqiang Zhao, Yaohui Cai, Zhiru Zhang, Zhuo Feng

HIGHLIGHT: By leveraging the generalized Courant-Fischer theorem, we propose a SPADE score for evaluating the adversarial robustness of a given model, which is proved to be an upper bound of the best Lipschitz constant under the manifold setting.

167, TITLE: Self-supervised and Supervised Joint Training for Resource-rich Machine Translation
http://proceedings.mlr.press/v139/cheng21b.html
AUTHORS: Yong Cheng, Wei Wang, Lu Jiang, Wolfgang Macherey
HIGHLIGHT: In this paper, we propose a joint training approach, F2-XEnDec, to combine self-supervised and supervised learning to optimize NMT models.

168, TITLE: Exact Optimization of Conformal Predictors via Incremental and Decremental Learning
http://proceedings.mlr.press/v139/cherubin21a.html
AUTHORS: Giovanni Cherubin, Konstantinos Chatzikokolakis, Martin Jaggi
HIGHLIGHT: In this work, we show that it is possible to speed up a CP classifier considerably, by studying it in conjunction with the underlying ML method, and by exploiting incremental&decremental learning.

169, TITLE: Problem Dependent View on Structured Thresholding Bandit Problems
http://proceedings.mlr.press/v139/cheshire21a.html
AUTHORS: James Cheshire, Pierre Menard, Alexandra Carpentier
HIGHLIGHT: We investigate the \textit{problem dependent regime} in the stochastic \emph{Thresholding Bandit problem} (\tbp) under several \emph{shape constraints}.

170, TITLE: Online Optimization in Games via Control Theory: Connecting Regret, Passivity and Poincar? Recurrence
http://proceedings.mlr.press/v139/cheung21a.html
AUTHORS: Yun Kuen Cheung, Georgios Piliouras
HIGHLIGHT: We present a novel control-theoretic understanding of online optimization and learning in games, via the notion of passivity.

171, TITLE: Understanding and Mitigating Accuracy Disparity in Regression
http://proceedings.mlr.press/v139/chi21a.html
AUTHORS: Jianfeng Chi, Yuan Tian, Geoffrey J. Gordon, Han Zhao
HIGHLIGHT: In this paper, we study the accuracy disparity problem in regression.

172, TITLE: Private Alternating Least Squares: Practical Private Matrix Completion with Tighter Rates
http://proceedings.mlr.press/v139/chien21a.html
AUTHORS: Steve Chien, Prateek Jain, Walid Krichene, Steffen Rendle, Shuang Song, Abhradeep Thakurta, Li Zhang
HIGHLIGHT: We study the problem of differentially private (DP) matrix completion under user-level privacy.

173, TITLE: Light RUMs
http://proceedings.mlr.press/v139/chierichetti21a.html
AUTHORS: Flavio Chierichetti, Ravi Kumar, Andrew Tomkins
HIGHLIGHT: In this paper we consider the question of the (lossy) compressibility of RUMs on a universe of size n, i.e., the minimum number of bits required to approximate the winning probabilities of each slate.

174, TITLE: Parallelizing Legendre Memory Unit Training
http://proceedings.mlr.press/v139/chilkuri21a.html
AUTHORS: Narsimha Reddy Chilkuri, Chris Eliasmith
HIGHLIGHT: Here we leverage the linear time-invariant (LTI) memory component of the LMU to construct a simplified variant that can be parallelized during training (and yet executed as an RNN during inference), resulting in up to 200 times faster training.

175, TITLE: Quantifying and Reducing Bias in Maximum Likelihood Estimation of Structured Anomalies
http://proceedings.mlr.press/v139/chitra21a.html
AUTHORS: Uthsav Chitra, Kimberly Ding, Jasper C.H. Lee, Benjamin J Raphael
HIGHLIGHT: In this work, we demonstrate that in the normal means setting, the bias of the MLE depends on the size of the anomaly family.

176, TITLE: Robust Learning-Augmented Caching: An Experimental Study
http://proceedings.mlr.press/v139/chledowski21a.html
AUTHORS: Jakub Chledowski, Adam Polak, Bartosz Szabucki, Konrad Tomasz Zolna

HIGHLIGHT: We show that a straightforward method – blindly following either a predictor or a classical robust algorithm, and switching whenever one becomes worse than the other – has only a low overhead over a well-performing predictor, while competing with classical methods when the coupled predictor fails, thus providing a cheap worst-case insurance.

177, TITLE: Unifying Vision-and-Language Tasks via Text Generation
http://proceedings.mlr.press/v139/cho21a.html
AUTHORS: Jaemin Cho, Jie Lei, Hao Tan, Mohit Bansal
HIGHLIGHT: To alleviate these hassles, in this work, we propose a unified framework that learns different tasks in a single architecture with the same language modeling objective, i.e., multimodal conditional text generation, where our models learn to generate labels in text based on the visual and textual inputs.

178, TITLE: Learning from Nested Data with Ornstein Auto-Encoders
http://proceedings.mlr.press/v139/choi21a.html
AUTHORS: Youngwon Choi, Sungdong Lee, Joong-Ho Won
HIGHLIGHT: After identifying several issues with RIOAE, we present the product-space OAE (PSOAE) that minimizes a tighter upper bound of the distance and achieves orthogonality in the representation space.

179, TITLE: Variational Empowerment as Representation Learning for Goal-Conditioned Reinforcement Learning
http://proceedings.mlr.press/v139/choi21b.html
AUTHORS: Jongwook Choi, Archit Sharma, Honglak Lee, Sergey Levine, Shixiang Shane Gu
HIGHLIGHT: In this paper, we discuss how these two approaches {—} goal-conditioned RL (GCRL) and MI-based RL {—} can be generalized into a single family of methods, interpreting mutual information maximization and variational empowerment as representation learning methods that acquire function-ally aware state representations for goal reaching.

180, TITLE: Label-Only Membership Inference Attacks
http://proceedings.mlr.press/v139/choquette-choo21a.html
AUTHORS: Christopher A. Choquette-Choo, Florian Tramer, Nicholas Carlini, Nicolas Papernot
HIGHLIGHT: Whereas current attack methods all require access to the model's predicted confidence score, we introduce a label-only attack that instead evaluates the robustness of the model's predicted (hard) labels under perturbations of the input, to infer membership.

181, TITLE: Modeling Hierarchical Structures with Continuous Recursive Neural Networks
http://proceedings.mlr.press/v139/chowdhury21a.html
AUTHORS: Jishnu Ray Chowdhury, Cornelia Caragea
HIGHLIGHT: In this work, we propose Continuous Recursive Neural Network (CRvNN) as a backpropagation-friendly alternative to address the aforementioned limitations.

182, TITLE: Scaling Multi-Agent Reinforcement Learning with Selective Parameter Sharing
http://proceedings.mlr.press/v139/christianos21a.html
AUTHORS: Filippos Christianos, Georgios Papoudakis, Muhammad A Rahman, Stefano V Albrecht
HIGHLIGHT: We propose a novel method to automatically identify agents which may benefit from sharing parameters by partitioning them based on their abilities and goals.

183, TITLE: Beyond Variance Reduction: Understanding the True Impact of Baselines on Policy Optimization
http://proceedings.mlr.press/v139/chung21a.html
AUTHORS: Wesley Chung, Valentin Thomas, Marlos C. Machado, Nicolas Le Roux
HIGHLIGHT: In this paper we demonstrate that the standard view is too limited for bandit and RL problems.

184, TITLE: First-Order Methods for Wasserstein Distributionally Robust MDP
http://proceedings.mlr.press/v139/clement21a.html
AUTHORS: Julien Grand Clement, Christian Kroer
HIGHLIGHT: We propose a framework for solving Distributionally robust MDPs via first-order methods, and instantiate it for several types of Wasserstein ambiguity sets.

185, TITLE: Phasic Policy Gradient
http://proceedings.mlr.press/v139/cobbe21a.html
AUTHORS: Karl W Cobbe, Jacob Hilton, Oleg Klimov, John Schulman
HIGHLIGHT: We introduce Phasic Policy Gradient (PPG), a reinforcement learning framework which modifies traditional on-policy actor-critic methods by separating policy and value function training into distinct phases.

186, TITLE: Riemannian Convex Potential Maps
http://proceedings.mlr.press/v139/cohen21a.html
AUTHORS: Samuel Cohen, Brandon Amos, Yaron Lipman
HIGHLIGHT: We propose and study a class of flows that uses convex potentials from Riemannian optimal transport.

187, TITLE: Scaling Properties of Deep Residual Networks
http://proceedings.mlr.press/v139/cohen21b.html
AUTHORS: Alain-Sam Cohen, Rama Cont, Alain Rossier, Renyuan Xu
HIGHLIGHT: We investigate the properties of weights trained by stochastic gradient descent and their scaling with network depth through detailed numerical experiments.

188, TITLE: Differentially-Private Clustering of Easy Instances
http://proceedings.mlr.press/v139/cohen21c.html
AUTHORS: Edith Cohen, Haim Kaplan, Yishay Mansour, Uri Stemmer, Eliad Tsfadia
HIGHLIGHT: In this work we aim at providing simple implementable differentrially private clustering algorithms when the the data is "easy," e.g., when there exists a significant separation between the clusters.

189, TITLE: Improving Ultrametrics Embeddings Through Coresets
http://proceedings.mlr.press/v139/cohen-addad21a.html
AUTHORS: Vincent Cohen-Addad, R?mi De Joannis De Verclos, Guillaume Lagarde
HIGHLIGHT: We improve the above result and show how to improve the above guarantee from 5c to $\sqrt{2}$c+e while achieving the same asymptotic running time.

190, TITLE: Correlation Clustering in Constant Many Parallel Rounds
http://proceedings.mlr.press/v139/cohen-addad21b.html
AUTHORS: Vincent Cohen-Addad, Silvio Lattanzi, Slobodan Mitrovic, Ashkan Norouzi-Fard, Nikos Parotsidis, Jakub Tarnawski
HIGHLIGHT: In this work we propose a massively parallel computation (MPC) algorithm for this problem that is considerably faster than prior work.

191, TITLE: Concentric mixtures of Mallows models for top-k rankings: sampling and identifiability
http://proceedings.mlr.press/v139/collas21a.html
AUTHORS: Fabien Collas, Ekhine Irurozki
HIGHLIGHT: In this paper, we study mixtures of two Mallows models for top-k rankings with equal location parameters but with different scale parameters (a mixture of concentric Mallows models).

192, TITLE: Exploiting Shared Representations for Personalized Federated Learning
http://proceedings.mlr.press/v139/collins21a.html
AUTHORS: Liam Collins, Hamed Hassani, Aryan Mokhtari, Sanjay Shakkottai
HIGHLIGHT: Based on this intuition, we propose a novel federated learning framework and algorithm for learning a shared data representation across clients and unique local heads for each client.

193, TITLE: Differentiable Particle Filtering via Entropy-Regularized Optimal Transport
http://proceedings.mlr.press/v139/corenflos21a.html
AUTHORS: Adrien Corenflos, James Thornton, George Deligiannidis, Arnaud Doucet
HIGHLIGHT: By leveraging optimal transport ideas, we introduce a principled differentiable particle filter and provide convergence results.

194, TITLE: Fairness and Bias in Online Selection
http://proceedings.mlr.press/v139/correa21a.html
AUTHORS: Jose Correa, Andres Cristi, Paul Duetting, Ashkan Norouzi-Fard
HIGHLIGHT: We address the issues of fairness and bias in online selection by introducing multi-color versions of the classic secretary and prophet problem.

195, TITLE: Relative Deviation Margin Bounds
http://proceedings.mlr.press/v139/cortes21a.html
AUTHORS: Corinna Cortes, Mehryar Mohri, Ananda Theertha Suresh
HIGHLIGHT: We present a series of new and more favorable margin-based learning guarantees that depend on the empirical margin loss of a predictor.

196, TITLE: A Discriminative Technique for Multiple-Source Adaptation
http://proceedings.mlr.press/v139/cortes21b.html
AUTHORS: Corinna Cortes, Mehryar Mohri, Ananda Theertha Suresh, Ningshan Zhang
HIGHLIGHT: We present a new discriminative technique for the multiple-source adaptation (MSA) problem.

197, TITLE: Characterizing Fairness Over the Set of Good Models Under Selective Labels
http://proceedings.mlr.press/v139/coston21a.html
AUTHORS: Amanda Coston, Ashesh Rambachan, Alexandra Chouldechova
HIGHLIGHT: We develop a framework for characterizing predictive fairness properties over the set of models that deliver similar overall performance, or “the set of good models.”

198, TITLE: Two-way kernel matrix puncturing: towards resource-efficient PCA and spectral clustering
http://proceedings.mlr.press/v139/couillet21a.html
AUTHORS: Romain Couillet, Florent Chatelain, Nicolas Le Bihan
HIGHLIGHT: The article introduces an elementary cost and storage reduction method for spectral clustering and principal component analysis.

199, TITLE: Explaining Time Series Predictions with Dynamic Masks
http://proceedings.mlr.press/v139/crabbe21a.html
AUTHORS: Jonathan Crabb?, Mihaela Van Der Schaar
HIGHLIGHT: To address these challenges, we propose dynamic masks (Dynamask).

200, TITLE: Generalised Lipschitz Regularisation Equals Distributional Robustness
http://proceedings.mlr.press/v139/cranko21a.html
AUTHORS: Zac Cranko, Zhan Shi, Xinhua Zhang, Richard Nock, Simon Kornblith
HIGHLIGHT: In response, we have been able to significantly sharpen existing results regarding the relationship between distributional robustness and regularisation, when defined with a transportation cost uncertainty set.

201, TITLE: Environment Inference for Invariant Learning
http://proceedings.mlr.press/v139/creager21a.html
AUTHORS: Elliot Creager, Joern-Henrik Jacobsen, Richard Zemel
HIGHLIGHT: We propose EIIL, a general framework for domain-invariant learning that incorporates Environment Inference to directly infer partitions that are maximally informative for downstream Invariant Learning.

202, TITLE: Mind the Box: l_1-APGD for Sparse Adversarial Attacks on Image Classifiers
http://proceedings.mlr.press/v139/croce21a.html
AUTHORS: Francesco Croce, Matthias Hein
HIGHLIGHT: We show that when taking into account also the image domain $[0,1]^d$, established l_1-projected gradient descent (PGD) attacks are suboptimal as they do not consider that the effective threat model is the intersection of the l_1-ball and $[0,1]^d$.

203, TITLE: Parameterless Transductive Feature Re-representation for Few-Shot Learning
http://proceedings.mlr.press/v139/cui21a.html
AUTHORS: Wentao Cui, Yuhong Guo
HIGHLIGHT: In this paper, we propose a parameterless transductive feature re-representation framework that differs from all existing solutions from the following perspectives.

204, TITLE: Randomized Algorithms for Submodular Function Maximization with a k-System Constraint
http://proceedings.mlr.press/v139/cui21b.html
AUTHORS: Shuang Cui, Kai Han, Tianshuai Zhu, Jing Tang, Benwei Wu, He Huang
HIGHLIGHT: In this paper, we study the problem of non-negative submodular function maximization subject to a k-system constraint, which generalizes many other important constraints in submodular optimization such as cardinality constraint, matroid constraint, and k-extendible system constraint.

205, TITLE: GBHT: Gradient Boosting Histogram Transform for Density Estimation
http://proceedings.mlr.press/v139/cui21c.html
AUTHORS: Jingyi Cui, Hanyuan Hang, Yisen Wang, Zhouchen Lin
HIGHLIGHT: In this paper, we propose a density estimation algorithm called \textit{Gradient Boosting Histogram Transform} (GBHT), where we adopt the \textit{Negative Log Likelihood} as the loss function to make the boosting procedure available for the unsupervised tasks.

206, TITLE: ProGraML: A Graph-based Program Representation for Data Flow Analysis and Compiler Optimizations
http://proceedings.mlr.press/v139/cummins21a.html
AUTHORS: Chris Cummins, Zacharias V. Fisches, Tal Ben-Nun, Torsten Hoefler, Michael F P O?Boyle, Hugh Leather
HIGHLIGHT: We propose ProGraML - Program Graphs for Machine Learning - a language-independent, portable representation of program semantics.

207, TITLE: Combining Pessimism with Optimism for Robust and Efficient Model-Based Deep Reinforcement Learning
http://proceedings.mlr.press/v139/curi21a.html
AUTHORS: Sebastian Curi, Ilija Bogunovic, Andreas Krause
HIGHLIGHT: We propose the Robust Hallucinated Upper-Confidence RL (RH-UCRL) algorithm to provably solve this problem while attaining near-optimal sample complexity guarantees.

208, TITLE: Quantifying Availability and Discovery in Recommender Systems via Stochastic Reachability
http://proceedings.mlr.press/v139/curmei21a.html
AUTHORS: Mihaela Curmei, Sarah Dean, Benjamin Recht
HIGHLIGHT: In this work, we consider how preference models in interactive recommendation systems determine the availability of content and users' opportunities for discovery.

209, TITLE: Dynamic Balancing for Model Selection in Bandits and RL
http://proceedings.mlr.press/v139/cutkosky21a.html
AUTHORS: Ashok Cutkosky, Christoph Dann, Abhimanyu Das, Claudio Gentile, Aldo Pacchiano, Manish Purohit
HIGHLIGHT: We propose a framework for model selection by combining base algorithms in stochastic bandits and reinforcement learning.

210, TITLE: ConViT: Improving Vision Transformers with Soft Convolutional Inductive Biases
http://proceedings.mlr.press/v139/d-ascoli21a.html
AUTHORS: St?phane D?Ascoli, Hugo Touvron, Matthew L Leavitt, Ari S Morcos, Giulio Biroli, Levent Sagun
HIGHLIGHT: To this end, we introduce gated positional self-attention (GPSA), a form of positional self-attention which can be equipped with a "soft" convolutional inductive bias.

211, TITLE: Consistent regression when oblivious outliers overwhelm
http://proceedings.mlr.press/v139/d-orsi21a.html
AUTHORS: Tommaso D?Orsi, Gleb Novikov, David Steurer
HIGHLIGHT: We consider a robust linear regression model $y=X\beta^* + \eta$, where an adversary oblivious to the design $X\in \mathbb{R}^{n\times d}$ may choose η to corrupt all but an α fraction of the observations y in an arbitrary way.

212, TITLE: Offline Reinforcement Learning with Pseudometric Learning
http://proceedings.mlr.press/v139/dadashi21a.html
AUTHORS: Robert Dadashi, Shideh Rezaeifar, Nino Vieillard, L?onard Hussenot, Olivier Pietquin, Matthieu Geist
HIGHLIGHT: In this work, we propose an iterative procedure to learn a pseudometric (closely related to bisimulation metrics) from logged transitions, and use it to define this notion of closeness.

213, TITLE: A Tale of Two Efficient and Informative Negative Sampling Distributions
http://proceedings.mlr.press/v139/daghaghi21a.html
AUTHORS: Shabnam Daghaghi, Tharun Medini, Nicholas Meisburger, Beidi Chen, Mengnan Zhao, Anshumali Shrivastava
HIGHLIGHT: In this paper, we show two classes of distributions where the sampling scheme is truly adaptive and provably generates negative samples in near-constant time.

214, TITLE: SiameseXML: Siamese Networks meet Extreme Classifiers with 100M Labels
http://proceedings.mlr.press/v139/dahiya21a.html
AUTHORS: Kunal Dahiya, Ananye Agarwal, Deepak Saini, Gururaj K, Jian Jiao, Amit Singh, Sumeet Agarwal, Purushottam Kar, Manik Varma
HIGHLIGHT: To address these, this paper develops the SiameseXML framework based on a novel probabilistic model that naturally motivates a modular approach melding Siamese architectures with high-capacity extreme classifiers, and a training pipeline that effortlessly scales to tasks with 100 million labels.

215, TITLE: Fixed-Parameter and Approximation Algorithms for PCA with Outliers
http://proceedings.mlr.press/v139/dahiya21b.html
AUTHORS: Yogesh Dahiya, Fedor Fomin, Fahad Panolan, Kirill Simonov

HIGHLIGHT: We study this problem from the perspective of parameterized complexity by investigating how parameters like the dimension of the data, the subspace dimension, the number of outliers and their structure, and approximation error, influence the computational complexity of the problem.

216, TITLE: Sliced Iterative Normalizing Flows
http://proceedings.mlr.press/v139/dai21a.html
AUTHORS: Biwei Dai, Uros Seljak
HIGHLIGHT: We develop an iterative (greedy) deep learning (DL) algorithm which is able to transform an arbitrary probability distribution function (PDF) into the target PDF.

217, TITLE: Convex Regularization in Monte-Carlo Tree Search
http://proceedings.mlr.press/v139/dam21a.html
AUTHORS: Tuan Q Dam, Carlo D?Eramo, Jan Peters, Joni Pajarinen
HIGHLIGHT: In this paper, we overcome these limitations by introducing the use of convex regularization in Monte-Carlo Tree Search (MCTS) to drive exploration efficiently and to improve policy updates.

218, TITLE: Demonstration-Conditioned Reinforcement Learning for Few-Shot Imitation
http://proceedings.mlr.press/v139/dance21a.html
AUTHORS: Christopher R. Dance, Julien Perez, Th?o Cachet
HIGHLIGHT: We propose a novel approach to learning few-shot-imitation agents that we call demonstration-conditioned reinforcement learning (DCRL).

219, TITLE: Re-understanding Finite-State Representations of Recurrent Policy Networks
http://proceedings.mlr.press/v139/danesh21a.html
AUTHORS: Mohamad H Danesh, Anurag Koul, Alan Fern, Saeed Khorram
HIGHLIGHT: We introduce an approach for understanding control policies represented as recurrent neural networks.

220, TITLE: Newton Method over Networks is Fast up to the Statistical Precision
http://proceedings.mlr.press/v139/daneshmand21a.html
AUTHORS: Amir Daneshmand, Gesualdo Scutari, Pavel Dvurechensky, Alexander Gasnikov
HIGHLIGHT: We propose a distributed cubic regularization of the Newton method for solving (constrained) empirical risk minimization problems over a network of agents, modeled as undirected graph.

221, TITLE: BasisDeVAE: Interpretable Simultaneous Dimensionality Reduction and Feature-Level Clustering with Derivative-Based Variational Autoencoders
http://proceedings.mlr.press/v139/danks21a.html
AUTHORS: Dominic Danks, Christopher Yau
HIGHLIGHT: We present DeVAE, a novel VAE-based model with a derivative-based forward mapping, allowing for greater control over decoder behaviour via specification of the decoder function in derivative space.

222, TITLE: Intermediate Layer Optimization for Inverse Problems using Deep Generative Models
http://proceedings.mlr.press/v139/daras21a.html
AUTHORS: Giannis Daras, Joseph Dean, Ajil Jalal, Alex Dimakis
HIGHLIGHT: We propose Intermediate Layer Optimization (ILO), a novel optimization algorithm for solving inverse problems with deep generative models.

223, TITLE: Measuring Robustness in Deep Learning Based Compressive Sensing
http://proceedings.mlr.press/v139/darestani21a.html
AUTHORS: Mohammad Zalbagi Darestani, Akshay S Chaudhari, Reinhard Heckel
HIGHLIGHT: In order to understand the sensitivity to such perturbations, in this work, we measure the robustness of different approaches for image reconstruction including trained and un-trained neural networks as well as traditional sparsity-based methods.

224, TITLE: SAINT-ACC: Safety-Aware Intelligent Adaptive Cruise Control for Autonomous Vehicles Using Deep Reinforcement Learning
http://proceedings.mlr.press/v139/das21a.html
AUTHORS: Lokesh Chandra Das, Myounggyu Won
HIGHLIGHT: We present a novel adaptive cruise control (ACC) system namely SAINT-ACC: {S}afety-{A}ware {Int}elligent {ACC} system (SAINT-ACC) that is designed to achieve simultaneous optimization of traffic efficiency, driving safety, and driving comfort through dynamic adaptation of the inter-vehicle gap based on deep reinforcement learning (RL).

225, TITLE: Lipschitz normalization for self-attention layers with application to graph neural networks
http://proceedings.mlr.press/v139/dasoulas21a.html
AUTHORS: George Dasoulas, Kevin Scaman, Aladin Virmaux
HIGHLIGHT: In this work, we show that enforcing Lipschitz continuity by normalizing the attention scores can significantly improve the performance of deep attention models.

226, TITLE: Householder Sketch for Accurate and Accelerated Least-Mean-Squares Solvers
http://proceedings.mlr.press/v139/dass21a.html
AUTHORS: Jyotikrishna Dass, Rabi Mahapatra
HIGHLIGHT: In retrospect, we explore classical Householder transformation as a candidate for sketching and accurately solving LMS problems.

227, TITLE: Byzantine-Resilient High-Dimensional SGD with Local Iterations on Heterogeneous Data
http://proceedings.mlr.press/v139/data21a.html
AUTHORS: Deepesh Data, Suhas Diggavi
HIGHLIGHT: We provide convergence analyses for both strongly-convex and non-convex smooth objectives in the heterogeneous data setting.

228, TITLE: Catformer: Designing Stable Transformers via Sensitivity Analysis
http://proceedings.mlr.press/v139/davis21a.html
AUTHORS: Jared Q Davis, Albert Gu, Krzysztof Choromanski, Tri Dao, Christopher Re, Chelsea Finn, Percy Liang
HIGHLIGHT: In this paper, we improve upon recent analysis of Transformers and formalize a notion of sensitivity to capture the difficulty of training.

229, TITLE: Diffusion Source Identification on Networks with Statistical Confidence
http://proceedings.mlr.press/v139/dawkins21a.html
AUTHORS: Quinlan E Dawkins, Tianxi Li, Haifeng Xu
HIGHLIGHT: We introduce a statistical framework for the study of this problem and develop a confidence set inference approach inspired by hypothesis testing.

230, TITLE: Bayesian Deep Learning via Subnetwork Inference
http://proceedings.mlr.press/v139/daxberger21a.html
AUTHORS: Erik Daxberger, Eric Nalisnick, James U Allingham, Javier Antoran, Jose Miguel Hernandez-Lobato
HIGHLIGHT: In this work, we show that it suffices to perform inference over a small subset of model weights in order to obtain accurate predictive posteriors.

231, TITLE: Adversarial Robustness Guarantees for Random Deep Neural Networks
http://proceedings.mlr.press/v139/de-palma21a.html
AUTHORS: Giacomo De Palma, Bobak Kiani, Seth Lloyd
HIGHLIGHT: We explore the properties of adversarial examples for deep neural networks with random weights and biases, and prove that for any p$\geq$1, the \ell^p distance of any given input from the classification boundary scales as one over the square root of the dimension of the input times the \ell^p norm of the input.

232, TITLE: High-Dimensional Gaussian Process Inference with Derivatives
http://proceedings.mlr.press/v139/de-roos21a.html
AUTHORS: Filip De Roos, Alexandra Gessner, Philipp Hennig
HIGHLIGHT: We show that in the \emph{low-data} regime $N < d$, the="" gram="" matrix="" can="" be="" decomposed="" in="" a="" manner="" that="" reduces="" cost="" of="" inference="" to="" $\mathcal{o}(n^2d="" +="" (n^2)^3)$="" (i.e., linear="" number="" dimensions)="" and,="" special="" cases,="" n^3)$.

233, TITLE: Transfer-Based Semantic Anomaly Detection
http://proceedings.mlr.press/v139/deecke21a.html
AUTHORS: Lucas Deecke, Lukas Ruff, Robert A. Vandermeulen, Hakan Bilen
HIGHLIGHT: In this paper, we show that a previously overlooked strategy for anomaly detection (AD) is to introduce an explicit inductive bias toward representations transferred over from some large and varied semantic task.

234, TITLE: Grid-Functioned Neural Networks
http://proceedings.mlr.press/v139/dehesa21a.html
AUTHORS: Javier Dehesa, Andrew Vidler, Julian Padget, Christof Lutteroth
HIGHLIGHT: We introduce a new neural network architecture that we call "grid-functioned" neural networks.

235, TITLE: Multidimensional Scaling: Approximation and Complexity
http://proceedings.mlr.press/v139/demaine21a.html
AUTHORS: Erik Demaine, Adam Hesterberg, Frederic Koehler, Jayson Lynch, John Urschel
HIGHLIGHT: In this paper, we prove that minimizing the Kamada-Kawai objective is NP-hard and give a provable approximation algorithm for optimizing it, which in particular is a PTAS on low-diameter graphs.

236, TITLE: What Does Rotation Prediction Tell Us about Classifier Accuracy under Varying Testing Environments?
http://proceedings.mlr.press/v139/deng21a.html
AUTHORS: Weijian Deng, Stephen Gould, Liang Zheng
HIGHLIGHT: In this work, we train semantic classification and rotation prediction in a multi-task way.

237, TITLE: Toward Better Generalization Bounds with Locally Elastic Stability
http://proceedings.mlr.press/v139/deng21b.html
AUTHORS: Zhun Deng, Hangfeng He, Weijie Su
HIGHLIGHT: Given that, we propose \emph{locally elastic stability} as a weaker and distribution-dependent stability notion, which still yields exponential generalization bounds.

238, TITLE: Revenue-Incentive Tradeoffs in Dynamic Reserve Pricing
http://proceedings.mlr.press/v139/deng21c.html
AUTHORS: Yuan Deng, Sebastien Lahaie, Vahab Mirrokni, Song Zuo
HIGHLIGHT: In this paper, we study how to set reserves to boost revenue based on the historical bids of strategic buyers, while controlling the impact of such a policy on the incentive compatibility of the repeated auctions.

239, TITLE: Heterogeneity for the Win: One-Shot Federated Clustering
http://proceedings.mlr.press/v139/dennis21a.html
AUTHORS: Don Kurian Dennis, Tian Li, Virginia Smith
HIGHLIGHT: In this work, we explore the unique challenges—and opportunities—of unsupervised federated learning (FL).

240, TITLE: Kernel Continual Learning
http://proceedings.mlr.press/v139/derakhshani21a.html
AUTHORS: Mohammad Mahdi Derakhshani, Xiantong Zhen, Ling Shao, Cees Snoek
HIGHLIGHT: This paper introduces kernel continual learning, a simple but effective variant of continual learning that leverages the non-parametric nature of kernel methods to tackle catastrophic forgetting.

241, TITLE: Bayesian Optimization over Hybrid Spaces
http://proceedings.mlr.press/v139/deshwal21a.html
AUTHORS: Aryan Deshwal, Syrine Belakaria, Janardhan Rao Doppa
HIGHLIGHT: In this paper, we propose a novel approach referred as Hybrid Bayesian Optimization (HyBO) by utilizing diffusion kernels, which are naturally defined over continuous and discrete variables.

242, TITLE: Navigation Turing Test (NTT): Learning to Evaluate Human-Like Navigation
http://proceedings.mlr.press/v139/devlin21a.html
AUTHORS: Sam Devlin, Raluca Georgescu, Ida Momennejad, Jaroslaw Rzepecki, Evelyn Zuniga, Gavin Costello, Guy Leroy, Ali Shaw, Katja Hofmann
HIGHLIGHT: We address these limitations through a novel automated Navigation Turing Test (ANTT) that learns to predict human judgments of human-likeness.

243, TITLE: Versatile Verification of Tree Ensembles
http://proceedings.mlr.press/v139/devos21a.html
AUTHORS: Laurens Devos, Wannes Meert, Jesse Davis
HIGHLIGHT: This paper introduces a generic algorithm called Veritas that enables tackling multiple different verification tasks for tree ensemble models like random forests (RFs) and gradient boosted decision trees (GBDTs).

244, TITLE: On the Inherent Regularization Effects of Noise Injection During Training
http://proceedings.mlr.press/v139/dhifallah21a.html
AUTHORS: Oussama Dhifallah, Yue Lu
HIGHLIGHT: In this paper, we present a theoretical study of one particular way of random perturbation, which corresponds to injecting artificial noise to the training data.

245, TITLE: Hierarchical Agglomerative Graph Clustering in Nearly-Linear Time
http://proceedings.mlr.press/v139/dhulipala21a.html
AUTHORS: Laxman Dhulipala, David Eisenstat, Jakub Lacki, Vahab Mirrokni, Jessica Shi
HIGHLIGHT: We study the widely-used hierarchical agglomerative clustering (HAC) algorithm on edge-weighted graphs.

246, TITLE: Learning Online Algorithms with Distributional Advice
http://proceedings.mlr.press/v139/diakonikolas21a.html
AUTHORS: Ilias Diakonikolas, Vasilis Kontonis, Christos Tzamos, Ali Vakilian, Nikos Zarifis
HIGHLIGHT: We study the problem of designing online algorithms given advice about the input.

247, TITLE: A Wasserstein Minimax Framework for Mixed Linear Regression
http://proceedings.mlr.press/v139/diamandis21a.html
AUTHORS: Theo Diamandis, Yonina Eldar, Alireza Fallah, Farzan Farnia, Asuman Ozdaglar
HIGHLIGHT: We propose an optimal transport-based framework for MLR problems, Wasserstein Mixed Linear Regression (WMLR), which minimizes the Wasserstein distance between the learned and target mixture regression models.

248, TITLE: Context-Aware Online Collective Inference for Templated Graphical Models
http://proceedings.mlr.press/v139/dickens21a.html
AUTHORS: Charles Dickens, Connor Pryor, Eriq Augustine, Alexander Miller, Lise Getoor
HIGHLIGHT: In this work, we examine online collective inference, the problem of maintaining and performing inference over a sequence of evolving graphical models.

249, TITLE: ARMS: Antithetic-REINFORCE-Multi-Sample Gradient for Binary Variables
http://proceedings.mlr.press/v139/dimitriev21a.html
AUTHORS: Aleksandar Dimitriev, Mingyuan Zhou
HIGHLIGHT: To better utilize more than two samples, we propose ARMS, an Antithetic REINFORCE-based Multi-Sample gradient estimator.

250, TITLE: XOR-CD: Linearly Convergent Constrained Structure Generation
http://proceedings.mlr.press/v139/ding21a.html
AUTHORS: Fan Ding, Jianzhu Ma, Jinbo Xu, Yexiang Xue
HIGHLIGHT: We propose XOR-Contrastive Divergence learning (XOR-CD), a provable approach for constrained structure generation, which remains difficult for state-of-the-art neural network and constraint reasoning approaches.

251, TITLE: Dual Principal Component Pursuit for Robust Subspace Learning: Theory and Algorithms for a Holistic Approach
http://proceedings.mlr.press/v139/ding21b.html
AUTHORS: Tianyu Ding, Zhihui Zhu, Rene Vidal, Daniel P Robinson
HIGHLIGHT: In this paper, we consider a DPCP approach for simultaneously computing the entire basis of the orthogonal complement subspace (we call this a holistic approach) by solving a non-convex non-smooth optimization problem over the Grassmannian.

252, TITLE: Coded-InvNet for Resilient Prediction Serving Systems
http://proceedings.mlr.press/v139/dinh21a.html
AUTHORS: Tuan Dinh, Kangwook Lee
HIGHLIGHT: Inspired by a new coded computation algorithm for invertible functions, we propose Coded-InvNet a new approach to design resilient prediction serving systems that can gracefully handle stragglers or node failures.

253, TITLE: Estimation and Quantization of Expected Persistence Diagrams
http://proceedings.mlr.press/v139/divol21a.html
AUTHORS: Vincent Divol, Theo Lacombe
HIGHLIGHT: In this article, we study two such summaries, the Expected Persistence Diagram (EPD), and its quantization.

254, TITLE: On Energy-Based Models with Overparametrized Shallow Neural Networks
http://proceedings.mlr.press/v139/domingo-enrich21a.html
AUTHORS: Carles Domingo-Enrich, Alberto Bietti, Eric Vanden-Eijnden, Joan Bruna
HIGHLIGHT: Building from the incipient theory of overparametrized neural networks, we show that models trained in the so-called 'active' regime provide a statistical advantage over their associated 'lazy' or kernel regime, leading to improved adaptivity to hidden low-dimensional structure in the data distribution, as already observed in supervised learning.

255, TITLE: Kernel-Based Reinforcement Learning: A Finite-Time Analysis
http://proceedings.mlr.press/v139/domingues21a.html
AUTHORS: Omar Darwiche Domingues, Pierre Menard, Matteo Pirotta, Emilie Kaufmann, Michal Valko
HIGHLIGHT: We introduce Kernel-UCBVI, a model-based optimistic algorithm that leverages the smoothness of the MDP and a non-parametric kernel estimator of the rewards and transitions to efficiently balance exploration and exploitation.

256, TITLE: Attention is not all you need: pure attention loses rank doubly exponentially with depth
http://proceedings.mlr.press/v139/dong21a.html
AUTHORS: Yihe Dong, Jean-Baptiste Cordonnier, Andreas Loukas
HIGHLIGHT: This work proposes a new way to understand self-attention networks: we show that their output can be decomposed into a sum of smaller terms—or paths—each involving the operation of a sequence of attention heads across layers.

257, TITLE: How rotational invariance of common kernels prevents generalization in high dimensions
http://proceedings.mlr.press/v139/donhauser21a.html
AUTHORS: Konstantin Donhauser, Mingqi Wu, Fanny Yang
HIGHLIGHT: In this paper, we show that in high dimensions, the rotational invariance property of commonly studied kernels (such as RBF, inner product kernels and fully-connected NTK of any depth) leads to inconsistent estimation unless the ground truth is a low-degree polynomial.

258, TITLE: Fast Stochastic Bregman Gradient Methods: Sharp Analysis and Variance Reduction
http://proceedings.mlr.press/v139/dragomir21a.html
AUTHORS: Radu Alexandru Dragomir, Mathieu Even, Hadrien Hendrikx
HIGHLIGHT: We study the problem of minimizing a relatively-smooth convex function using stochastic Bregman gradient methods.

259, TITLE: Bilinear Classes: A Structural Framework for Provable Generalization in RL
http://proceedings.mlr.press/v139/du21a.html
AUTHORS: Simon Du, Sham Kakade, Jason Lee, Shachar Lovett, Gaurav Mahajan, Wen Sun, Ruosong Wang
HIGHLIGHT: This work introduces Bilinear Classes, a new structural framework, which permit generalization in reinforcement learning in a wide variety of settings through the use of function approximation.

260, TITLE: Improved Contrastive Divergence Training of Energy-Based Models
http://proceedings.mlr.press/v139/du21b.html
AUTHORS: Yilun Du, Shuang Li, Joshua Tenenbaum, Igor Mordatch
HIGHLIGHT: We propose an adaptation to improve contrastive divergence training by scrutinizing a gradient term that is difficult to calculate and is often left out for convenience.

261, TITLE: Order-Agnostic Cross Entropy for Non-Autoregressive Machine Translation
http://proceedings.mlr.press/v139/du21c.html
AUTHORS: Cunxiao Du, Zhaopeng Tu, Jing Jiang
HIGHLIGHT: We propose a new training objective named order-agnostic cross entropy (OaXE) for fully non-autoregressive translation (NAT) models.

262, TITLE: Putting the "Learning" into Learning-Augmented Algorithms for Frequency Estimation
http://proceedings.mlr.press/v139/du21d.html
AUTHORS: Elbert Du, Franklyn Wang, Michael Mitzenmacher
HIGHLIGHT: Learning here is used to predict heavy hitters from a data stream, which are counted explicitly outside the sketch.

263, TITLE: Estimating a-Rank from A Few Entries with Low Rank Matrix Completion
http://proceedings.mlr.press/v139/du21e.html
AUTHORS: Yali Du, Xue Yan, Xu Chen, Jun Wang, Haifeng Zhang
HIGHLIGHT: In this paper, we aim to reduce the number of pairwise comparisons in recovering a satisfying ranking for n strategies in two-player meta-games, by exploring the fact that agents with similar skills may achieve similar payoffs against others.

264, TITLE: Learning Diverse-Structured Networks for Adversarial Robustness
http://proceedings.mlr.press/v139/du21f.html
AUTHORS: Xuefeng Du, Jingfeng Zhang, Bo Han, Tongliang Liu, Yu Rong, Gang Niu, Junzhou Huang, Masashi Sugiyama
HIGHLIGHT: In this paper, we argue that NA and AT cannot be handled independently, since given a dataset, the optimal NA in ST would be no longer optimal in AT.

265, TITLE: Risk Bounds and Rademacher Complexity in Batch Reinforcement Learning
http://proceedings.mlr.press/v139/duan21a.html
AUTHORS: Yaqi Duan, Chi Jin, Zhiyuan Li
HIGHLIGHT: This paper considers batch Reinforcement Learning (RL) with general value function approximation.

266, TITLE: Sawtooth Factorial Topic Embeddings Guided Gamma Belief Network
http://proceedings.mlr.press/v139/duan21b.html
AUTHORS: Zhibin Duan, Dongsheng Wang, Bo Chen, Chaojie Wang, Wenchao Chen, Yewen Li, Jie Ren, Mingyuan Zhou
HIGHLIGHT: To relax this assumption, we propose sawtooth factorial topic embedding guided GBN, a deep generative model of documents that captures the dependencies and semantic similarities between the topics in the embedding space.

267, TITLE: Exponential Reduction in Sample Complexity with Learning of Ising Model Dynamics
http://proceedings.mlr.press/v139/dutt21a.html
AUTHORS: Arkopal Dutt, Andrey Lokhov, Marc D Vuffray, Sidhant Misra
HIGHLIGHT: We study the problem of reconstructing binary graphical models from correlated samples produced by a dynamical process, which is natural in many applications.

268, TITLE: Reinforcement Learning Under Moral Uncertainty
http://proceedings.mlr.press/v139/ecoffet21a.html
AUTHORS: Adrien Ecoffet, Joel Lehman
HIGHLIGHT: This paper translates such insights to the field of reinforcement learning, proposes two training methods that realize different points among competing desiderata, and trains agents in simple environments to act under moral uncertainty.

269, TITLE: Confidence-Budget Matching for Sequential Budgeted Learning
http://proceedings.mlr.press/v139/efroni21a.html
AUTHORS: Yonathan Efroni, Nadav Merlis, Aadirupa Saha, Shie Mannor
HIGHLIGHT: In this work, we formalize decision-making problems with querying budget, where there is a (possibly time-dependent) hard limit on the number of reward queries allowed.

270, TITLE: Self-Paced Context Evaluation for Contextual Reinforcement Learning
http://proceedings.mlr.press/v139/eimer21a.html
AUTHORS: Theresa Eimer, Andr? Biedenkapp, Frank Hutter, Marius Lindauer
HIGHLIGHT: To improve sample efficiency for learning on such instances of a problem domain, we present Self-Paced Context Evaluation (SPaCE).

271, TITLE: Provably Strict Generalisation Benefit for Equivariant Models
http://proceedings.mlr.press/v139/elesedy21a.html
AUTHORS: Bryn Elesedy, Sheheryar Zaidi
HIGHLIGHT: By considering the simplest case of linear models, this paper provides the first provably non-zero improvement in generalisation for invariant/equivariant models when the target distribution is invariant/equivariant with respect to a compact group.

272, TITLE: Efficient Iterative Amortized Inference for Learning Symmetric and Disentangled Multi-Object Representations
http://proceedings.mlr.press/v139/emami21a.html
AUTHORS: Patrick Emami, Pan He, Sanjay Ranka, Anand Rangarajan
HIGHLIGHT: In this work, we introduce EfficientMORL, an efficient framework for the unsupervised learning of object-centric representations.

273, TITLE: Implicit Bias of Linear RNNs
http://proceedings.mlr.press/v139/emami21b.html
AUTHORS: Melikasadat Emami, Mojtaba Sahraee-Ardakan, Parthe Pandit, Sundeep Rangan, Alyson K Fletcher
HIGHLIGHT: However, RNNs' poor ability to capture long-term dependencies has not been fully understood. This paper provides a rigorous explanation of this property in the special case of linear RNNs.

274, TITLE: Global Optimality Beyond Two Layers: Training Deep ReLU Networks via Convex Programs
http://proceedings.mlr.press/v139/ergen21a.html
AUTHORS: Tolga Ergen, Mert Pilanci
HIGHLIGHT: In this paper, we develop a novel unified framework to reveal a hidden regularization mechanism through the lens of convex optimization.

275, TITLE: Revealing the Structure of Deep Neural Networks via Convex Duality
http://proceedings.mlr.press/v139/ergen21b.html
AUTHORS: Tolga Ergen, Mert Pilanci
HIGHLIGHT: We study regularized deep neural networks (DNNs) and introduce a convex analytic framework to characterize the structure of the hidden layers.

276, TITLE: Whitening for Self-Supervised Representation Learning
http://proceedings.mlr.press/v139/ermolov21a.html
AUTHORS: Aleksandr Ermolov, Aliaksandr Siarohin, Enver Sangineto, Nicu Sebe
HIGHLIGHT: In this paper, we propose a different direction and a new loss function for SSL, which is based on the whitening of the latent-space features.

277, TITLE: Graph Mixture Density Networks
http://proceedings.mlr.press/v139/errica21a.html
AUTHORS: Federico Errica, Davide Bacciu, Alessio Micheli
HIGHLIGHT: We introduce the Graph Mixture Density Networks, a new family of machine learning models that can fit multimodal output distributions conditioned on graphs of arbitrary topology.

278, TITLE: Cross-Gradient Aggregation for Decentralized Learning from Non-IID Data
http://proceedings.mlr.press/v139/esfandiari21a.html
AUTHORS: Yasaman Esfandiari, Sin Yong Tan, Zhanhong Jiang, Aditya Balu, Ethan Herron, Chinmay Hegde, Soumik Sarkar
HIGHLIGHT: Inspired by ideas from continual learning, we propose Cross-Gradient Aggregation (CGA), a novel decentralized learning algorithm where (i) each agent aggregates cross-gradient information, i.e., derivatives of its model with respect to its neighbors' datasets, and (ii) updates its model using a projected gradient based on quadratic programming (QP).

279, TITLE: Weight-covariance alignment for adversarially robust neural networks
http://proceedings.mlr.press/v139/eustratiadis21a.html
AUTHORS: Panagiotis Eustratiadis, Henry Gouk, Da Li, Timothy Hospedales
HIGHLIGHT: We propose a new SNN that achieves state-of-the-art performance without relying on adversarial training, and enjoys solid theoretical justification.

280, TITLE: Data augmentation for deep learning based accelerated MRI reconstruction with limited data
http://proceedings.mlr.press/v139/fabian21a.html
AUTHORS: Zalan Fabian, Reinhard Heckel, Mahdi Soltanolkotabi
HIGHLIGHT: Inspired by the success of Data Augmentation (DA) for classification problems, in this paper, we propose a pipeline for data augmentation for accelerated MRI reconstruction and study its effectiveness at reducing the required training data in a variety of settings.

281, TITLE: Poisson-Randomised DirBN: Large Mutation is Needed in Dirichlet Belief Networks
http://proceedings.mlr.press/v139/fan21a.html
AUTHORS: Xuhui Fan, Bin Li, Yaqiong Li, Scott A. Sisson
HIGHLIGHT: In this work, we propose Poisson-randomised Dirichlet Belief Networks (Pois-DirBN), which allows large mutations for the latent distributions across layers to enlarge the representation capability.

282, TITLE: Model-based Reinforcement Learning for Continuous Control with Posterior Sampling
http://proceedings.mlr.press/v139/fan21b.html
AUTHORS: Ying Fan, Yifei Ming
HIGHLIGHT: In this paper, we study model-based posterior sampling for reinforcement learning (PSRL) in continuous state-action spaces theoretically and empirically.

283, TITLE: SECANT: Self-Expert Cloning for Zero-Shot Generalization of Visual Policies
http://proceedings.mlr.press/v139/fan21c.html
AUTHORS: Linxi Fan, Guanzhi Wang, De-An Huang, Zhiding Yu, Li Fei-Fei, Yuke Zhu, Animashree Anandkumar
HIGHLIGHT: In this work, we consider robust policy learning which targets zero-shot generalization to unseen visual environments with large distributional shift.

284, TITLE: On Estimation in Latent Variable Models
http://proceedings.mlr.press/v139/fang21a.html
AUTHORS: Guanhua Fang, Ping Li

HIGHLIGHT: In this paper, we consider a gradient based method via using variance reduction technique to accelerate estimation procedure.

285, TITLE: On Variational Inference in Biclustering Models
http://proceedings.mlr.press/v139/fang21b.html
AUTHORS: Guanhua Fang, Ping Li
HIGHLIGHT: In this paper, we develop a theory for the estimation of general biclustering models, where the data is assumed to follow certain statistical distribution with underlying biclustering structure.

286, TITLE: Learning Bounds for Open-Set Learning
http://proceedings.mlr.press/v139/fang21c.html
AUTHORS: Zhen Fang, Jie Lu, Anjin Liu, Feng Liu, Guangquan Zhang
HIGHLIGHT: In this paper, we target a more challenging and re_x0002_alistic setting: open-set learning (OSL), where there exist test samples from the classes that are unseen during training.

287, TITLE: Streaming Bayesian Deep Tensor Factorization
http://proceedings.mlr.press/v139/fang21d.html
AUTHORS: Shikai Fang, Zheng Wang, Zhimeng Pan, Ji Liu, Shandian Zhe
HIGHLIGHT: More important, for highly expressive, deep factorization, we lack an effective approach to handle streaming data, which are ubiquitous in real-world applications. To address these issues, we propose SBTD, a Streaming Bayesian Deep Tensor factorization method.

288, TITLE: PID Accelerated Value Iteration Algorithm
http://proceedings.mlr.press/v139/farahmand21a.html
AUTHORS: Amir-Massoud Farahmand, Mohammad Ghavamzadeh
HIGHLIGHT: We propose modifications to VI in order to potentially accelerate its convergence behaviour.

289, TITLE: Near-Optimal Entrywise Anomaly Detection for Low-Rank Matrices with Sub-Exponential Noise
http://proceedings.mlr.press/v139/farias21a.html
AUTHORS: Vivek Farias, Andrew A Li, Tianyi Peng
HIGHLIGHT: So motivated, we propose a conceptually simple entrywise approach to anomaly detection in low-rank matrices.

290, TITLE: Connecting Optimal Ex-Ante Collusion in Teams to Extensive-Form Correlation: Faster Algorithms and Positive Complexity Results
http://proceedings.mlr.press/v139/farina21a.html
AUTHORS: Gabriele Farina, Andrea Celli, Nicola Gatti, Tuomas Sandholm
HIGHLIGHT: We focus on the problem of finding an optimal strategy for a team of players that faces an opponent in an imperfect-information zero-sum extensive-form game.

291, TITLE: Train simultaneously, generalize better: Stability of gradient-based minimax learners
http://proceedings.mlr.press/v139/farnia21a.html
AUTHORS: Farzan Farnia, Asuman Ozdaglar
HIGHLIGHT: In this paper, we show that the optimization algorithm also plays a key role in the generalization performance of the trained minimax model.

292, TITLE: Unbalanced minibatch Optimal Transport; applications to Domain Adaptation
http://proceedings.mlr.press/v139/fatras21a.html
AUTHORS: Kilian Fatras, Thibault Sejourne, R?mi Flamary, Nicolas Courty
HIGHLIGHT: As an alternative, we suggest that the same minibatch strategy coupled with unbalanced optimal transport can yield more robust behaviors.

293, TITLE: Risk-Sensitive Reinforcement Learning with Function Approximation: A Debiasing Approach
http://proceedings.mlr.press/v139/fei21a.html
AUTHORS: Yingjie Fei, Zhuoran Yang, Zhaoran Wang
HIGHLIGHT: We study function approximation for episodic reinforcement learning with entropic risk measure.

294, TITLE: Lossless Compression of Efficient Private Local Randomizers
http://proceedings.mlr.press/v139/feldman21a.html
AUTHORS: Vitaly Feldman, Kunal Talwar

HIGHLIGHT: Here we demonstrate a general approach that, under standard cryptographic assumptions, compresses every efficient LDP algorithm with negligible loss in privacy and utility guarantees.

295, TITLE: Dimensionality Reduction for the Sum-of-Distances Metric
http://proceedings.mlr.press/v139/feng21a.html
AUTHORS: Zhili Feng, Praneeth Kacham, David Woodruff
HIGHLIGHT: We give a dimensionality reduction procedure to approximate the sum of distances of a given set of n points in Rd to any "shape" that lies in a k-dimensional subspace.

296, TITLE: Reserve Price Optimization for First Price Auctions in Display Advertising
http://proceedings.mlr.press/v139/feng21b.html
AUTHORS: Zhe Feng, Sebastien Lahaie, Jon Schneider, Jinchao Ye
HIGHLIGHT: In this paper, we propose a gradient-based algorithm to adaptively update and optimize reserve prices based on estimates of bidders' responsiveness to experimental shocks in reserves.

297, TITLE: Uncertainty Principles of Encoding GANs
http://proceedings.mlr.press/v139/feng21c.html
AUTHORS: Ruili Feng, Zhouchen Lin, Jiapeng Zhu, Deli Zhao, Jingren Zhou, Zheng-Jun Zha
HIGHLIGHT: In this paper we study this predicament of encoding GANs, which is indispensable research for the GAN community.

298, TITLE: Pointwise Binary Classification with Pairwise Confidence Comparisons
http://proceedings.mlr.press/v139/feng21d.html
AUTHORS: Lei Feng, Senlin Shu, Nan Lu, Bo Han, Miao Xu, Gang Niu, Bo An, Masashi Sugiyama
HIGHLIGHT: Thus, in this paper, we propose a novel setting called pairwise comparison (Pcomp) classification, where we have only pairs of unlabeled data that we know one is more likely to be positive than the other.

299, TITLE: Provably Correct Optimization and Exploration with Non-linear Policies
http://proceedings.mlr.press/v139/feng21e.html
AUTHORS: Fei Feng, Wotao Yin, Alekh Agarwal, Lin Yang
HIGHLIGHT: In this paper, we address this question by designing ENIAC, an actor-critic method that allows non-linear function approximation in the critic.

300, TITLE: KD3A: Unsupervised Multi-Source Decentralized Domain Adaptation via Knowledge Distillation
http://proceedings.mlr.press/v139/feng21f.html
AUTHORS: Haozhe Feng, Zhaoyang You, Minghao Chen, Tianye Zhang, Minfeng Zhu, Fei Wu, Chao Wu, Wei Chen
HIGHLIGHT: To address the above problems, we propose a privacy-preserving UMDA paradigm named Knowledge Distillation based Decentralized Domain Adaptation (KD3A), which performs domain adaptation through the knowledge distillation on models from different source domains.

301, TITLE: Understanding Noise Injection in GANs
http://proceedings.mlr.press/v139/feng21g.html
AUTHORS: Ruili Feng, Deli Zhao, Zheng-Jun Zha
HIGHLIGHT: In this paper, we propose a geometric framework to theoretically analyze the role of noise injection in GANs.

302, TITLE: GNNAutoScale: Scalable and Expressive Graph Neural Networks via Historical Embeddings
http://proceedings.mlr.press/v139/fey21a.html
AUTHORS: Matthias Fey, Jan E. Lenssen, Frank Weichert, Jure Leskovec
HIGHLIGHT: We present GNNAutoScale (GAS), a framework for scaling arbitrary message-passing GNNs to large graphs.

303, TITLE: PsiPhi-Learning: Reinforcement Learning with Demonstrations using Successor Features and Inverse Temporal Difference Learning
http://proceedings.mlr.press/v139/filos21a.html
AUTHORS: Angelos Filos, Clare Lyle, Yarin Gal, Sergey Levine, Natasha Jaques, Gregory Farquhar
HIGHLIGHT: We propose a multi-task inverse reinforcement learning (IRL) algorithm, called \emph{inverse temporal difference learning} (ITD), that learns shared state features, alongside per-agent successor features and preference vectors, purely from demonstrations without reward labels.

304, TITLE: A Practical Method for Constructing Equivariant Multilayer Perceptrons for Arbitrary Matrix Groups
http://proceedings.mlr.press/v139/finzi21a.html

AUTHORS: Marc Finzi, Max Welling, Andrew Gordon Wilson
HIGHLIGHT: In this work we provide a completely general algorithm for solving for the equivariant layers of matrix groups.

305, TITLE: Few-Shot Conformal Prediction with Auxiliary Tasks
http://proceedings.mlr.press/v139/fisch21a.html
AUTHORS: Adam Fisch, Tal Schuster, Tommi Jaakkola, Dr.Regina Barzilay
HIGHLIGHT: In this work, we obtain substantially tighter prediction sets while maintaining desirable marginal guarantees by casting conformal prediction as a meta-learning paradigm over exchangeable collections of auxiliary tasks.

306, TITLE: Scalable Certified Segmentation via Randomized Smoothing
http://proceedings.mlr.press/v139/fischer21a.html
AUTHORS: Marc Fischer, Maximilian Baader, Martin Vechev
HIGHLIGHT: We present a new certification method for image and point cloud segmentation based on randomized smoothing.

307, TITLE: What?s in the Box? Exploring the Inner Life of Neural Networks with Robust Rules
http://proceedings.mlr.press/v139/fischer21b.html
AUTHORS: Jonas Fischer, Anna Olah, Jilles Vreeken
HIGHLIGHT: We propose a novel method for exploring how neurons within neural networks interact.

308, TITLE: Online Learning with Optimism and Delay
http://proceedings.mlr.press/v139/flaspohler21a.html
AUTHORS: Genevieve E Flaspohler, Francesco Orabona, Judah Cohen, Soukayna Mouatadid, Miruna Oprescu, Paulo Orenstein, Lester Mackey
HIGHLIGHT: Inspired by the demands of real-time climate and weather forecasting, we develop optimistic online learning algorithms that require no parameter tuning and have optimal regret guarantees under delayed feedback.

309, TITLE: Online A-Optimal Design and Active Linear Regression
http://proceedings.mlr.press/v139/fontaine21a.html
AUTHORS: Xavier Fontaine, Pierre Perrault, Michal Valko, Vianney Perchet
HIGHLIGHT: We consider in this paper the problem of optimal experiment design where a decision maker can choose which points to sample to obtain an estimate $\hat{\beta}$ of the hidden parameter $\beta^{\star}$ of an underlying linear model.

310, TITLE: Deep Adaptive Design: Amortizing Sequential Bayesian Experimental Design
http://proceedings.mlr.press/v139/foster21a.html
AUTHORS: Adam Foster, Desi R Ivanova, Ilyas Malik, Tom Rainforth
HIGHLIGHT: We introduce Deep Adaptive Design (DAD), a method for amortizing the cost of adaptive Bayesian experimental design that allows experiments to be run in real-time.

311, TITLE: Efficient Online Learning for Dynamic k-Clustering
http://proceedings.mlr.press/v139/fotakis21a.html
AUTHORS: Dimitris Fotakis, Georgios Piliouras, Stratis Skoulakis
HIGHLIGHT: In this work, we study dynamic clustering problems from the perspective of online learning.

312, TITLE: Clustered Sampling: Low-Variance and Improved Representativity for Clients Selection in Federated Learning
http://proceedings.mlr.press/v139/fraboni21a.html
AUTHORS: Yann Fraboni, Richard Vidal, Laetitia Kameni, Marco Lorenzi
HIGHLIGHT: This work addresses the problem of optimizing communications between server and clients in federated learning (FL).

313, TITLE: Agnostic Learning of Halfspaces with Gradient Descent via Soft Margins
http://proceedings.mlr.press/v139/frei21a.html
AUTHORS: Spencer Frei, Yuan Cao, Quanquan Gu
HIGHLIGHT: We show that when a quantity we refer to as the \textit{soft margin} is well-behaved—a condition satisfied by log-concave isotropic distributions among others—minimizers of convex surrogates for the zero-one loss are approximate minimizers for the zero-one loss itself.

314, TITLE: Provable Generalization of SGD-trained Neural Networks of Any Width in the Presence of Adversarial Label Noise
http://proceedings.mlr.press/v139/frei21b.html

AUTHORS: Spencer Frei, Yuan Cao, Quanquan Gu
HIGHLIGHT: To the best of our knowledge, this is the first work to show that overparameterized neural networks trained by SGD can generalize when the data is corrupted with adversarial label noise.

315, TITLE: Post-selection inference with HSIC-Lasso
http://proceedings.mlr.press/v139/freidling21a.html
AUTHORS: Tobias Freidling, Benjamin Poignard, H?ctor Climente-Gonz?lez, Makoto Yamada
HIGHLIGHT: We propose a selective inference procedure using the so-called model-free "HSIC-Lasso" based on the framework of truncated Gaussians combined with the polyhedral lemma.

316, TITLE: Variational Data Assimilation with a Learned Inverse Observation Operator
http://proceedings.mlr.press/v139/frerix21a.html
AUTHORS: Thomas Frerix, Dmitrii Kochkov, Jamie Smith, Daniel Cremers, Michael Brenner, Stephan Hoyer
HIGHLIGHT: We learn a mapping from observational data to physical states and show how it can be used to improve optimizability.

317, TITLE: Bayesian Quadrature on Riemannian Data Manifolds
http://proceedings.mlr.press/v139/frohlich21a.html
AUTHORS: Christian Fr?hlich, Alexandra Gessner, Philipp Hennig, Bernhard Sch?lkopf, Georgios Arvanitidis
HIGHLIGHT: To ease this computational burden, we advocate probabilistic numerical methods for Riemannian statistics.

318, TITLE: Learn-to-Share: A Hardware-friendly Transfer Learning Framework Exploiting Computation and Parameter Sharing
http://proceedings.mlr.press/v139/fu21a.html
AUTHORS: Cheng Fu, Hanxian Huang, Xinyun Chen, Yuandong Tian, Jishen Zhao
HIGHLIGHT: In this work, we propose LeTS, a framework that leverages both computation and parameter sharing across multiple tasks.

319, TITLE: Learning Task Informed Abstractions
http://proceedings.mlr.press/v139/fu21b.html
AUTHORS: Xiang Fu, Ge Yang, Pulkit Agrawal, Tommi Jaakkola
HIGHLIGHT: To mitigate this problem, we propose learning Task Informed Abstractions (TIA) that explicitly separates reward-correlated visual features from distractors.

320, TITLE: Double-Win Quant: Aggressively Winning Robustness of Quantized Deep Neural Networks via Random Precision Training and Inference
http://proceedings.mlr.press/v139/fu21c.html
AUTHORS: Yonggan Fu, Qixuan Yu, Meng Li, Vikas Chandra, Yingyan Lin
HIGHLIGHT: In this work, we demonstrate a new perspective regarding quantization’s role in DNNs’ robustness, advocating that quantization can be leveraged to largely boost DNNs’ robustness, and propose a framework dubbed Double-Win Quant that can boost the robustness of quantized DNNs over their full precision counterparts by a large margin.

321, TITLE: Auto-NBA: Efficient and Effective Search Over the Joint Space of Networks, Bitwidths, and Accelerators
http://proceedings.mlr.press/v139/fu21d.html
AUTHORS: Yonggan Fu, Yongan Zhang, Yang Zhang, David Cox, Yingyan Lin
HIGHLIGHT: To tackle these daunting challenges towards optimal and fast development of DNN accelerators, we propose a framework dubbed Auto-NBA to enable jointly searching for the Networks, Bitwidths, and Accelerators, by efficiently localizing the optimal design within the huge joint design space for each target dataset and acceleration specification.

322, TITLE: A Deep Reinforcement Learning Approach to Marginalized Importance Sampling with the Successor Representation
http://proceedings.mlr.press/v139/fujimoto21a.html
AUTHORS: Scott Fujimoto, David Meger, Doina Precup
HIGHLIGHT: We bridge the gap between MIS and deep reinforcement learning by observing that the density ratio can be computed from the successor representation of the target policy.

323, TITLE: Learning disentangled representations via product manifold projection
http://proceedings.mlr.press/v139/fumero21a.html
AUTHORS: Marco Fumero, Luca Cosmo, Simone Melzi, Emanuele Rodola
HIGHLIGHT: We propose a novel approach to disentangle the generative factors of variation underlying a given set of observations.

324, TITLE: Policy Information Capacity: Information-Theoretic Measure for Task Complexity in Deep Reinforcement Learning
http://proceedings.mlr.press/v139/furuta21a.html
AUTHORS: Hiroki Furuta, Tatsuya Matsushima, Tadashi Kozuno, Yutaka Matsuo, Sergey Levine, Ofir Nachum, Shixiang Shane Gu
HIGHLIGHT: In this work, we propose policy information capacity (PIC) – the mutual information between policy parameters and episodic return – and policy-optimal information capacity (POIC) – between policy parameters and episodic optimality – as two environment-agnostic, algorithm-agnostic quantitative metrics for task difficulty.

325, TITLE: An Information-Geometric Distance on the Space of Tasks
http://proceedings.mlr.press/v139/gao21a.html
AUTHORS: Yansong Gao, Pratik Chaudhari
HIGHLIGHT: We develop an algorithm to compute the distance which iteratively transports the marginal on the data of the source task to that of the target task while updating the weights of the classifier to track this evolving data distribution.

326, TITLE: Maximum Mean Discrepancy Test is Aware of Adversarial Attacks
http://proceedings.mlr.press/v139/gao21b.html
AUTHORS: Ruize Gao, Feng Liu, Jingfeng Zhang, Bo Han, Tongliang Liu, Gang Niu, Masashi Sugiyama
HIGHLIGHT: Given this phenomenon, we raise a question: are natural and adversarial data really from different distributions? The answer is affirmative- the previous use of the MMD test on the purpose missed three key factors, and accordingly, we propose three components.

327, TITLE: Unsupervised Co-part Segmentation through Assembly
http://proceedings.mlr.press/v139/gao21c.html
AUTHORS: Qingzhe Gao, Bin Wang, Libin Liu, Baoquan Chen
HIGHLIGHT: We propose an unsupervised learning approach for co-part segmentation from images.

328, TITLE: Discriminative Complementary-Label Learning with Weighted Loss
http://proceedings.mlr.press/v139/gao21d.html
AUTHORS: Yi Gao, Min-Ling Zhang
HIGHLIGHT: In this paper, we derive a simple and theoretically-sound \emph{discriminative} model towards $P(\bar y\mid {\bm x})$, which naturally leads to a risk estimator with estimation error bound at $\mathcal{O}(1/\sqrt{n})$ convergence rate.

329, TITLE: RATT: Leveraging Unlabeled Data to Guarantee Generalization
http://proceedings.mlr.press/v139/garg21a.html
AUTHORS: Saurabh Garg, Sivaraman Balakrishnan, Zico Kolter, Zachary Lipton
HIGHLIGHT: In this paper, we leverage unlabeled data to produce generalization bounds.

330, TITLE: On Proximal Policy Optimization?s Heavy-tailed Gradients
http://proceedings.mlr.press/v139/garg21b.html
AUTHORS: Saurabh Garg, Joshua Zhanson, Emilio Parisotto, Adarsh Prasad, Zico Kolter, Zachary Lipton, Sivaraman Balakrishnan, Ruslan Salakhutdinov, Pradeep Ravikumar
HIGHLIGHT: In this paper, we present a detailed empirical study to characterize the heavy-tailed nature of the gradients of the PPO surrogate reward function.

331, TITLE: What does LIME really see in images?
http://proceedings.mlr.press/v139/garreau21a.html
AUTHORS: Damien Garreau, Dina Mardaoui
HIGHLIGHT: On the theoretical side, we show that when the number of generated examples is large, LIME explanations are concentrated around a limit explanation for which we give an explicit expression.

332, TITLE: Parametric Graph for Unimodal Ranking Bandit
http://proceedings.mlr.press/v139/gauthier21a.html
AUTHORS: Camille-Sovanneary Gauthier, Romaric Gaudel, Elisa Fromont, Boammani Aser Lompo
HIGHLIGHT: We propose an original algorithm, easy to implement and with strong theoretical guarantees to tackle this problem in the Position-Based Model (PBM) setting, well suited for applications where items are displayed on a grid.

333, TITLE: Let?s Agree to Degree: Comparing Graph Convolutional Networks in the Message-Passing Framework
http://proceedings.mlr.press/v139/geerts21a.html

AUTHORS: Floris Geerts, Filip Mazowiecki, Guillermo Perez
HIGHLIGHT: In this paper we cast neural networks defined on graphs as message-passing neural networks (MPNNs) to study the distinguishing power of different classes of such models.

334, TITLE: On the difficulty of unbiased alpha divergence minimization
http://proceedings.mlr.press/v139/geffner21a.html
AUTHORS: Tomas Geffner, Justin Domke
HIGHLIGHT: In this work we study unbiased methods for alpha-divergence minimization through the Signal-to-Noise Ratio (SNR) of the gradient estimator.

335, TITLE: How and Why to Use Experimental Data to Evaluate Methods for Observational Causal Inference
http://proceedings.mlr.press/v139/gentzel21a.html
AUTHORS: Amanda M Gentzel, Purva Pruthi, David Jensen
HIGHLIGHT: We describe and analyze observational sampling from randomized controlled trials (OSRCT), a method for evaluating causal inference methods using data from randomized controlled trials (RCTs).

336, TITLE: Strategic Classification in the Dark
http://proceedings.mlr.press/v139/ghalme21a.html
AUTHORS: Ganesh Ghalme, Vineet Nair, Itay Eilat, Inbal Talgam-Cohen, Nir Rosenfeld
HIGHLIGHT: In this paper we generalize the strategic classification model to such scenarios and analyze the effect of an unknown classifier.

337, TITLE: EMaQ: Expected-Max Q-Learning Operator for Simple Yet Effective Offline and Online RL
http://proceedings.mlr.press/v139/ghasemipour21a.html
AUTHORS: Seyed Kamyar Seyed Ghasemipour, Dale Schuurmans, Shixiang Shane Gu
HIGHLIGHT: In this work, we closely investigate an important simplification of BCQ (Fujimoto et al., 2018) - a prior approach for offline RL - removing a heuristic design choice.

338, TITLE: Differentially Private Aggregation in the Shuffle Model: Almost Central Accuracy in Almost a Single Message
http://proceedings.mlr.press/v139/ghazi21a.html
AUTHORS: Badih Ghazi, Ravi Kumar, Pasin Manurangsi, Rasmus Pagh, Amer Sinha
HIGHLIGHT: In this work, we study the problem of summing (aggregating) real numbers or integers, a basic primitive in numerous machine learning tasks, in the shuffle model.

339, TITLE: The Power of Adaptivity for Stochastic Submodular Cover
http://proceedings.mlr.press/v139/ghuge21a.html
AUTHORS: Rohan Ghuge, Anupam Gupta, Viswanath Nagarajan
HIGHLIGHT: We ask: how well can solutions with only a few adaptive rounds approximate fully-adaptive solutions?

340, TITLE: Differentially Private Quantiles
http://proceedings.mlr.press/v139/gillenwater21a.html
AUTHORS: Jennifer Gillenwater, Matthew Joseph, Alex Kulesza
HIGHLIGHT: In this work we propose an instance of the exponential mechanism that simultaneously estimates exactly m quantiles from n data points while guaranteeing differential privacy.

341, TITLE: Query Complexity of Adversarial Attacks
http://proceedings.mlr.press/v139/gluch21a.html
AUTHORS: Grzegorz Gluch, R?diger Urbanke
HIGHLIGHT: There are two main attack models considered in the adversarial robustness literature: black-box and white-box. We consider these threat models as two ends of a fine-grained spectrum, indexed by the number of queries the adversary can ask.

342, TITLE: Spectral Normalisation for Deep Reinforcement Learning: An Optimisation Perspective
http://proceedings.mlr.press/v139/gogianu21a.html
AUTHORS: Florin Gogianu, Tudor Berariu, Mihaela C Rosca, Claudia Clopath, Lucian Busoniu, Razvan Pascanu
HIGHLIGHT: We diverge from this view and show we can recover the performance of these developments not by changing the objective, but by regularising the value-function estimator.

343, TITLE: 12-Lead ECG Reconstruction via Koopman Operators
http://proceedings.mlr.press/v139/golany21a.html
AUTHORS: Tomer Golany, Kira Radinsky, Daniel Freedman, Saar Minha

HIGHLIGHT: In this work, we present a methodology to reconstruct missing or noisy leads using the theory of Koopman Operators.

344, TITLE: Function Contrastive Learning of Transferable Meta-Representations
http://proceedings.mlr.press/v139/gondal21a.html
AUTHORS: Muhammad Waleed Gondal, Shruti Joshi, Nasim Rahaman, Stefan Bauer, Manuel Wuthrich, Bernhard Sch?lkopf
HIGHLIGHT: In this work, we study the implications of this joint training on the transferability of the meta-representations.

345, TITLE: Active Slices for Sliced Stein Discrepancy
http://proceedings.mlr.press/v139/gong21a.html
AUTHORS: Wenbo Gong, Kaibo Zhang, Yingzhen Li, Jose Miguel Hernandez-Lobato
HIGHLIGHT: First, we show in theory that the requirement of using optimal slicing directions in the kernelized version of SSD can be relaxed, validating the resulting discrepancy with finite random slicing directions. Second, given that good slicing directions are crucial for practical performance, we propose a fast algorithm for finding good slicing directions based on ideas of active sub-space construction and spectral decomposition.

346, TITLE: On the Problem of Underranking in Group-Fair Ranking
http://proceedings.mlr.press/v139/gorantla21a.html
AUTHORS: Sruthi Gorantla, Amit Deshpande, Anand Louis
HIGHLIGHT: In this paper, we formulate the problem of underranking in group-fair rankings based on how close the group-fair rank of each item is to its original rank, and prove a lower bound on the trade-off achievable for simultaneous underranking and group fairness in ranking.

347, TITLE: MARINA: Faster Non-Convex Distributed Learning with Compression
http://proceedings.mlr.press/v139/gorbunov21a.html
AUTHORS: Eduard Gorbunov, Konstantin P. Burlachenko, Zhize Li, Peter Richtarik
HIGHLIGHT: We develop and analyze MARINA: a new communication efficient method for non-convex distributed learning over heterogeneous datasets.

348, TITLE: Systematic Analysis of Cluster Similarity Indices: How to Validate Validation Measures
http://proceedings.mlr.press/v139/gosgens21a.html
AUTHORS: Martijn M G?sgens, Alexey Tikhonov, Liudmila Prokhorenkova
HIGHLIGHT: We propose a theoretical framework to tackle this problem: we develop a list of desirable properties and conduct an extensive theoretical analysis to verify which indices satisfy them.

349, TITLE: Revisiting Point Cloud Shape Classification with a Simple and Effective Baseline
http://proceedings.mlr.press/v139/goyal21a.html
AUTHORS: Ankit Goyal, Hei Law, Bowei Liu, Alejandro Newell, Jia Deng
HIGHLIGHT: First, we find that auxiliary factors like different evaluation schemes, data augmentation strategies, and loss functions, which are independent of the model architecture, make a large difference in performance.

350, TITLE: Dissecting Supervised Constrastive Learning
http://proceedings.mlr.press/v139/graf21a.html
AUTHORS: Florian Graf, Christoph Hofer, Marc Niethammer, Roland Kwitt
HIGHLIGHT: In this work, we address the question whether there are fundamental differences in the sought-for representation geometry in the output space of the encoder at minimal loss.

351, TITLE: Oops I Took A Gradient: Scalable Sampling for Discrete Distributions
http://proceedings.mlr.press/v139/grathwohl21a.html
AUTHORS: Will Grathwohl, Kevin Swersky, Milad Hashemi, David Duvenaud, Chris Maddison
HIGHLIGHT: We propose a general and scalable approximate sampling strategy for probabilistic models with discrete variables.

352, TITLE: Detecting Rewards Deterioration in Episodic Reinforcement Learning
http://proceedings.mlr.press/v139/greenberg21a.html
AUTHORS: Ido Greenberg, Shie Mannor
HIGHLIGHT: In this paper, we address this problem by focusing directly on the rewards and testing for degradation. We present this problem as a multivariate mean-shift detection problem with possibly partial observations.

353, TITLE: Crystallization Learning with the Delaunay Triangulation
http://proceedings.mlr.press/v139/gu21a.html
AUTHORS: Jiaqi Gu, Guosheng Yin
HIGHLIGHT: Based on the Delaunay triangulation, we propose the crystallization learning to estimate the conditional expectation function in the framework of nonparametric regression.

354, TITLE: AutoAttend: Automated Attention Representation Search
http://proceedings.mlr.press/v139/guan21a.html
AUTHORS: Chaoyu Guan, Xin Wang, Wenwu Zhu
HIGHLIGHT: In this paper, we automate Key, Query and Value representation design, which is one of the most important steps to obtain effective self-attentions.

355, TITLE: Operationalizing Complex Causes: A Pragmatic View of Mediation
http://proceedings.mlr.press/v139/gultchin21a.html
AUTHORS: Limor Gultchin, David Watson, Matt Kusner, Ricardo Silva
HIGHLIGHT: Given a collection of candidate mediators, we propose (a) a two-step method for predicting the causal responses of crude interventions; and (b) a testing procedure to identify mediators of crude interventions.

356, TITLE: On a Combination of Alternating Minimization and Nesterov?s Momentum
http://proceedings.mlr.press/v139/guminov21a.html
AUTHORS: Sergey Guminov, Pavel Dvurechensky, Nazarii Tupitsa, Alexander Gasnikov
HIGHLIGHT: In this paper we combine AM and Nesterov’s acceleration to propose an accelerated alternating minimization algorithm.

357, TITLE: Decentralized Single-Timescale Actor-Critic on Zero-Sum Two-Player Stochastic Games
http://proceedings.mlr.press/v139/guo21a.html
AUTHORS: Hongyi Guo, Zuyue Fu, Zhuoran Yang, Zhaoran Wang
HIGHLIGHT: We study the global convergence and global optimality of the actor-critic algorithm applied for the zero-sum two-player stochastic games in a decentralized manner.

358, TITLE: Adversarial Policy Learning in Two-player Competitive Games
http://proceedings.mlr.press/v139/guo21b.html
AUTHORS: Wenbo Guo, Xian Wu, Sui Huang, Xinyu Xing
HIGHLIGHT: In this work, we propose a new adversarial learning algorithm.

359, TITLE: Soft then Hard: Rethinking the Quantization in Neural Image Compression
http://proceedings.mlr.press/v139/guo21c.html
AUTHORS: Zongyu Guo, Zhizheng Zhang, Runsen Feng, Zhibo Chen
HIGHLIGHT: We thus propose a novel soft-then-hard quantization strategy for neural image compression that first learns an expressive latent space softly, then closes the train-test mismatch with hard quantization.

360, TITLE: UneVEn: Universal Value Exploration for Multi-Agent Reinforcement Learning
http://proceedings.mlr.press/v139/gupta21a.html
AUTHORS: Tarun Gupta, Anuj Mahajan, Bei Peng, Wendelin Boehmer, Shimon Whiteson
HIGHLIGHT: Specifically, we propose a novel MARL approach called Universal Value Exploration (UneVEn) that learns a set of related tasks simultaneously with a linear decomposition of universal successor features.

361, TITLE: Distribution-Free Calibration Guarantees for Histogram Binning without Sample Splitting
http://proceedings.mlr.press/v139/gupta21b.html
AUTHORS: Chirag Gupta, Aaditya Ramdas
HIGHLIGHT: We prove calibration guarantees for the popular histogram binning (also called uniform-mass binning) method of Zadrozny and Elkan (2001).

362, TITLE: Correcting Exposure Bias for Link Recommendation
http://proceedings.mlr.press/v139/gupta21c.html
AUTHORS: Shantanu Gupta, Hao Wang, Zachary Lipton, Yuyang Wang
HIGHLIGHT: We propose estimators that leverage known exposure probabilities to mitigate this bias and consequent feedback loops.

363, TITLE: The Heavy-Tail Phenomenon in SGD

http://proceedings.mlr.press/v139/gurbuzbalaban21a.html
AUTHORS: Mert Gurbuzbalaban, Umut Simsekli, Lingjiong Zhu
HIGHLIGHT: In this paper, we argue that these three seemingly unrelated perspectives for generalization are deeply linked to each other.

364, TITLE: Knowledge Enhanced Machine Learning Pipeline against Diverse Adversarial Attacks
http://proceedings.mlr.press/v139/gurel21a.html
AUTHORS: Nezihe Merve G?rel, Xiangyu Qi, Luka Rimanic, Ce Zhang, Bo Li
HIGHLIGHT: In this work, we aim to enhance the ML robustness from a different perspective by leveraging domain knowledge: We propose a Knowledge Enhanced Machine Learning Pipeline (KEMLP) to integrate domain knowledge (i.e., logic relationships among different predictions) into a probabilistic graphical model via first-order logic rules.

365, TITLE: Adapting to Delays and Data in Adversarial Multi-Armed Bandits
http://proceedings.mlr.press/v139/gyorgy21a.html
AUTHORS: Andras Gyorgy, Pooria Joulani
HIGHLIGHT: We consider the adversarial multi-armed bandit problem under delayed feedback.

366, TITLE: Rate-Distortion Analysis of Minimum Excess Risk in Bayesian Learning
http://proceedings.mlr.press/v139/hafez-kolahi21a.html
AUTHORS: Hassan Hafez-Kolahi, Behrad Moniri, Shohreh Kasaei, Mahdieh Soleymani Baghshah
HIGHLIGHT: In this paper, we build upon and extend the recent results of (Xu & Raginsky, 2020) to analyze the MER in Bayesian learning and derive information-theoretic bounds on it.

367, TITLE: Regret Minimization in Stochastic Non-Convex Learning via a Proximal-Gradient Approach
http://proceedings.mlr.press/v139/hallak21a.html
AUTHORS: Nadav Hallak, Panayotis Mertikopoulos, Volkan Cevher
HIGHLIGHT: On that account, we propose a conceptual approach that leverages non-convex optimality measures, leading to a suitable generalization of the learner’s local regret.

368, TITLE: Diversity Actor-Critic: Sample-Aware Entropy Regularization for Sample-Efficient Exploration
http://proceedings.mlr.press/v139/han21a.html
AUTHORS: Seungyul Han, Youngchul Sung
HIGHLIGHT: In this paper, sample-aware policy entropy regularization is proposed to enhance the conventional policy entropy regularization for better exploration.

369, TITLE: Adversarial Combinatorial Bandits with General Non-linear Reward Functions
http://proceedings.mlr.press/v139/han21b.html
AUTHORS: Yanjun Han, Yining Wang, Xi Chen
HIGHLIGHT: In this paper we study the adversarial combinatorial bandit with a known non-linear reward function, extending existing work on adversarial linear combinatorial bandit.

370, TITLE: A Collective Learning Framework to Boost GNN Expressiveness for Node Classification
http://proceedings.mlr.press/v139/hang21a.html
AUTHORS: Mengyue Hang, Jennifer Neville, Bruno Ribeiro
HIGHLIGHT: In this work, we investigate this question and propose {\em collective learning} for GNNs —a general collective classification approach for node representation learning that increases their representation power.

371, TITLE: Grounding Language to Entities and Dynamics for Generalization in Reinforcement Learning
http://proceedings.mlr.press/v139/hanjie21a.html
AUTHORS: Austin W. Hanjie, Victor Y Zhong, Karthik Narasimhan
HIGHLIGHT: We develop a new model, EMMA (Entity Mapper with Multi-modal Attention) which uses an entity-conditioned attention module that allows for selective focus over relevant descriptions in the manual for each entity in the environment.

372, TITLE: Sparse Feature Selection Makes Batch Reinforcement Learning More Sample Efficient
http://proceedings.mlr.press/v139/hao21a.html
AUTHORS: Botao Hao, Yaqi Duan, Tor Lattimore, Csaba Szepesvari, Mengdi Wang
HIGHLIGHT: This paper provides a statistical analysis of high-dimensional batch reinforcement learning (RL) using sparse linear function approximation.

373, TITLE: Bootstrapping Fitted Q-Evaluation for Off-Policy Inference
http://proceedings.mlr.press/v139/hao21b.html
AUTHORS: Botao Hao, Xiang Ji, Yaqi Duan, Hao Lu, Csaba Szepesvari, Mengdi Wang
HIGHLIGHT: In this paper, we study the use of bootstrapping in off-policy evaluation (OPE), and in particular, we focus on the fitted Q-evaluation (FQE) that is known to be minimax-optimal in the tabular and linear-model cases.

374, TITLE: Compressed Maximum Likelihood
http://proceedings.mlr.press/v139/hao21c.html
AUTHORS: Yi Hao, Alon Orlitsky
HIGHLIGHT: Inspired by recent advances in estimating distribution functionals, we propose $\textit{compressed maximum likelihood}$ (CML) that applies ML to the compressed samples.

375, TITLE: Valid Causal Inference with (Some) Invalid Instruments
http://proceedings.mlr.press/v139/hartford21a.html
AUTHORS: Jason S Hartford, Victor Veitch, Dhanya Sridhar, Kevin Leyton-Brown
HIGHLIGHT: In this paper, we show how to perform consistent IV estimation despite violations of the exclusion assumption.

376, TITLE: Model Performance Scaling with Multiple Data Sources
http://proceedings.mlr.press/v139/hashimoto21a.html
AUTHORS: Tatsunori Hashimoto
HIGHLIGHT: We show that there is a simple scaling law that predicts the loss incurred by a model even under varying dataset composition.

377, TITLE: Hierarchical VAEs Know What They Don't Know
http://proceedings.mlr.press/v139/havtorn21a.html
AUTHORS: Jakob D. Drachmann Havtorn, Jes Frellsen, Soren Hauberg, Lars Maal?e
HIGHLIGHT: In the context of hierarchical variational autoencoders, we provide evidence to explain this behavior by out-of-distribution data having in-distribution low-level features.

378, TITLE: Defense against backdoor attacks via robust covariance estimation
http://proceedings.mlr.press/v139/hayase21a.html
AUTHORS: Jonathan Hayase, Weihao Kong, Raghav Somani, Sewoong Oh
HIGHLIGHT: We propose a novel defense algorithm using robust covariance estimation to amplify the spectral signature of corrupted data.

379, TITLE: Boosting for Online Convex Optimization
http://proceedings.mlr.press/v139/hazan21a.html
AUTHORS: Elad Hazan, Karan Singh
HIGHLIGHT: We consider the decision-making framework of online convex optimization with a very large number of experts.

380, TITLE: PipeTransformer: Automated Elastic Pipelining for Distributed Training of Large-scale Models
http://proceedings.mlr.press/v139/he21a.html
AUTHORS: Chaoyang He, Shen Li, Mahdi Soltanolkotabi, Salman Avestimehr
HIGHLIGHT: In this paper, we propose PipeTransformer, which leverages automated elastic pipelining for efficient distributed training of Transformer models.

381, TITLE: SoundDet: Polyphonic Moving Sound Event Detection and Localization from Raw Waveform
http://proceedings.mlr.press/v139/he21b.html
AUTHORS: Yuhang He, Niki Trigoni, Andrew Markham
HIGHLIGHT: We present a new framework SoundDet, which is an end-to-end trainable and light-weight framework, for polyphonic moving sound event detection and localization.

382, TITLE: Logarithmic Regret for Reinforcement Learning with Linear Function Approximation
http://proceedings.mlr.press/v139/he21c.html
AUTHORS: Jiafan He, Dongruo Zhou, Quanquan Gu
HIGHLIGHT: In this paper, we show that logarithmic regret is attainable under two recently proposed linear MDP assumptions provided that there exists a positive sub-optimality gap for the optimal action-value function.

383, TITLE: Finding Relevant Information via a Discrete Fourier Expansion

http://proceedings.mlr.press/v139/heidari21a.html
AUTHORS: Mohsen Heidari, Jithin Sreedharan, Gil I Shamir, Wojciech Szpankowski
HIGHLIGHT: To address this, we propose a Fourier-based approach to extract relevant information in the supervised setting.

384, TITLE: Zeroth-Order Non-Convex Learning via Hierarchical Dual Averaging
http://proceedings.mlr.press/v139/heliou21a.html
AUTHORS: Am?lie H?liou, Matthieu Martin, Panayotis Mertikopoulos, Thibaud Rahier
HIGHLIGHT: We propose a hierarchical version of dual averaging for zeroth-order online non-convex optimization {–} i.e., learning processes where, at each stage, the optimizer is facing an unknown non-convex loss function and only receives the incurred loss as feedback.

385, TITLE: Improving Molecular Graph Neural Network Explainability with Orthonormalization and Induced Sparsity
http://proceedings.mlr.press/v139/henderson21a.html
AUTHORS: Ryan Henderson, Djork-Arn? Clevert, Floriane Montanari
HIGHLIGHT: To help, we propose two simple regularization techniques to apply during the training of GCNNs: Batch Representation Orthonormalization (BRO) and Gini regularization.

386, TITLE: Muesli: Combining Improvements in Policy Optimization
http://proceedings.mlr.press/v139/hessel21a.html
AUTHORS: Matteo Hessel, Ivo Danihelka, Fabio Viola, Arthur Guez, Simon Schmitt, Laurent Sifre, Theophane Weber, David Silver, Hado Van Hasselt
HIGHLIGHT: We propose a novel policy update that combines regularized policy optimization with model learning as an auxiliary loss.

387, TITLE: Learning Representations by Humans, for Humans
http://proceedings.mlr.press/v139/hilgard21a.html
AUTHORS: Sophie Hilgard, Nir Rosenfeld, Mahzarin R Banaji, Jack Cao, David Parkes
HIGHLIGHT: Here we propose a framework to directly support human decision-making, in which the role of machines is to reframe problems rather than to prescribe actions through prediction.

388, TITLE: Optimizing Black-box Metrics with Iterative Example Weighting
http://proceedings.mlr.press/v139/hiranandani21a.html
AUTHORS: Gaurush Hiranandani, Jatin Mathur, Harikrishna Narasimhan, Mahdi Milani Fard, Sanmi Koyejo
HIGHLIGHT: Our approach is to adaptively learn example weights on the training dataset such that the resulting weighted objective best approximates the metric on the validation sample.

389, TITLE: Trees with Attention for Set Prediction Tasks
http://proceedings.mlr.press/v139/hirsch21a.html
AUTHORS: Roy Hirsch, Ran Gilad-Bachrach
HIGHLIGHT: Set-Tree, presented in this work, extends the support for sets to tree-based models, such as Random-Forest and Gradient-Boosting, by introducing an attention mechanism and set-compatible split criteria.

390, TITLE: Multiplicative Noise and Heavy Tails in Stochastic Optimization
http://proceedings.mlr.press/v139/hodgkinson21a.html
AUTHORS: Liam Hodgkinson, Michael Mahoney
HIGHLIGHT: Modeling stochastic optimization algorithms as discrete random recurrence relations, we show that multiplicative noise, as it commonly arises due to variance in local rates of convergence, results in heavy-tailed stationary behaviour in the parameters.

391, TITLE: MC-LSTM: Mass-Conserving LSTM
http://proceedings.mlr.press/v139/hoedt21a.html
AUTHORS: Pieter-Jan Hoedt, Frederik Kratzert, Daniel Klotz, Christina Halmich, Markus Holzleitner, Grey S Nearing, Sepp Hochreiter, Guenter Klambauer
HIGHLIGHT: Our novel Mass-Conserving LSTM (MC-LSTM) adheres to these conservation laws by extending the inductive bias of LSTM to model the redistribution of those stored quantities.

392, TITLE: Learning Curves for Analysis of Deep Networks
http://proceedings.mlr.press/v139/hoiem21a.html
AUTHORS: Derek Hoiem, Tanmay Gupta, Zhizhong Li, Michal Shlapentokh-Rothman
HIGHLIGHT: We propose a method to robustly estimate learning curves, abstract their parameters into error and data-reliance, and evaluate the effectiveness of different parameterizations.

393, TITLE: Equivariant Learning of Stochastic Fields: Gaussian Processes and Steerable Conditional Neural Processes
http://proceedings.mlr.press/v139/holderrieth21a.html
AUTHORS: Peter Holderrieth, Michael J Hutchinson, Yee Whye Teh
HIGHLIGHT: Motivated by objects such as electric fields or fluid streams, we study the problem of learning stochastic fields, i.e. stochastic processes whose samples are fields like those occurring in physics and engineering.

394, TITLE: Latent Programmer: Discrete Latent Codes for Program Synthesis
http://proceedings.mlr.press/v139/hong21a.html
AUTHORS: Joey Hong, David Dohan, Rishabh Singh, Charles Sutton, Manzil Zaheer
HIGHLIGHT: Based on these insights, we introduce the Latent Programmer (LP), a program synthesis method that first predicts a discrete latent code from input/output examples, and then generates the program in the target language.

395, TITLE: Chebyshev Polynomial Codes: Task Entanglement-based Coding for Distributed Matrix Multiplication
http://proceedings.mlr.press/v139/hong21b.html
AUTHORS: Sangwoo Hong, Heecheol Yang, Youngseok Yoon, Taehyun Cho, Jungwoo Lee
HIGHLIGHT: We propose Chebyshev polynomial codes, which can achieve order-wise improvement in encoding complexity at the master and communication load in distributed matrix multiplication using task entanglement.

396, TITLE: Federated Learning of User Verification Models Without Sharing Embeddings
http://proceedings.mlr.press/v139/hosseini21a.html
AUTHORS: Hossein Hosseini, Hyunsin Park, Sungrack Yun, Christos Louizos, Joseph Soriaga, Max Welling
HIGHLIGHT: To address this problem, we propose Federated User Verification (FedUV), a framework in which users jointly learn a set of vectors and maximize the correlation of their instance embeddings with a secret linear combination of those vectors.

397, TITLE: The Limits of Min-Max Optimization Algorithms: Convergence to Spurious Non-Critical Sets
http://proceedings.mlr.press/v139/hsieh21a.html
AUTHORS: Ya-Ping Hsieh, Panayotis Mertikopoulos, Volkan Cevher
HIGHLIGHT: In particular, we show that a wide class of state-of-the-art schemes and heuristics may converge with arbitrarily high probability to attractors that are in no way min-max optimal or even stationary.

398, TITLE: Near-Optimal Representation Learning for Linear Bandits and Linear RL
http://proceedings.mlr.press/v139/hu21a.html
AUTHORS: Jiachen Hu, Xiaoyu Chen, Chi Jin, Lihong Li, Liwei Wang
HIGHLIGHT: We propose a sample-efficient algorithm, MTLR-OFUL, which leverages the shared representation to achieve $\tilde{O}(M\sqrt{dkT} + d\sqrt{kMT})$ regret, with T being the number of total steps.

399, TITLE: On the Random Conjugate Kernel and Neural Tangent Kernel
http://proceedings.mlr.press/v139/hu21b.html
AUTHORS: Zhengmian Hu, Heng Huang
HIGHLIGHT: We investigate the distributions of Conjugate Kernel (CK) and Neural Tangent Kernel (NTK) for ReLU networks with random initialization.

400, TITLE: Off-Belief Learning
http://proceedings.mlr.press/v139/hu21c.html
AUTHORS: Hengyuan Hu, Adam Lerer, Brandon Cui, Luis Pineda, Noam Brown, Jakob Foerster
HIGHLIGHT: Policies learned through self-play may adopt arbitrary conventions and implicitly rely on multi-step reasoning based on fragile assumptions about other agents' actions and thus fail when paired with humans or independently trained agents at test time. To address this, we present off-belief learning (OBL).

401, TITLE: Generalizable Episodic Memory for Deep Reinforcement Learning
http://proceedings.mlr.press/v139/hu21d.html
AUTHORS: Hao Hu, Jianing Ye, Guangxiang Zhu, Zhizhou Ren, Chongjie Zhang
HIGHLIGHT: To address this problem, we propose Generalizable Episodic Memory (GEM), which effectively organizes the state-action values of episodic memory in a generalizable manner and supports implicit planning on memorized trajectories.

402, TITLE: A Scalable Deterministic Global Optimization Algorithm for Clustering Problems
http://proceedings.mlr.press/v139/hua21a.html
AUTHORS: Kaixun Hua, Mingfei Shi, Yankai Cao

HIGHLIGHT: In this paper, we modelled the MSSC task as a two-stage optimization problem and proposed a tailed reduced-space branch and bound (BB) algorithm.

403, TITLE: On Recovering from Modeling Errors Using Testing Bayesian Networks
http://proceedings.mlr.press/v139/huang21a.html
AUTHORS: Haiying Huang, Adnan Darwiche
HIGHLIGHT: We consider the problem of supervised learning with Bayesian Networks when the used dependency structure is incomplete due to missing edges or missing variable states.

404, TITLE: A Novel Sequential Coreset Method for Gradient Descent Algorithms
http://proceedings.mlr.press/v139/huang21b.html
AUTHORS: Jiawei Huang, Ruomin Huang, Wenjie Liu, Nikolaos Freris, Hu Ding
HIGHLIGHT: In this paper, based on the “locality” property of gradient descent algorithms, we propose a new framework, termed “sequential coreset”, which effectively avoids these obstacles.

405, TITLE: FL-NTK: A Neural Tangent Kernel-based Framework for Federated Learning Analysis
http://proceedings.mlr.press/v139/huang21c.html
AUTHORS: Baihe Huang, Xiaoxiao Li, Zhao Song, Xin Yang
HIGHLIGHT: The current paper presents a new class of convergence analysis for FL, Federated Neural Tangent Kernel (FL-NTK), which corresponds to overparamterized ReLU neural networks trained by gradient descent in FL and is inspired by the analysis in Neural Tangent Kernel (NTK).

406, TITLE: STRODE: Stochastic Boundary Ordinary Differential Equation
http://proceedings.mlr.press/v139/huang21d.html
AUTHORS: Hengguan Huang, Hongfu Liu, Hao Wang, Chang Xiao, Ye Wang
HIGHLIGHT: In this paper, we present a probabilistic ordinary differential equation (ODE), called STochastic boundaRy ODE (STRODE), that learns both the timings and the dynamics of time series data without requiring any timing annotations during training.

407, TITLE: A Riemannian Block Coordinate Descent Method for Computing the Projection Robust Wasserstein Distance
http://proceedings.mlr.press/v139/huang21e.html
AUTHORS: Minhui Huang, Shiqian Ma, Lifeng Lai
HIGHLIGHT: In this paper, we propose a Riemannian block coordinate descent (RBCD) method to solve this problem, which is based on a novel reformulation of the regularized max-min problem over the Stiefel manifold.

408, TITLE: Projection Robust Wasserstein Barycenters
http://proceedings.mlr.press/v139/huang21f.html
AUTHORS: Minhui Huang, Shiqian Ma, Lifeng Lai
HIGHLIGHT: This paper proposes the projection robust Wasserstein barycenter (PRWB) that has the potential to mitigate the curse of dimensionality, and a relaxed PRWB (RPRWB) model that is computationally more tractable.

409, TITLE: Accurate Post Training Quantization With Small Calibration Sets
http://proceedings.mlr.press/v139/hubara21a.html
AUTHORS: Itay Hubara, Yury Nahshan, Yair Hanani, Ron Banner, Daniel Soudry
HIGHLIGHT: To this end, we minimize the quantization errors of each layer or block separately by optimizing its parameters over the calibration set.

410, TITLE: Learning and Planning in Complex Action Spaces
http://proceedings.mlr.press/v139/hubert21a.html
AUTHORS: Thomas Hubert, Julian Schrittwieser, Ioannis Antonoglou, Mohammadamin Barekatain, Simon Schmitt, David Silver
HIGHLIGHT: In this paper, we propose a general framework to reason in a principled way about policy evaluation and improvement over such sampled action subsets.

411, TITLE: Generative Adversarial Transformers
http://proceedings.mlr.press/v139/hudson21a.html
AUTHORS: Drew A Hudson, Larry Zitnick
HIGHLIGHT: We introduce the GANsformer, a novel and efficient type of transformer, and explore it for the task of visual generative modeling.

412, TITLE: Neural Pharmacodynamic State Space Modeling

http://proceedings.mlr.press/v139/hussain21a.html
AUTHORS: Zeshan M Hussain, Rahul G. Krishnan, David Sontag
HIGHLIGHT: We propose a deep generative model that makes use of a novel attention-based neural architecture inspired by the physics of how treatments affect disease state.

413, TITLE: Hyperparameter Selection for Imitation Learning
http://proceedings.mlr.press/v139/hussenot21a.html
AUTHORS: L?onard Hussenot, Marcin Andrychowicz, Damien Vincent, Robert Dadashi, Anton Raichuk, Sabela Ramos, Nikola Momchev, Sertan Girgin, Raphael Marinier, Lukasz Stafiniak, Manu Orsini, Olivier Bachem, Matthieu Geist, Olivier Pietquin
HIGHLIGHT: We address the issue of tuning hyperparameters (HPs) for imitation learning algorithms in the context of continuous-control, when the underlying reward function of the demonstrating expert cannot be observed at any time.

414, TITLE: Pareto GAN: Extending the Representational Power of GANs to Heavy-Tailed Distributions
http://proceedings.mlr.press/v139/huster21a.html
AUTHORS: Todd Huster, Jeremy Cohen, Zinan Lin, Kevin Chan, Charles Kamhoua, Nandi O. Leslie, Cho-Yu Jason Chiang, Vyas Sekar
HIGHLIGHT: We identify issues with standard loss functions and propose the use of alternative metric spaces that enable stable and efficient learning.

415, TITLE: LieTransformer: Equivariant Self-Attention for Lie Groups
http://proceedings.mlr.press/v139/hutchinson21a.html
AUTHORS: Michael J Hutchinson, Charline Le Lan, Sheheryar Zaidi, Emilien Dupont, Yee Whye Teh, Hyunjik Kim
HIGHLIGHT: We propose the LieTransformer, an architecture composed of LieSelfAttention layers that are equivariant to arbitrary Lie groups and their discrete subgroups.

416, TITLE: Crowdsourcing via Annotator Co-occurrence Imputation and Provable Symmetric Nonnegative Matrix Factorization
http://proceedings.mlr.press/v139/ibrahim21a.html
AUTHORS: Shahana Ibrahim, Xiao Fu
HIGHLIGHT: This work recasts the pairwise co-occurrence based D&S model learning problem as a symmetric NMF (SymNMF) problem—which offers enhanced identifiability relative to CNMF.

417, TITLE: Selecting Data Augmentation for Simulating Interventions
http://proceedings.mlr.press/v139/ilse21a.html
AUTHORS: Maximilian Ilse, Jakub M Tomczak, Patrick Forr?
HIGHLIGHT: In this paper, we focus on the case where the problem arises through spurious correlation between the observed domains and the actual task labels.

418, TITLE: Scalable Marginal Likelihood Estimation for Model Selection in Deep Learning
http://proceedings.mlr.press/v139/immer21a.html
AUTHORS: Alexander Immer, Matthias Bauer, Vincent Fortuin, Gunnar R?tsch, Khan Mohammad Emtiyaz
HIGHLIGHT: In this work, we present a scalable marginal-likelihood estimation method to select both hyperparameters and network architectures, based on the training data alone.

419, TITLE: Active Learning for Distributionally Robust Level-Set Estimation
http://proceedings.mlr.press/v139/inatsu21a.html
AUTHORS: Yu Inatsu, Shogo Iwazaki, Ichiro Takeuchi
HIGHLIGHT: In this study, we addressed this problem by considering the \textit{distributionally robust PTR} (DRPTR) measure, which considers the worst-case PTR within given candidate distributions.

420, TITLE: Learning Randomly Perturbed Structured Predictors for Direct Loss Minimization
http://proceedings.mlr.press/v139/indelman21a.html
AUTHORS: Hedda Cohen Indelman, Tamir Hazan
HIGHLIGHT: In this work, we interpolate between these techniques by learning the variance of randomized structured predictors as well as their mean, in order to balance between the learned score function and the randomized noise.

421, TITLE: Randomized Entity-wise Factorization for Multi-Agent Reinforcement Learning
http://proceedings.mlr.press/v139/iqbal21a.html
AUTHORS: Shariq Iqbal, Christian A Schroeder De Witt, Bei Peng, Wendelin Boehmer, Shimon Whiteson, Fei Sha
HIGHLIGHT: Our method aims to leverage these commonalities by asking the question: “What is the expected utility of each agent when only considering a randomly selected sub-group of its observed entities?”

422, TITLE: Randomized Exploration in Reinforcement Learning with General Value Function Approximation
http://proceedings.mlr.press/v139/ishfaq21a.html
AUTHORS: Haque Ishfaq, Qiwen Cui, Viet Nguyen, Alex Ayoub, Zhuoran Yang, Zhaoran Wang, Doina Precup, Lin Yang
HIGHLIGHT: We propose a model-free reinforcement learning algorithm inspired by the popular randomized least squares value iteration (RLSVI) algorithm as well as the optimism principle.

423, TITLE: Distributed Second Order Methods with Fast Rates and Compressed Communication
http://proceedings.mlr.press/v139/islamov21a.html
AUTHORS: Rustem Islamov, Xun Qian, Peter Richtarik
HIGHLIGHT: We develop several new communication-efficient second-order methods for distributed optimization.

424, TITLE: What Are Bayesian Neural Network Posteriors Really Like?
http://proceedings.mlr.press/v139/izmailov21a.html
AUTHORS: Pavel Izmailov, Sharad Vikram, Matthew D Hoffman, Andrew Gordon Gordon Wilson
HIGHLIGHT: To investigate foundational questions in Bayesian deep learning, we instead use full batch Hamiltonian Monte Carlo (HMC) on modern architectures.

425, TITLE: How to Learn when Data Reacts to Your Model: Performative Gradient Descent
http://proceedings.mlr.press/v139/izzo21a.html
AUTHORS: Zachary Izzo, Lexing Ying, James Zou
HIGHLIGHT: Here we introduce \emph{performative gradient descent} (PerfGD), an algorithm for computing performatively optimal points.

426, TITLE: Perceiver: General Perception with Iterative Attention
http://proceedings.mlr.press/v139/jaegle21a.html
AUTHORS: Andrew Jaegle, Felix Gimeno, Andy Brock, Oriol Vinyals, Andrew Zisserman, Joao Carreira
HIGHLIGHT: In this paper we introduce the Perceiver {–} a model that builds upon Transformers and hence makes few architectural assumptions about the relationship between its inputs, but that also scales to hundreds of thousands of inputs, like ConvNets.

427, TITLE: Imitation by Predicting Observations
http://proceedings.mlr.press/v139/jaegle21b.html
AUTHORS: Andrew Jaegle, Yury Sulsky, Arun Ahuja, Jake Bruce, Rob Fergus, Greg Wayne
HIGHLIGHT: We present a new method for imitation solely from observations that achieves comparable performance to experts on challenging continuous control tasks while also exhibiting robustness in the presence of observations unrelated to the task.

428, TITLE: Local Correlation Clustering with Asymmetric Classification Errors
http://proceedings.mlr.press/v139/jafarov21a.html
AUTHORS: Jafar Jafarov, Sanchit Kalhan, Konstantin Makarychev, Yury Makarychev
HIGHLIGHT: We study the ℓ_p objective in Correlation Clustering under the following assumption: Every similar edge has weight in $[\alpha\mathbf{w},\mathbf{w}]$ and every dissimilar edge has weight at least $\alpha\mathbf{w}$ (where $\alpha \leq 1$ and $\mathbf{w}>0$ is a scaling parameter).

429, TITLE: Alternative Microfoundations for Strategic Classification
http://proceedings.mlr.press/v139/jagadeesan21a.html
AUTHORS: Meena Jagadeesan, Celestine Mendler-D?nner, Moritz Hardt
HIGHLIGHT: In this work, we argue that a direct combination of these ingredients leads to brittle solution concepts of limited descriptive and prescriptive value.

430, TITLE: Robust Density Estimation from Batches: The Best Things in Life are (Nearly) Free
http://proceedings.mlr.press/v139/jain21a.html
AUTHORS: Ayush Jain, Alon Orlitsky
HIGHLIGHT: We answer this question, showing that, perhaps surprisingly, up to logarithmic factors, the optimal sample complexity is the same as for genuine, non-adversarial, data!

431, TITLE: Instance-Optimal Compressed Sensing via Posterior Sampling
http://proceedings.mlr.press/v139/jalal21a.html
AUTHORS: Ajil Jalal, Sushrut Karmalkar, Alex Dimakis, Eric Price

HIGHLIGHT: We show for Gaussian measurements and \emph{any} prior distribution on the signal, that the posterior sampling estimator achieves near-optimal recovery guarantees.

432, TITLE: Fairness for Image Generation with Uncertain Sensitive Attributes
http://proceedings.mlr.press/v139/jalal21b.html
AUTHORS: Ajil Jalal, Sushrut Karmalkar, Jessica Hoffmann, Alex Dimakis, Eric Price
HIGHLIGHT: This work tackles the issue of fairness in the context of generative procedures, such as image super-resolution, which entail different definitions from the standard classification setting.

433, TITLE: Feature Clustering for Support Identification in Extreme Regions
http://proceedings.mlr.press/v139/jalalzai21a.html
AUTHORS: Hamid Jalalzai, R?mi Leluc
HIGHLIGHT: The present paper develops a novel optimization-based approach to assess the dependence structure of extremes.

434, TITLE: Improved Regret Bounds of Bilinear Bandits using Action Space Analysis
http://proceedings.mlr.press/v139/jang21a.html
AUTHORS: Kyoungseok Jang, Kwang-Sung Jun, Se-Young Yun, Wanmo Kang
HIGHLIGHT: In this paper, we make progress towards closing the gap between the upper and lower bound on the optimal regret.

435, TITLE: Inverse Decision Modeling: Learning Interpretable Representations of Behavior
http://proceedings.mlr.press/v139/jarrett21a.html
AUTHORS: Daniel Jarrett, Alihan H?y?k, Mihaela Van Der Schaar
HIGHLIGHT: In this paper, we develop an expressive, unifying perspective on *inverse decision modeling*: a framework for learning parameterized representations of sequential decision behavior.

436, TITLE: Catastrophic Fisher Explosion: Early Phase Fisher Matrix Impacts Generalization
http://proceedings.mlr.press/v139/jastrzebski21a.html
AUTHORS: Stanislaw Jastrzebski, Devansh Arpit, Oliver Astrand, Giancarlo B Kerg, Huan Wang, Caiming Xiong, Richard Socher, Kyunghyun Cho, Krzysztof J Geras
HIGHLIGHT: We highlight that poor final generalization coincides with the trace of the FIM attaining a large value early in training, to which we refer as catastrophic Fisher explosion.

437, TITLE: Policy Gradient Bayesian Robust Optimization for Imitation Learning
http://proceedings.mlr.press/v139/javed21a.html
AUTHORS: Zaynah Javed, Daniel S Brown, Satvik Sharma, Jerry Zhu, Ashwin Balakrishna, Marek Petrik, Anca Dragan, Ken Goldberg
HIGHLIGHT: We derive a novel policy gradient-style robust optimization approach, PG-BROIL, that optimizes a soft-robust objective that balances expected performance and risk.

438, TITLE: In-Database Regression in Input Sparsity Time
http://proceedings.mlr.press/v139/jayaram21a.html
AUTHORS: Rajesh Jayaram, Alireza Samadian, David Woodruff, Peng Ye
HIGHLIGHT: In this work, we design subspace embeddings for database joins which can be computed significantly faster than computing the join.

439, TITLE: Parallel and Flexible Sampling from Autoregressive Models via Langevin Dynamics
http://proceedings.mlr.press/v139/jayaram21b.html
AUTHORS: Vivek Jayaram, John Thickstun
HIGHLIGHT: This paper introduces an alternative approach to sampling from autoregressive models.

440, TITLE: Objective Bound Conditional Gaussian Process for Bayesian Optimization
http://proceedings.mlr.press/v139/jeong21a.html
AUTHORS: Taewon Jeong, Heeyoung Kim
HIGHLIGHT: In this paper, we propose a new surrogate model, called the objective bound conditional Gaussian process (OBCGP), to condition a Gaussian process on a bound on the optimal function value.

441, TITLE: Quantifying Ignorance in Individual-Level Causal-Effect Estimates under Hidden Confounding
http://proceedings.mlr.press/v139/jesson21a.html

AUTHORS: Andrew Jesson, S?ren Mindermann, Yarin Gal, Uri Shalit
HIGHLIGHT: We present a new parametric interval estimator suited for high-dimensional data, that estimates a range of possible CATE values when given a predefined bound on the level of hidden confounding.

442, TITLE: DeepReDuce: ReLU Reduction for Fast Private Inference
http://proceedings.mlr.press/v139/jha21a.html
AUTHORS: Nandan Kumar Jha, Zahra Ghodsi, Siddharth Garg, Brandon Reagen
HIGHLIGHT: This paper proposes DeepReDuce: a set of optimizations for the judicious removal of ReLUs to reduce private inference latency.

443, TITLE: Factor-analytic inverse regression for high-dimension, small-sample dimensionality reduction
http://proceedings.mlr.press/v139/jha21b.html
AUTHORS: Aditi Jha, Michael J. Morais, Jonathan W Pillow
HIGHLIGHT: To overcome this limitation, we propose Class-conditional Factor Analytic Dimensions (CFAD), a model-based dimensionality reduction method for high-dimensional, small-sample data.

444, TITLE: Fast margin maximization via dual acceleration
http://proceedings.mlr.press/v139/ji21a.html
AUTHORS: Ziwei Ji, Nathan Srebro, Matus Telgarsky
HIGHLIGHT: We present and analyze a momentum-based gradient method for training linear classifiers with an exponentially-tailed loss (e.g., the exponential or logistic loss), which maximizes the classification margin on separable data at a rate of O(1/t^2).

445, TITLE: Marginalized Stochastic Natural Gradients for Black-Box Variational Inference
http://proceedings.mlr.press/v139/ji21b.html
AUTHORS: Geng Ji, Debora Sujono, Erik B Sudderth
HIGHLIGHT: We propose a stochastic natural gradient estimator that is as broadly applicable and unbiased, but improves efficiency by exploiting the curvature of the variational bound, and provably reduces variance by marginalizing discrete latent variables.

446, TITLE: Bilevel Optimization: Convergence Analysis and Enhanced Design
http://proceedings.mlr.press/v139/ji21c.html
AUTHORS: Kaiyi Ji, Junjie Yang, Yingbin Liang
HIGHLIGHT: In this paper, we investigate the nonconvex-strongly-convex bilevel optimization problem.

447, TITLE: Efficient Statistical Tests: A Neural Tangent Kernel Approach
http://proceedings.mlr.press/v139/jia21a.html
AUTHORS: Sheng Jia, Ehsan Nezhadarya, Yuhuai Wu, Jimmy Ba
HIGHLIGHT: We propose a shift-invariant convolutional neural tangent kernel (SCNTK) based outlier detector and two-sample tests with maximum mean discrepancy (MMD) that is O(n) in the number of samples due to using the random feature approximation.

448, TITLE: Scaling Up Visual and Vision-Language Representation Learning With Noisy Text Supervision
http://proceedings.mlr.press/v139/jia21b.html
AUTHORS: Chao Jia, Yinfei Yang, Ye Xia, Yi-Ting Chen, Zarana Parekh, Hieu Pham, Quoc Le, Yun-Hsuan Sung, Zhen Li, Tom Duerig
HIGHLIGHT: In this paper, we leverage a noisy dataset of over one billion image alt-text pairs, obtained without expensive filtering or post-processing steps in the Conceptual Captions dataset.

449, TITLE: Multi-Dimensional Classification via Sparse Label Encoding
http://proceedings.mlr.press/v139/jia21c.html
AUTHORS: Bin-Bin Jia, Min-Ling Zhang
HIGHLIGHT: In this paper, we propose a novel MDC approach named SLEM which learns the predictive model in an encoded label space instead of the original heterogeneous one.

450, TITLE: Self-Damaging Contrastive Learning
http://proceedings.mlr.press/v139/jiang21a.html
AUTHORS: Ziyu Jiang, Tianlong Chen, Bobak J Mortazavi, Zhangyang Wang
HIGHLIGHT: This paper proposes to explicitly tackle this challenge, via a principled framework called Self-Damaging Contrastive Learning (SDCLR), to automatically balance the representation learning without knowing the classes.

451, TITLE: Prioritized Level Replay
http://proceedings.mlr.press/v139/jiang21b.html
AUTHORS: Minqi Jiang, Edward Grefenstette, Tim Rockt?schel
HIGHLIGHT: We introduce Prioritized Level Replay (PLR), a general framework for selectively sampling the next training level by prioritizing those with higher estimated learning potential when revisited in the future.

452, TITLE: Monotonic Robust Policy Optimization with Model Discrepancy
http://proceedings.mlr.press/v139/jiang21c.html
AUTHORS: Yuankun Jiang, Chenglin Li, Wenrui Dai, Junni Zou, Hongkai Xiong
HIGHLIGHT: Since the average and worst-case performance are both important for generalization in RL, in this paper, we propose a policy optimization approach for concurrently improving the policy's performance in the average and worst-case environment.

453, TITLE: Approximation Theory of Convolutional Architectures for Time Series Modelling
http://proceedings.mlr.press/v139/jiang21d.html
AUTHORS: Haotian Jiang, Zhong Li, Qianxiao Li
HIGHLIGHT: In this paper, we derive parallel results for convolutional architectures, with WaveNet being a prime example.

454, TITLE: Streaming and Distributed Algorithms for Robust Column Subset Selection
http://proceedings.mlr.press/v139/jiang21e.html
AUTHORS: Shuli Jiang, Dennis Li, Irene Mengze Li, Arvind V Mahankali, David Woodruff
HIGHLIGHT: We give the first single-pass streaming algorithm for Column Subset Selection with respect to the entrywise ℓ_p-norm with $1 \leq p < 2$.

455, TITLE: Single Pass Entrywise-Transformed Low Rank Approximation
http://proceedings.mlr.press/v139/jiang21f.html
AUTHORS: Yifei Jiang, Yi Li, Yiming Sun, Jiaxin Wang, David Woodruff
HIGHLIGHT: In this paper we resolve this open question, obtaining the first single-pass algorithm for this problem and for the same class of functions f studied by Liang et al.

456, TITLE: The Emergence of Individuality
http://proceedings.mlr.press/v139/jiang21g.html
AUTHORS: Jiechuan Jiang, Zongqing Lu
HIGHLIGHT: Inspired by that individuality is of being an individual separate from others, we propose a simple yet efficient method for the emergence of individuality (EOI) in multi-agent reinforcement learning (MARL).

457, TITLE: Online Selection Problems against Constrained Adversary
http://proceedings.mlr.press/v139/jiang21h.html
AUTHORS: Zhihao Jiang, Pinyan Lu, Zhihao Gavin Tang, Yuhao Zhang
HIGHLIGHT: Inspired by a recent line of work in online algorithms with predictions, we study the constrained adversary model that utilizes predictions from a different perspective.

458, TITLE: Active Covering
http://proceedings.mlr.press/v139/jiang21i.html
AUTHORS: Heinrich Jiang, Afshin Rostamizadeh
HIGHLIGHT: We analyze the problem of active covering, where the learner is given an unlabeled dataset and can sequentially label query examples.

459, TITLE: Emphatic Algorithms for Deep Reinforcement Learning
http://proceedings.mlr.press/v139/jiang21j.html
AUTHORS: Ray Jiang, Tom Zahavy, Zhongwen Xu, Adam White, Matteo Hessel, Charles Blundell, Hado Van Hasselt
HIGHLIGHT: In this paper, we extend the use of emphatic methods to deep reinforcement learning agents.

460, TITLE: Characterizing Structural Regularities of Labeled Data in Overparameterized Models
http://proceedings.mlr.press/v139/jiang21k.html
AUTHORS: Ziheng Jiang, Chiyuan Zhang, Kunal Talwar, Michael C Mozer
HIGHLIGHT: We analyze how individual instances are treated by a model via a consistency score. The score characterizes the expected accuracy for a held-out instance given training sets of varying size sampled from the data distribution.

461, TITLE: Optimal Streaming Algorithms for Multi-Armed Bandits
http://proceedings.mlr.press/v139/jin21a.html
AUTHORS: Tianyuan Jin, Keke Huang, Jing Tang, Xiaokui Xiao
HIGHLIGHT: We propose an algorithm that works for any k and achieves the optimal sample complexity O(\frac{n}{\epsilon^2} \log\frac{k}{\delta}) using a single-arm memory and a single pass of the stream.

462, TITLE: Towards Tight Bounds on the Sample Complexity of Average-reward MDPs
http://proceedings.mlr.press/v139/jin21b.html
AUTHORS: Yujia Jin, Aaron Sidford
HIGHLIGHT: When the mixing time of the probability transition matrix of all policies is at most t_mix, we provide an algorithm that solves the problem using $\widetilde{O}(t_\mathrm{mix} \epsilon^{-3})$ (oblivious) samples per state-action pair.

463, TITLE: Almost Optimal Anytime Algorithm for Batched Multi-Armed Bandits
http://proceedings.mlr.press/v139/jin21c.html
AUTHORS: Tianyuan Jin, Jing Tang, Pan Xu, Keke Huang, Xiaokui Xiao, Quanquan Gu
HIGHLIGHT: In this paper, we study the anytime batched multi-armed bandit problem.

464, TITLE: MOTS: Minimax Optimal Thompson Sampling
http://proceedings.mlr.press/v139/jin21d.html
AUTHORS: Tianyuan Jin, Pan Xu, Jieming Shi, Xiaokui Xiao, Quanquan Gu
HIGHLIGHT: In this paper we fill this long open gap by proposing a new Thompson sampling algorithm called MOTS that adaptively truncates the sampling result of the chosen arm at each time step.

465, TITLE: Is Pessimism Provably Efficient for Offline RL?
http://proceedings.mlr.press/v139/jin21e.html
AUTHORS: Ying Jin, Zhuoran Yang, Zhaoran Wang
HIGHLIGHT: In this paper, we propose a pessimistic variant of the value iteration algorithm (PEVI), which incorporates an uncertainty quantifier as the penalty function.

466, TITLE: Adversarial Option-Aware Hierarchical Imitation Learning
http://proceedings.mlr.press/v139/jing21a.html
AUTHORS: Mingxuan Jing, Wenbing Huang, Fuchun Sun, Xiaojian Ma, Tao Kong, Chuang Gan, Lei Li
HIGHLIGHT: In this paper, we propose Option-GAIL, a novel method to learn skills at long horizon.

467, TITLE: Discrete-Valued Latent Preference Matrix Estimation with Graph Side Information
http://proceedings.mlr.press/v139/jo21a.html
AUTHORS: Changhun Jo, Kangwook Lee
HIGHLIGHT: In this work, we propose a new model in which 1) the unknown latent preference matrix can have any discrete values, and 2) users can be clustered into multiple clusters, thereby relaxing the assumptions made in prior work.

468, TITLE: Provable Lipschitz Certification for Generative Models
http://proceedings.mlr.press/v139/jordan21a.html
AUTHORS: Matt Jordan, Alex Dimakis
HIGHLIGHT: We present a scalable technique for upper bounding the Lipschitz constant of generative models.

469, TITLE: Isometric Gaussian Process Latent Variable Model for Dissimilarity Data
http://proceedings.mlr.press/v139/jorgensen21a.html
AUTHORS: Martin J?rgensen, Soren Hauberg
HIGHLIGHT: We present a probabilistic model where the latent variable respects both the distances and the topology of the modeled data.

470, TITLE: On the Generalization Power of Overfitted Two-Layer Neural Tangent Kernel Models
http://proceedings.mlr.press/v139/ju21a.html
AUTHORS: Peizhong Ju, Xiaojun Lin, Ness Shroff
HIGHLIGHT: In this paper, we study the generalization performance of min ℓ_2-norm overfitting solutions for the neural tangent kernel (NTK) model of a two-layer neural network with ReLU activation that has no bias term.

471, TITLE: Improved Confidence Bounds for the Linear Logistic Model and Applications to Bandits
http://proceedings.mlr.press/v139/jun21a.html

AUTHORS: Kwang-Sung Jun, Lalit Jain, Houssam Nassif, Blake Mason
HIGHLIGHT: We propose improved fixed-design confidence bounds for the linear logistic model.

472, TITLE: Detection of Signal in the Spiked Rectangular Models
http://proceedings.mlr.press/v139/jung21a.html
AUTHORS: Ji Hyung Jung, Hye Won Chung, Ji Oon Lee
HIGHLIGHT: We consider the problem of detecting signals in the rank-one signal-plus-noise data matrix models that generalize the spiked Wishart matrices.

473, TITLE: Estimating Identifiable Causal Effects on Markov Equivalence Class through Double Machine Learning
http://proceedings.mlr.press/v139/jung21b.html
AUTHORS: Yonghan Jung, Jin Tian, Elias Bareinboim
HIGHLIGHT: In this paper, we study the problem of causal estimation from a MEC represented by a partial ancestral graph (PAG), which is learnable from observational data.

474, TITLE: A Nullspace Property for Subspace-Preserving Recovery
http://proceedings.mlr.press/v139/kaba21a.html
AUTHORS: Mustafa D Kaba, Chong You, Daniel P Robinson, Enrique Mallada, Rene Vidal
HIGHLIGHT: This paper derives a necessary and sufficient condition for subspace-preserving recovery that is inspired by the classical nullspace property.Based on this novel condition, called here the subspace nullspace property, we derive equivalent characterizations that either admit a clear geometric interpretation that relates data distribution and subspace separation to the recovery success, or can be verified using a finite set of extreme points of a properly defined set.

475, TITLE: Training Recurrent Neural Networks via Forward Propagation Through Time
http://proceedings.mlr.press/v139/kag21a.html
AUTHORS: Anil Kag, Venkatesh Saligrama
HIGHLIGHT: We propose a novel forward-propagation algorithm, FPTT, where at each time, for an instance, we update RNN parameters by optimizing an instantaneous risk function.

476, TITLE: The Distributed Discrete Gaussian Mechanism for Federated Learning with Secure Aggregation
http://proceedings.mlr.press/v139/kairouz21a.html
AUTHORS: Peter Kairouz, Ziyu Liu, Thomas Steinke
HIGHLIGHT: We present a comprehensive end-to-end system, which appropriately discretizes the data and adds discrete Gaussian noise before performing secure aggregation.

477, TITLE: Practical and Private (Deep) Learning Without Sampling or Shuffling
http://proceedings.mlr.press/v139/kairouz21b.html
AUTHORS: Peter Kairouz, Brendan Mcmahan, Shuang Song, Om Thakkar, Abhradeep Thakurta, Zheng Xu
HIGHLIGHT: We consider training models with differential privacy (DP) using mini-batch gradients.

478, TITLE: A Differentiable Point Process with Its Application to Spiking Neural Networks
http://proceedings.mlr.press/v139/kajino21a.html
AUTHORS: Hiroshi Kajino
HIGHLIGHT: This paper is concerned about a learning algorithm for a probabilistic model of spiking neural networks (SNNs).

479, TITLE: Projection techniques to update the truncated SVD of evolving matrices with applications
http://proceedings.mlr.press/v139/kalantzis21a.html
AUTHORS: Vasileios Kalantzis, Georgios Kollias, Shashanka Ubaru, Athanasios N. Nikolakopoulos, Lior Horesh, Kenneth Clarkson
HIGHLIGHT: The algorithm presented in this paper undertakes a projection viewpoint and focuses on building a pair of subspaces which approximate the linear span of the sought singular vectors of the updated matrix.

480, TITLE: Optimal Off-Policy Evaluation from Multiple Logging Policies
http://proceedings.mlr.press/v139/kallus21a.html
AUTHORS: Nathan Kallus, Yuta Saito, Masatoshi Uehara
HIGHLIGHT: In this paper, we resolve this dilemma by finding the OPE estimator for multiple loggers with minimum variance for any instance, i.e., the efficient one.

481, TITLE: Efficient Performance Bounds for Primal-Dual Reinforcement Learning from Demonstrations
http://proceedings.mlr.press/v139/kamoutsi21a.html

AUTHORS: Angeliki Kamoutsi, Goran Banjac, John Lygeros
HIGHLIGHT: To bridge the gap between theory and practice, we introduce a novel bilinear saddle-point framework using Lagrangian duality.

482, TITLE: Statistical Estimation from Dependent Data
http://proceedings.mlr.press/v139/kandiros21a.html
AUTHORS: Vardis Kandiros, Yuval Dagan, Nishanth Dikkala, Surbhi Goel, Constantinos Daskalakis
HIGHLIGHT: As our main contribution we provide algorithms and statistically efficient estimation rates for this model, giving several instantiations of our bounds in logistic regression, sparse logistic regression, and neural network regression settings with dependent data.

483, TITLE: SKIing on Simplices: Kernel Interpolation on the Permutohedral Lattice for Scalable Gaussian Processes
http://proceedings.mlr.press/v139/kapoor21a.html
AUTHORS: Sanyam Kapoor, Marc Finzi, Ke Alexander Wang, Andrew Gordon Gordon Wilson
HIGHLIGHT: In this work, we develop a connection between SKI and the permutohedral lattice used for high-dimensional fast bilateral filtering.

484, TITLE: Variational Auto-Regressive Gaussian Processes for Continual Learning
http://proceedings.mlr.press/v139/kapoor21b.html
AUTHORS: Sanyam Kapoor, Theofanis Karaletsos, Thang D Bui
HIGHLIGHT: By relying on sparse inducing point approximations for scalable posteriors, we propose a novel auto-regressive variational distribution which reveals two fruitful connections to existing results in Bayesian inference, expectation propagation and orthogonal inducing points.

485, TITLE: Off-Policy Confidence Sequences
http://proceedings.mlr.press/v139/karampatziakis21a.html
AUTHORS: Nikos Karampatziakis, Paul Mineiro, Aaditya Ramdas
HIGHLIGHT: We develop confidence bounds that hold uniformly over time for off-policy evaluation in the contextual bandit setting.

486, TITLE: Learning from History for Byzantine Robust Optimization
http://proceedings.mlr.press/v139/karimireddy21a.html
AUTHORS: Sai Praneeth Karimireddy, Lie He, Martin Jaggi
HIGHLIGHT: To address these issues, we present two surprisingly simple strategies: a new robust iterative clipping procedure, and incorporating worker momentum to overcome time-coupled attacks.

487, TITLE: Non-Negative Bregman Divergence Minimization for Deep Direct Density Ratio Estimation
http://proceedings.mlr.press/v139/kato21a.html
AUTHORS: Masahiro Kato, Takeshi Teshima
HIGHLIGHT: In this paper, to mitigate train-loss hacking, we propose non-negative correction for empirical BD estimators.

488, TITLE: Improved Algorithms for Agnostic Pool-based Active Classification
http://proceedings.mlr.press/v139/katz-samuels21a.html
AUTHORS: Julian Katz-Samuels, Jifan Zhang, Lalit Jain, Kevin Jamieson
HIGHLIGHT: In this work we propose an algorithm that, in contrast to uniform sampling over the disagreement region, solves an experimental design problem to determine a distribution over examples from which to request labels.

489, TITLE: When Does Data Augmentation Help With Membership Inference Attacks?
http://proceedings.mlr.press/v139/kaya21a.html
AUTHORS: Yigitcan Kaya, Tudor Dumitras
HIGHLIGHT: Employing two recent MIAs, we explore the lower bound on the risk in the absence of formal upper bounds.

490, TITLE: Regularized Submodular Maximization at Scale
http://proceedings.mlr.press/v139/kazemi21a.html
AUTHORS: Ehsan Kazemi, Shervin Minaee, Moran Feldman, Amin Karbasi
HIGHLIGHT: In this paper, we propose scalable methods for maximizing a regularized submodular function $f \triangleq g-\ell$ expressed as the difference between a monotone submodular function g and a modular function ℓ.

491, TITLE: Prior Image-Constrained Reconstruction using Style-Based Generative Models
http://proceedings.mlr.press/v139/kelkar21a.html

AUTHORS: Varun A Kelkar, Mark Anastasio
HIGHLIGHT: In this study, a framework for estimating an object of interest that is semantically related to a known prior image, is proposed.

492, TITLE: Self Normalizing Flows
http://proceedings.mlr.press/v139/keller21a.html
AUTHORS: Thomas A Keller, Jorn W.T. Peters, Priyank Jaini, Emiel Hoogeboom, Patrick Forr?, Max Welling
HIGHLIGHT: In this work, we propose \emph{Self Normalizing Flows}, a flexible framework for training normalizing flows by replacing expensive terms in the gradient by learned approximate inverses at each layer.

493, TITLE: Interpretable Stability Bounds for Spectral Graph Filters
http://proceedings.mlr.press/v139/kenlay21a.html
AUTHORS: Henry Kenlay, Dorina Thanou, Xiaowen Dong
HIGHLIGHT: In this paper, we study filter stability and provide a novel and interpretable upper bound on the change of filter output, where the bound is expressed in terms of the endpoint degrees of the deleted and newly added edges, as well as the spatial proximity of those edges.

494, TITLE: Affine Invariant Analysis of Frank-Wolfe on Strongly Convex Sets
http://proceedings.mlr.press/v139/kerdreux21a.html
AUTHORS: Thomas Kerdreux, Lewis Liu, Simon Lacoste-Julien, Damien Scieur
HIGHLIGHT: In this work, we introduce new structural assumptions on the problem (such as the directional smoothness) and derive an affine invariant, norm-independent analysis of Frank-Wolfe.

495, TITLE: Markpainting: Adversarial Machine Learning meets Inpainting
http://proceedings.mlr.press/v139/khachaturov21a.html
AUTHORS: David Khachaturov, Ilia Shumailov, Yiren Zhao, Nicolas Papernot, Ross Anderson
HIGHLIGHT: In this paper we study how to manipulate it using our markpainting technique.

496, TITLE: Finite-Sample Analysis of Off-Policy Natural Actor-Critic Algorithm
http://proceedings.mlr.press/v139/khodadadian21a.html
AUTHORS: Sajad Khodadadian, Zaiwei Chen, Siva Theja Maguluri
HIGHLIGHT: In this paper, we provide finite-sample convergence guarantees for an off-policy variant of the natural actor-critic (NAC) algorithm based on Importance Sampling.

497, TITLE: Functional Space Analysis of Local GAN Convergence
http://proceedings.mlr.press/v139/khrulkov21a.html
AUTHORS: Valentin Khrulkov, Artem Babenko, Ivan Oseledets
HIGHLIGHT: We propose a novel perspective where we study the local dynamics of adversarial training in the general functional space and show how it can be represented as a system of partial differential equations.

498, TITLE: "Hey, that's not an ODE": Faster ODE Adjoints via Seminorms
http://proceedings.mlr.press/v139/kidger21a.html
AUTHORS: Patrick Kidger, Ricky T. Q. Chen, Terry J Lyons
HIGHLIGHT: Here, we demonstrate that the particular structure of the adjoint equations makes the usual choices of norm (such as L^2) unnecessarily stringent.

499, TITLE: Neural SDEs as Infinite-Dimensional GANs
http://proceedings.mlr.press/v139/kidger21b.html
AUTHORS: Patrick Kidger, James Foster, Xuechen Li, Terry J Lyons
HIGHLIGHT: Here, we show that the current classical approach to fitting SDEs may be approached as a special case of (Wasserstein) GANs, and in doing so the neural and classical regimes may be brought together.

500, TITLE: GRAD-MATCH: Gradient Matching based Data Subset Selection for Efficient Deep Model Training
http://proceedings.mlr.press/v139/killamsetty21a.html
AUTHORS: Krishnateja Killamsetty, Durga S, Ganesh Ramakrishnan, Abir De, Rishabh Iyer
HIGHLIGHT: In this work, we propose a general framework, GRAD-MATCH, which finds subsets that closely match the gradient of the \emph{training or validation} set.

501, TITLE: Improving Predictors via Combination Across Diverse Task Categories
http://proceedings.mlr.press/v139/kim21a.html

AUTHORS: Kwang In Kim
HIGHLIGHT: Our algorithm aligns the heterogeneous domains of different predictors in a shared latent space to facilitate comparisons of predictors independently of the domains on which they are originally defined.

502, TITLE: Self-Improved Retrosynthetic Planning
http://proceedings.mlr.press/v139/kim21b.html
AUTHORS: Junsu Kim, Sungsoo Ahn, Hankook Lee, Jinwoo Shin
HIGHLIGHT: Motivated by this, we propose an end-to-end framework for directly training the DNNs towards generating reaction pathways with the desirable properties.

503, TITLE: Reward Identification in Inverse Reinforcement Learning
http://proceedings.mlr.press/v139/kim21c.html
AUTHORS: Kuno Kim, Shivam Garg, Kirankumar Shiragur, Stefano Ermon
HIGHLIGHT: In this work, we formalize the reward identification problem in IRL and study how identifiability relates to properties of the MDP model.

504, TITLE: I-BERT: Integer-only BERT Quantization
http://proceedings.mlr.press/v139/kim21d.html
AUTHORS: Sehoon Kim, Amir Gholami, Zhewei Yao, Michael W. Mahoney, Kurt Keutzer
HIGHLIGHT: In this work, we propose I-BERT, a novel quantization scheme for Transformer based models that quantizes the entire inference with integer-only arithmetic.

505, TITLE: Message Passing Adaptive Resonance Theory for Online Active Semi-supervised Learning
http://proceedings.mlr.press/v139/kim21e.html
AUTHORS: Taehyeong Kim, Injune Hwang, Hyundo Lee, Hyunseo Kim, Won-Seok Choi, Joseph J Lim, Byoung-Tak Zhang
HIGHLIGHT: In this study, we propose Message Passing Adaptive Resonance Theory (MPART) that learns the distribution and topology of input data online.

506, TITLE: Conditional Variational Autoencoder with Adversarial Learning for End-to-End Text-to-Speech
http://proceedings.mlr.press/v139/kim21f.html
AUTHORS: Jaehyeon Kim, Jungil Kong, Juhee Son
HIGHLIGHT: In this work, we present a parallel end-to-end TTS method that generates more natural sounding audio than current two-stage models.

507, TITLE: A Policy Gradient Algorithm for Learning to Learn in Multiagent Reinforcement Learning
http://proceedings.mlr.press/v139/kim21g.html
AUTHORS: Dong Ki Kim, Miao Liu, Matthew D Riemer, Chuangchuang Sun, Marwa Abdulhai, Golnaz Habibi, Sebastian Lopez-Cot, Gerald Tesauro, Jonathan How
HIGHLIGHT: In this paper, we propose a novel meta-multiagent policy gradient theorem that directly accounts for the non-stationary policy dynamics inherent to multiagent learning settings.

508, TITLE: Inferring Latent Dynamics Underlying Neural Population Activity via Neural Differential Equations
http://proceedings.mlr.press/v139/kim21h.html
AUTHORS: Timothy D Kim, Thomas Z Luo, Jonathan W Pillow, Carlos Brody
HIGHLIGHT: Here we address this problem by introducing a low-dimensional nonlinear model for latent neural population dynamics using neural ordinary differential equations (neural ODEs), with noisy sensory inputs and Poisson spike train outputs.

509, TITLE: The Lipschitz Constant of Self-Attention
http://proceedings.mlr.press/v139/kim21i.html
AUTHORS: Hyunjik Kim, George Papamakarios, Andriy Mnih
HIGHLIGHT: In this paper, we investigate the Lipschitz constant of self-attention, a non-linear neural network module widely used in sequence modelling.

510, TITLE: Unsupervised Skill Discovery with Bottleneck Option Learning
http://proceedings.mlr.press/v139/kim21j.html
AUTHORS: Jaekyeom Kim, Seohong Park, Gunhee Kim
HIGHLIGHT: We propose a novel unsupervised skill discovery method named Information Bottleneck Option Learning (IBOL).

511, TITLE: ViLT: Vision-and-Language Transformer Without Convolution or Region Supervision
http://proceedings.mlr.press/v139/kim21k.html
AUTHORS: Wonjae Kim, Bokyung Son, Ildoo Kim
HIGHLIGHT: In this paper, we present a minimal VLP model, Vision-and-Language Transformer (ViLT), monolithic in the sense that the processing of visual inputs is drastically simplified to just the same convolution-free manner that we process textual inputs.

512, TITLE: Bias-Robust Bayesian Optimization via Dueling Bandits
http://proceedings.mlr.press/v139/kirschner21a.html
AUTHORS: Johannes Kirschner, Andreas Krause
HIGHLIGHT: Our first contribution is a reduction of the confounded setting to the dueling bandit model. Then we propose a novel approach for dueling bandits based on information-directed sampling (IDS).

513, TITLE: CLOCS: Contrastive Learning of Cardiac Signals Across Space, Time, and Patients
http://proceedings.mlr.press/v139/kiyasseh21a.html
AUTHORS: Dani Kiyasseh, Tingting Zhu, David A Clifton
HIGHLIGHT: We propose a family of contrastive learning methods, CLOCS, that encourages representations across space, time, \textit{and} patients to be similar to one another.

514, TITLE: Scalable Optimal Transport in High Dimensions for Graph Distances, Embedding Alignment, and More
http://proceedings.mlr.press/v139/klicpera21a.html
AUTHORS: Johannes Klicpera, Marten Lienen, Stephan G?nnemann
HIGHLIGHT: In this work we propose two effective log-linear time approximations of the cost matrix: First, a sparse approximation based on locality sensitive hashing (LSH) and, second, a Nystr{ö}m approximation with LSH-based sparse corrections, which we call locally corrected Nystr{ö}m (LCN).

515, TITLE: Representational aspects of depth and conditioning in normalizing flows
http://proceedings.mlr.press/v139/koehler21a.html
AUTHORS: Frederic Koehler, Viraj Mehta, Andrej Risteski
HIGHLIGHT: In our paper, we tackle representational aspects around depth and conditioning of normalizing flows: both for general invertible architectures, and for a particular common architecture, affine couplings.

516, TITLE: WILDS: A Benchmark of in-the-Wild Distribution Shifts
http://proceedings.mlr.press/v139/koh21a.html
AUTHORS: Pang Wei Koh, Shiori Sagawa, Henrik Marklund, Sang Michael Xie, Marvin Zhang, Akshay Balsubramani, Weihua Hu, Michihiro Yasunaga, Richard Lanas Phillips, Irena Gao, Tony Lee, Etienne David, Ian Stavness, Wei Guo, Berton Earnshaw, Imran Haque, Sara M Beery, Jure Leskovec, Anshul Kundaje, Emma Pierson, Sergey Levine, Chelsea Finn, Percy Liang
HIGHLIGHT: To address this gap, we present WILDS, a curated benchmark of 10 datasets reflecting a diverse range of distribution shifts that naturally arise in real-world applications, such as shifts across hospitals for tumor identification; across camera traps for wildlife monitoring; and across time and location in satellite imaging and poverty mapping.

517, TITLE: One-sided Frank-Wolfe algorithms for saddle problems
http://proceedings.mlr.press/v139/kolmogorov21a.html
AUTHORS: Vladimir Kolmogorov, Thomas Pock
HIGHLIGHT: We study a class of convex-concave saddle-point problems of the form $\min_x\max_y ?Kx,y?+f_{\cal P}(x)-h^*(y)$ where K is a linear operator, $f_{\cal P}$ is the sum of a convex function f with a Lipschitz-continuous gradient and the indicator function of a bounded convex polytope ${\cal P}$, and $h^\ast$ is a convex (possibly nonsmooth) function.

518, TITLE: A Lower Bound for the Sample Complexity of Inverse Reinforcement Learning
http://proceedings.mlr.press/v139/komanduru21a.html
AUTHORS: Abi Komanduru, Jean Honorio
HIGHLIGHT: This paper develops an information-theoretic lower bound for the sample complexity of the finite state, finite action IRL problem.

519, TITLE: Consensus Control for Decentralized Deep Learning
http://proceedings.mlr.press/v139/kong21a.html
AUTHORS: Lingjing Kong, Tao Lin, Anastasia Koloskova, Martin Jaggi, Sebastian Stich
HIGHLIGHT: We identify the changing consensus distance between devices as a key parameter to explain the gap between centralized and decentralized training.

520, TITLE: A Distribution-dependent Analysis of Meta Learning

http://proceedings.mlr.press/v139/konobeev21a.html
AUTHORS: Mikhail Konobeev, Ilja Kuzborskij, Csaba Szepesvari
HIGHLIGHT: For this case we propose to adopt the EM method, which is shown to enjoy efficient updates in our case.

521, TITLE: Evaluating Robustness of Predictive Uncertainty Estimation: Are Dirichlet-based Models Reliable?
http://proceedings.mlr.press/v139/kopetzki21a.html
AUTHORS: Anna-Kathrin Kopetzki, Bertrand Charpentier, Daniel Z?gner, Sandhya Giri, Stephan G?nnemann
HIGHLIGHT: In this work, we present the first large-scale, in-depth study of the robustness of DBU models under adversarial attacks.

522, TITLE: Kernel Stein Discrepancy Descent
http://proceedings.mlr.press/v139/korba21a.html
AUTHORS: Anna Korba, Pierre-Cyril Aubin-Frankowski, Szymon Majewski, Pierre Ablin
HIGHLIGHT: We investigate the properties of its Wasserstein gradient flow to approximate a target probability distribution π on $\mathbb{R}^d$, known up to a normalization constant.

523, TITLE: Boosting the Throughput and Accelerator Utilization of Specialized CNN Inference Beyond Increasing Batch Size
http://proceedings.mlr.press/v139/kosaian21a.html
AUTHORS: Jack Kosaian, Amar Phanishayee, Matthai Philipose, Debadeepta Dey, Rashmi Vinayak
HIGHLIGHT: We propose FoldedCNNs, a new approach to CNN design that increases inference throughput and utilization beyond large batch size.

524, TITLE: NeRF-VAE: A Geometry Aware 3D Scene Generative Model
http://proceedings.mlr.press/v139/kosiorek21a.html
AUTHORS: Adam R Kosiorek, Heiko Strathmann, Daniel Zoran, Pol Moreno, Rosalia Schneider, Sona Mokra, Danilo Jimenez Rezende
HIGHLIGHT: We propose NeRF-VAE, a 3D scene generative model that incorporates geometric structure via Neural Radiance Fields (NeRF) and differentiable volume rendering.

525, TITLE: Active Testing: Sample-Efficient Model Evaluation
http://proceedings.mlr.press/v139/kossen21a.html
AUTHORS: Jannik Kossen, Sebastian Farquhar, Yarin Gal, Tom Rainforth
HIGHLIGHT: We introduce a new framework for sample-efficient model evaluation that we call active testing.

526, TITLE: High Confidence Generalization for Reinforcement Learning
http://proceedings.mlr.press/v139/kostas21a.html
AUTHORS: James Kostas, Yash Chandak, Scott M Jordan, Georgios Theocharous, Philip Thomas
HIGHLIGHT: We present several classes of reinforcement learning algorithms that safely generalize to Markov decision processes (MDPs) not seen during training.

527, TITLE: Offline Reinforcement Learning with Fisher Divergence Critic Regularization
http://proceedings.mlr.press/v139/kostrikov21a.html
AUTHORS: Ilya Kostrikov, Rob Fergus, Jonathan Tompson, Ofir Nachum
HIGHLIGHT: In this work, we propose an alternative approach to encouraging the learned policy to stay close to the data, namely parameterizing the critic as the log-behavior-policy, which generated the offline data, plus a state-action value offset term, which can be learned using a neural network.

528, TITLE: ADOM: Accelerated Decentralized Optimization Method for Time-Varying Networks
http://proceedings.mlr.press/v139/kovalev21a.html
AUTHORS: Dmitry Kovalev, Egor Shulgin, Peter Richtarik, Alexander V Rogozin, Alexander Gasnikov
HIGHLIGHT: We propose ADOM - an accelerated method for smooth and strongly convex decentralized optimization over time-varying networks.

529, TITLE: Revisiting Peng's Q($?$) for Modern Reinforcement Learning
http://proceedings.mlr.press/v139/kozuno21a.html
AUTHORS: Tadashi Kozuno, Yunhao Tang, Mark Rowland, Remi Munos, Steven Kapturowski, Will Dabney, Michal Valko, David Abel
HIGHLIGHT: Motivated by the empirical results and the lack of theory, we carry out theoretical analyses of Peng’s Q(\lambda), a representative example of non-conservative algorithms.

530, TITLE: Adapting to misspecification in contextual bandits with offline regression oracles
http://proceedings.mlr.press/v139/krishnamurthy21a.html
AUTHORS: Sanath Kumar Krishnamurthy, Vitor Hadad, Susan Athey
HIGHLIGHT: We propose a simple family of contextual bandit algorithms that adapt to misspecification error by reverting to a good safe policy when there is evidence that misspecification is causing a regret increase.

531, TITLE: Out-of-Distribution Generalization via Risk Extrapolation (REx)
http://proceedings.mlr.press/v139/krueger21a.html
AUTHORS: David Krueger, Ethan Caballero, Joern-Henrik Jacobsen, Amy Zhang, Jonathan Binas, Dinghuai Zhang, Remi Le Priol, Aaron Courville
HIGHLIGHT: We motivate this approach, Risk Extrapolation (REx), as a form of robust optimization over a perturbation set of extrapolated domains (MM-REx), and propose a penalty on the variance of training risks (V-REx) as a simpler variant.

532, TITLE: Near-Optimal Confidence Sequences for Bounded Random Variables
http://proceedings.mlr.press/v139/kuchibhotla21a.html
AUTHORS: Arun K Kuchibhotla, Qinqing Zheng
HIGHLIGHT: To address this question, we provide a near-optimal confidence sequence for bounded random variables by utilizing Bentkus' concentration results.

533, TITLE: Differentially Private Bayesian Inference for Generalized Linear Models
http://proceedings.mlr.press/v139/kulkarni21a.html
AUTHORS: Tejas Kulkarni, Joonas J?lk?, Antti Koskela, Samuel Kaski, Antti Honkela
HIGHLIGHT: In this work, with logistic and Poisson regression as running examples, we introduce a generic noise-aware DP Bayesian inference method for a GLM at hand, given a noisy sum of summary statistics.

534, TITLE: Bayesian Structural Adaptation for Continual Learning
http://proceedings.mlr.press/v139/kumar21a.html
AUTHORS: Abhishek Kumar, Sunabha Chatterjee, Piyush Rai
HIGHLIGHT: We present a novel Bayesian framework based on continually learning the structure of deep neural networks, to unify these distinct yet complementary approaches.

535, TITLE: Implicit rate-constrained optimization of non-decomposable objectives
http://proceedings.mlr.press/v139/kumar21b.html
AUTHORS: Abhishek Kumar, Harikrishna Narasimhan, Andrew Cotter
HIGHLIGHT: Our key idea is to formulate a rate-constrained optimization that expresses the threshold parameter as a function of the model parameters via the Implicit Function theorem.

536, TITLE: A Scalable Second Order Method for Ill-Conditioned Matrix Completion from Few Samples
http://proceedings.mlr.press/v139/kummerle21a.html
AUTHORS: Christian K?mmerle, Claudio M. Verdun
HIGHLIGHT: We propose an iterative algorithm for low-rank matrix completion with that can be interpreted as an iteratively reweighted least squares (IRLS) algorithm, a saddle-escaping smoothing Newton method or a variable metric proximal gradient method applied to a non-convex rank surrogate.

537, TITLE: Meta-Thompson Sampling
http://proceedings.mlr.press/v139/kveton21a.html
AUTHORS: Branislav Kveton, Mikhail Konobeev, Manzil Zaheer, Chih-Wei Hsu, Martin Mladenov, Craig Boutilier, Csaba Szepesvari
HIGHLIGHT: We propose several efficient implementations of MetaTS and analyze it in Gaussian bandits.

538, TITLE: Targeted Data Acquisition for Evolving Negotiation Agents
http://proceedings.mlr.press/v139/kwon21a.html
AUTHORS: Minae Kwon, Siddharth Karamcheti, Mariano-Florentino Cuellar, Dorsa Sadigh
HIGHLIGHT: To address this, we introduce a targeted data acquisition framework where we guide the exploration of a reinforcement learning agent using annotations from an expert oracle.

539, TITLE: ASAM: Adaptive Sharpness-Aware Minimization for Scale-Invariant Learning of Deep Neural Networks
http://proceedings.mlr.press/v139/kwon21b.html
AUTHORS: Jungmin Kwon, Jeongseop Kim, Hyunseo Park, In Kwon Choi

HIGHLIGHT: In this paper, we introduce the concept of adaptive sharpness which is scale-invariant and propose the corresponding generalization bound.

540, TITLE: On the price of explainability for some clustering problems
http://proceedings.mlr.press/v139/laber21a.html
AUTHORS: Eduardo S Laber, Lucas Murtinho
HIGHLIGHT: Here, we study this price for the following clustering problems: k-means, k-medians, k-centers and maximum-spacing.

541, TITLE: Adaptive Newton Sketch: Linear-time Optimization with Quadratic Convergence and Effective Hessian Dimensionality
http://proceedings.mlr.press/v139/lacotte21a.html
AUTHORS: Jonathan Lacotte, Yifei Wang, Mert Pilanci
HIGHLIGHT: We propose a randomized algorithm with quadratic convergence rate for convex optimization problems with a self-concordant, composite, strongly convex objective function.

542, TITLE: Generalization Bounds in the Presence of Outliers: a Median-of-Means Study
http://proceedings.mlr.press/v139/laforgue21a.html
AUTHORS: Pierre Laforgue, Guillaume Staerman, Stephan Cl?men?on
HIGHLIGHT: In this context, the present work proposes a general study of MoM's concentration properties under the contamination regime, that provides a clear understanding on the impact of the outlier proportion and the number of blocks chosen.

543, TITLE: Model Fusion for Personalized Learning
http://proceedings.mlr.press/v139/lam21a.html
AUTHORS: Thanh Chi Lam, Nghia Hoang, Bryan Kian Hsiang Low, Patrick Jaillet
HIGHLIGHT: To accommodate for such scenarios, we develop a new personalized learning framework that synthesizes customized models for unseen tasks via fusion of independently pre-trained models of related tasks.

544, TITLE: Gradient Disaggregation: Breaking Privacy in Federated Learning by Reconstructing the User Participant Matrix
http://proceedings.mlr.press/v139/lam21b.html
AUTHORS: Maximilian Lam, Gu-Yeon Wei, David Brooks, Vijay Janapa Reddi, Michael Mitzenmacher
HIGHLIGHT: Our method revolves around reconstructing participant information (e.g: which rounds of training users participated in) from aggregated model updates by leveraging summary information from device analytics commonly used to monitor, debug, and manage federated learning systems.

545, TITLE: Stochastic Multi-Armed Bandits with Unrestricted Delay Distributions
http://proceedings.mlr.press/v139/lancewicki21a.html
AUTHORS: Tal Lancewicki, Shahar Segal, Tomer Koren, Yishay Mansour
HIGHLIGHT: Our main contribution is algorithms that achieve near-optimal regret in each of the settings, with an additional additive dependence on the quantiles of the delay distribution.

546, TITLE: Discovering symbolic policies with deep reinforcement learning
http://proceedings.mlr.press/v139/landajuela21a.html
AUTHORS: Mikel Landajuela, Brenden K Petersen, Sookyung Kim, Claudio P Santiago, Ruben Glatt, Nathan Mundhenk, Jacob F Pettit, Daniel Faissol
HIGHLIGHT: To this end, we propose deep symbolic policy, a novel approach to directly search the space of symbolic policies.

547, TITLE: Graph Cuts Always Find a Global Optimum for Potts Models (With a Catch)
http://proceedings.mlr.press/v139/lang21a.html
AUTHORS: Hunter Lang, David Sontag, Aravindan Vijayaraghavan
HIGHLIGHT: We prove that the alpha-expansion algorithm for MAP inference always returns a globally optimal assignment for Markov Random Fields with Potts pairwise potentials, with a catch: the returned assignment is only guaranteed to be optimal for an instance within a small perturbation of the original problem instance.

548, TITLE: Efficient Message Passing for 0-1 ILPs with Binary Decision Diagrams
http://proceedings.mlr.press/v139/lange21a.html
AUTHORS: Jan-Hendrik Lange, Paul Swoboda
HIGHLIGHT: We present a message passing method for 0{-}1 integer linear programs.

549, TITLE: CountSketches, Feature Hashing and the Median of Three
http://proceedings.mlr.press/v139/larsen21a.html
AUTHORS: Kasper Green Larsen, Rasmus Pagh, Jakub Tetek
HIGHLIGHT: In this paper, we revisit the classic CountSketch method, which is a sparse, random projection that transforms a (high-dimensional) Euclidean vector v to a vector of dimension $(2t-1) s$, where $t, s > 0$ are integer parameters.

550, TITLE: MorphVAE: Generating Neural Morphologies from 3D-Walks using a Variational Autoencoder with Spherical Latent Space
http://proceedings.mlr.press/v139/laturnus21a.html
AUTHORS: Sophie C. Laturnus, Philipp Berens
HIGHLIGHT: Here we propose MorphVAE, a sequence-to-sequence variational autoencoder with spherical latent space as a generative model for neural morphologies.

551, TITLE: Improved Regret Bound and Experience Replay in Regularized Policy Iteration
http://proceedings.mlr.press/v139/lazic21a.html
AUTHORS: Nevena Lazic, Dong Yin, Yasin Abbasi-Yadkori, Csaba Szepesvari
HIGHLIGHT: In this work, we study algorithms for learning in infinite-horizon undiscounted Markov decision processes (MDPs) with function approximation.

552, TITLE: LAMDA: Label Matching Deep Domain Adaptation
http://proceedings.mlr.press/v139/le21a.html
AUTHORS: Trung Le, Tuan Nguyen, Nhat Ho, Hung Bui, Dinh Phung
HIGHLIGHT: In this paper, we propose and study a new challenging setting that allows us to use a Wasserstein distance (WS) to not only quantify the data shift but also to define the label shift directly.

553, TITLE: Gaussian Process-Based Real-Time Learning for Safety Critical Applications
http://proceedings.mlr.press/v139/lederer21a.html
AUTHORS: Armin Lederer, Alejandro J Ord??ez Conejo, Korbinian A Maier, Wenxin Xiao, Jonas Umlauft, Sandra Hirche
HIGHLIGHT: Due to its high computational complexity, Gaussian process regression must be used offline on batches of data, which prevents applications, where a fast adaptation through online learning is necessary to ensure safety. In order to overcome this issue, we propose the LoG-GP.

554, TITLE: Sharing Less is More: Lifelong Learning in Deep Networks with Selective Layer Transfer
http://proceedings.mlr.press/v139/lee21a.html
AUTHORS: Seungwon Lee, Sima Behpour, Eric Eaton
HIGHLIGHT: We first show that the lifelong learning performance of several current deep learning architectures can be significantly improved by transfer at the appropriate layers. We then develop an expectation-maximization (EM) method to automatically select the appropriate transfer configuration and optimize the task network weights.

555, TITLE: Fair Selective Classification Via Sufficiency
http://proceedings.mlr.press/v139/lee21b.html
AUTHORS: Joshua K Lee, Yuheng Bu, Deepta Rajan, Prasanna Sattigeri, Rameswar Panda, Subhro Das, Gregory W Wornell
HIGHLIGHT: We prove that the sufficiency criterion can be used to mitigate these disparities by ensuring that selective classification increases performance on all groups, and introduce a method for mitigating the disparity in precision across the entire coverage scale based on this criterion.

556, TITLE: On-the-fly Rectification for Robust Large-Vocabulary Topic Inference
http://proceedings.mlr.press/v139/lee21c.html
AUTHORS: Moontae Lee, Sungjun Cho, Kun Dong, David Mimno, David Bindel
HIGHLIGHT: We propose novel methods that simultaneously compress and rectify co-occurrence statistics, scaling gracefully with the size of vocabulary and the dimension of latent space.

557, TITLE: Unsupervised Embedding Adaptation via Early-Stage Feature Reconstruction for Few-Shot Classification
http://proceedings.mlr.press/v139/lee21d.html
AUTHORS: Dong Hoon Lee, Sae-Young Chung
HIGHLIGHT: We propose unsupervised embedding adaptation for the downstream few-shot classification task.

558, TITLE: Continual Learning in the Teacher-Student Setup: Impact of Task Similarity
http://proceedings.mlr.press/v139/lee21e.html

AUTHORS: Sebastian Lee, Sebastian Goldt, Andrew Saxe
HIGHLIGHT: Here, we attempt to narrow this gap between theory and practice by studying continual learning in the teacher-student setup.

559, TITLE: OptiDICE: Offline Policy Optimization via Stationary Distribution Correction Estimation
http://proceedings.mlr.press/v139/lee21f.html
AUTHORS: Jongmin Lee, Wonseok Jeon, Byungjun Lee, Joelle Pineau, Kee-Eung Kim
HIGHLIGHT: In this paper, we present an offline RL algorithm that prevents overestimation in a more principled way.

560, TITLE: SUNRISE: A Simple Unified Framework for Ensemble Learning in Deep Reinforcement Learning
http://proceedings.mlr.press/v139/lee21g.html
AUTHORS: Kimin Lee, Michael Laskin, Aravind Srinivas, Pieter Abbeel
HIGHLIGHT: To mitigate these issues, we present SUNRISE, a simple unified ensemble method, which is compatible with various off-policy RL algorithms.

561, TITLE: Achieving Near Instance-Optimality and Minimax-Optimality in Stochastic and Adversarial Linear Bandits Simultaneously
http://proceedings.mlr.press/v139/lee21h.html
AUTHORS: Chung-Wei Lee, Haipeng Luo, Chen-Yu Wei, Mengxiao Zhang, Xiaojin Zhang
HIGHLIGHT: In this work, we develop linear bandit algorithms that automatically adapt to different environments.

562, TITLE: PEBBLE: Feedback-Efficient Interactive Reinforcement Learning via Relabeling Experience and Unsupervised Pre-training
http://proceedings.mlr.press/v139/lee21i.html
AUTHORS: Kimin Lee, Laura M Smith, Pieter Abbeel
HIGHLIGHT: We present an off-policy, interactive RL algorithm that capitalizes on the strengths of both feedback and off-policy learning.

563, TITLE: Near-Optimal Linear Regression under Distribution Shift
http://proceedings.mlr.press/v139/lei21a.html
AUTHORS: Qi Lei, Wei Hu, Jason Lee
HIGHLIGHT: We develop estimators that achieve minimax linear risk for linear regression problems under distribution shift.

564, TITLE: Stability and Generalization of Stochastic Gradient Methods for Minimax Problems
http://proceedings.mlr.press/v139/lei21b.html
AUTHORS: Yunwen Lei, Zhenhuan Yang, Tianbao Yang, Yiming Ying
HIGHLIGHT: In this paper, we provide a comprehensive generalization analysis of stochastic gradient methods for minimax problems under both convex-concave and nonconvex-nonconcave cases through the lens of algorithmic stability.

565, TITLE: Scalable Evaluation of Multi-Agent Reinforcement Learning with Melting Pot
http://proceedings.mlr.press/v139/leibo21a.html
AUTHORS: Joel Z Leibo, Edgar A Due?ez-Guzman, Alexander Vezhnevets, John P Agapiou, Peter Sunehag, Raphael Koster, Jayd Matyas, Charlie Beattie, Igor Mordatch, Thore Graepel
HIGHLIGHT: Our contribution, Melting Pot, is a MARL evaluation suite that fills this gap and uses reinforcement learning to reduce the human labor required to create novel test scenarios.

566, TITLE: Better Training using Weight-Constrained Stochastic Dynamics
http://proceedings.mlr.press/v139/leimkuhler21a.html
AUTHORS: Benedict Leimkuhler, Tiffany J Vlaar, Timoth?e Pouchon, Amos Storkey
HIGHLIGHT: We provide a general approach to efficiently incorporate constraints into a stochastic gradient Langevin framework, allowing enhanced exploration of the loss landscape.

567, TITLE: Globally-Robust Neural Networks
http://proceedings.mlr.press/v139/leino21a.html
AUTHORS: Klas Leino, Zifan Wang, Matt Fredrikson
HIGHLIGHT: We show that widely-used architectures can be easily adapted to this objective by incorporating efficient global Lipschitz bounds into the network, yielding certifiably-robust models by construction that achieve state-of-the-art verifiable accuracy.

568, TITLE: Learning to Price Against a Moving Target
http://proceedings.mlr.press/v139/leme21a.html

AUTHORS: Renato Paes Leme, Balasubramanian Sivan, Yifeng Teng, Pratik Worah
HIGHLIGHT: Here we study the problem where the buyer's value is a moving target, i.e., they change over time either by a stochastic process or adversarially with bounded variation.

569, TITLE: SigGPDE: Scaling Sparse Gaussian Processes on Sequential Data
http://proceedings.mlr.press/v139/lemercier21a.html
AUTHORS: Maud Lemercier, Cristopher Salvi, Thomas Cass, Edwin V. Bonilla, Theodoros Damoulas, Terry J Lyons
HIGHLIGHT: We develop SigGPDE, a new scalable sparse variational inference framework for Gaussian Processes (GPs) on sequential data.

570, TITLE: Strategic Classification Made Practical
http://proceedings.mlr.press/v139/levanon21a.html
AUTHORS: Sagi Levanon, Nir Rosenfeld
HIGHLIGHT: In this paper we present a learning framework for strategic classification that is practical.

571, TITLE: Improved, Deterministic Smoothing for L_1 Certified Robustness
http://proceedings.mlr.press/v139/levine21a.html
AUTHORS: Alexander J Levine, Soheil Feizi
HIGHLIGHT: In this work, we propose a non-additive and deterministic smoothing method, Deterministic Smoothing with Splitting Noise (DSSN).

572, TITLE: BASE Layers: Simplifying Training of Large, Sparse Models
http://proceedings.mlr.press/v139/lewis21a.html
AUTHORS: Mike Lewis, Shruti Bhosale, Tim Dettmers, Naman Goyal, Luke Zettlemoyer
HIGHLIGHT: We introduce a new balanced assignment of experts (BASE) layer for large language models that greatly simplifies existing high capacity sparse layers.

573, TITLE: Run-Sort-ReRun: Escaping Batch Size Limitations in Sliced Wasserstein Generative Models
http://proceedings.mlr.press/v139/lezama21a.html
AUTHORS: Jose Lezama, Wei Chen, Qiang Qiu
HIGHLIGHT: In this paper, we build upon recent progress in sliced Wasserstein distances, a family of differentiable metrics for distribution discrepancy based on the Optimal Transport paradigm.

574, TITLE: PAGE: A Simple and Optimal Probabilistic Gradient Estimator for Nonconvex Optimization
http://proceedings.mlr.press/v139/li21a.html
AUTHORS: Zhize Li, Hongyan Bao, Xiangliang Zhang, Peter Richtarik
HIGHLIGHT: In this paper, we propose a novel stochastic gradient estimator—ProbAbilistic Gradient Estimator (PAGE)—for nonconvex optimization.

575, TITLE: Tightening the Dependence on Horizon in the Sample Complexity of Q-Learning
http://proceedings.mlr.press/v139/li21b.html
AUTHORS: Gen Li, Changxiao Cai, Yuxin Chen, Yuantao Gu, Yuting Wei, Yuejie Chi
HIGHLIGHT: In this work, we sharpen the sample complexity of synchronous Q-learning to the order of $\frac{|S||A|}{(1-\gamma)^4\varepsilon^2}$ (up to some logarithmic factor) for any $0<\varepsilon <1$, leading to an order-wise improvement in $\frac{1}{1-\gamma}$.

576, TITLE: Winograd Algorithm for AdderNet
http://proceedings.mlr.press/v139/li21c.html
AUTHORS: Wenshuo Li, Hanting Chen, Mingqiang Huang, Xinghao Chen, Chunjing Xu, Yunhe Wang
HIGHLIGHT: To further optimize the hardware overhead of using AdderNet, this paper studies the winograd algorithm, which is a widely used fast algorithm for accelerating convolution and saving the computational costs.

577, TITLE: A Free Lunch From ANN: Towards Efficient, Accurate Spiking Neural Networks Calibration
http://proceedings.mlr.press/v139/li21d.html
AUTHORS: Yuhang Li, Shikuang Deng, Xin Dong, Ruihao Gong, Shi Gu
HIGHLIGHT: We introduce SNN Calibration, a cheap but extraordinarily effective method by leveraging the knowledge within a pre-trained Artificial Neural Network (ANN).

578, TITLE: Privacy-Preserving Feature Selection with Secure Multiparty Computation
http://proceedings.mlr.press/v139/li21e.html

AUTHORS: Xiling Li, Rafael Dowsley, Martine De Cock
HIGHLIGHT: In this work, we propose the first MPC based protocol for private feature selection based on the filter method, which is independent of model training, and can be used in combination with any MPC protocol to rank features.

579, TITLE: Theory of Spectral Method for Union of Subspaces-Based Random Geometry Graph
http://proceedings.mlr.press/v139/li21f.html
AUTHORS: Gen Li, Yuantao Gu
HIGHLIGHT: This paper establishes a theory to show the power of this method for the first time, in which we demonstrate the mechanism of spectral clustering by analyzing a simplified algorithm under the widely used semi-random model.

580, TITLE: MURAL: Meta-Learning Uncertainty-Aware Rewards for Outcome-Driven Reinforcement Learning
http://proceedings.mlr.press/v139/li21g.html
AUTHORS: Kevin Li, Abhishek Gupta, Ashwin Reddy, Vitchyr H Pong, Aurick Zhou, Justin Yu, Sergey Levine
HIGHLIGHT: In this work, we study a more tractable class of reinforcement learning problems defined simply by examples of successful outcome states, which can be much easier to provide while still making the exploration problem more tractable.

581, TITLE: Ditto: Fair and Robust Federated Learning Through Personalization
http://proceedings.mlr.press/v139/li21h.html
AUTHORS: Tian Li, Shengyuan Hu, Ahmad Beirami, Virginia Smith
HIGHLIGHT: In this work, we identify that robustness to data and model poisoning attacks and fairness, measured as the uniformity of performance across devices, are competing constraints in statistically heterogeneous networks.

582, TITLE: Quantization Algorithms for Random Fourier Features
http://proceedings.mlr.press/v139/li21i.html
AUTHORS: Xiaoyun Li, Ping Li
HIGHLIGHT: In this paper, we focus on developing quantization algorithms for RFF.

583, TITLE: Approximate Group Fairness for Clustering
http://proceedings.mlr.press/v139/li21j.html
AUTHORS: Bo Li, Lijun Li, Ankang Sun, Chenhao Wang, Yingfan Wang
HIGHLIGHT: Particularly, we propose two dimensions to relax core requirements: one is on the degree of distance improvement, and the other is on the size of deviating coalition.

584, TITLE: Sharper Generalization Bounds for Clustering
http://proceedings.mlr.press/v139/li21k.html
AUTHORS: Shaojie Li, Yong Liu
HIGHLIGHT: In this paper, we propose a unified clustering learning framework and investigate its excess risk bounds, obtaining state-of-the-art upper bounds under mild assumptions.

585, TITLE: Provably End-to-end Label-noise Learning without Anchor Points
http://proceedings.mlr.press/v139/li21l.html
AUTHORS: Xuefeng Li, Tongliang Liu, Bo Han, Gang Niu, Masashi Sugiyama
HIGHLIGHT: In this paper, we propose an end-to-end framework for solving label-noise learning without anchor points, in which we simultaneously optimize two objectives: the cross entropy loss between the noisy label and the predicted probability by the neural network, and the volume of the simplex formed by the columns of the transition matrix.

586, TITLE: A Novel Method to Solve Neural Knapsack Problems
http://proceedings.mlr.press/v139/li21m.html
AUTHORS: Duanshun Li, Jing Liu, Dongeun Lee, Ali Seyedmazloom, Giridhar Kaushik, Kookjin Lee, Noseong Park
HIGHLIGHT: In this paper, we present a game-theoretic method to solve 0-1 knapsack problems (KPs) where the number of items (products) is large and the values of items are not predetermined but decided by an external value assignment function (e.g., a neural network in our case) during the optimization process.

587, TITLE: Mixed Cross Entropy Loss for Neural Machine Translation
http://proceedings.mlr.press/v139/li21n.html
AUTHORS: Haoran Li, Wei Lu
HIGHLIGHT: In this paper, we propose mixed Cross Entropy loss (mixed CE) as a substitute for CE in both training approaches.

588, TITLE: Training Graph Neural Networks with 1000 Layers

http://proceedings.mlr.press/v139/li21o.html
AUTHORS: Guohao Li, Matthias M?ller, Bernard Ghanem, Vladlen Koltun
HIGHLIGHT: In this work, we study reversible connections, group convolutions, weight tying, and equilibrium models to advance the memory and parameter efficiency of GNNs.

589, TITLE: Active Feature Acquisition with Generative Surrogate Models
http://proceedings.mlr.press/v139/li21p.html
AUTHORS: Yang Li, Junier Oliva
HIGHLIGHT: In this work, we consider models that perform active feature acquisition (AFA) and query the environment for unobserved features to improve the prediction assessments at evaluation time.

590, TITLE: Partially Observed Exchangeable Modeling
http://proceedings.mlr.press/v139/li21q.html
AUTHORS: Yang Li, Junier Oliva
HIGHLIGHT: In this work, we propose a novel framework, partially observed exchangeable modeling (POEx) that takes in a set of related partially observed instances and infers the conditional distribution for the unobserved dimensions over multiple elements.

591, TITLE: Testing DNN-based Autonomous Driving Systems under Critical Environmental Conditions
http://proceedings.mlr.press/v139/li21r.html
AUTHORS: Zhong Li, Minxue Pan, Tian Zhang, Xuandong Li
HIGHLIGHT: In this paper, we propose to test DNN-based ADS under different environmental conditions to identify the critical ones, that is, the environmental conditions under which the ADS are more prone to errors.

592, TITLE: The Symmetry between Arms and Knapsacks: A Primal-Dual Approach for Bandits with Knapsacks
http://proceedings.mlr.press/v139/li21s.html
AUTHORS: Xiaocheng Li, Chunlin Sun, Yinyu Ye
HIGHLIGHT: In this paper, we study the bandits with knapsacks (BwK) problem and develop a primal-dual based algorithm that achieves a problem-dependent logarithmic regret bound.

593, TITLE: Distributionally Robust Optimization with Markovian Data
http://proceedings.mlr.press/v139/li21t.html
AUTHORS: Mengmeng Li, Tobias Sutter, Daniel Kuhn
HIGHLIGHT: We propose a data-driven distributionally robust optimization model to estimate the problem’s objective function and optimal solution.

594, TITLE: Communication-Efficient Distributed SVD via Local Power Iterations
http://proceedings.mlr.press/v139/li21u.html
AUTHORS: Xiang Li, Shusen Wang, Kun Chen, Zhihua Zhang
HIGHLIGHT: In the aggregation, we propose to weight each local eigenvector matrix with orthogonal Procrustes transformation (OPT).

595, TITLE: FILTRA: Rethinking Steerable CNN by Filter Transform
http://proceedings.mlr.press/v139/li21v.html
AUTHORS: Bo Li, Qili Wang, Gim Hee Lee
HIGHLIGHT: In this paper, we show that kernel constructed by filter transform can also be interpreted in the group representation theory.

596, TITLE: Online Unrelated Machine Load Balancing with Predictions Revisited
http://proceedings.mlr.press/v139/li21w.html
AUTHORS: Shi Li, Jiayi Xian
HIGHLIGHT: We study the online load balancing problem with machine learned predictions, and give results that improve upon and extend those in a recent paper by Lattanzi et al. (2020).

597, TITLE: Asymptotic Normality and Confidence Intervals for Prediction Risk of the Min-Norm Least Squares Estimator
http://proceedings.mlr.press/v139/li21x.html
AUTHORS: Zeng Li, Chuanlong Xie, Qinwen Wang
HIGHLIGHT: This paper quantifies the uncertainty of prediction risk for the min-norm least squares estimator in high-dimensional linear regression models.

598, TITLE: TeraPipe: Token-Level Pipeline Parallelism for Training Large-Scale Language Models
http://proceedings.mlr.press/v139/li21y.html
AUTHORS: Zhuohan Li, Siyuan Zhuang, Shiyuan Guo, Danyang Zhuo, Hao Zhang, Dawn Song, Ion Stoica
HIGHLIGHT: In this work, we identify a new and orthogonal dimension from existing model parallel approaches: it is possible to perform pipeline parallelism within a single training sequence for Transformer-based language models thanks to its autoregressive property.

599, TITLE: A Second look at Exponential and Cosine Step Sizes: Simplicity, Adaptivity, and Performance
http://proceedings.mlr.press/v139/li21z.html
AUTHORS: Xiaoyu Li, Zhenxun Zhuang, Francesco Orabona
HIGHLIGHT: In this paper, we study two step size schedules whose power has been repeatedly confirmed in practice: the exponential and the cosine step sizes.

600, TITLE: Towards Understanding and Mitigating Social Biases in Language Models
http://proceedings.mlr.press/v139/liang21a.html
AUTHORS: Paul Pu Liang, Chiyu Wu, Louis-Philippe Morency, Ruslan Salakhutdinov
HIGHLIGHT: With these tools, we propose steps towards mitigating social biases during text generation.

601, TITLE: Uncovering the Connections Between Adversarial Transferability and Knowledge Transferability
http://proceedings.mlr.press/v139/liang21b.html
AUTHORS: Kaizhao Liang, Jacky Y Zhang, Boxin Wang, Zhuolin Yang, Sanmi Koyejo, Bo Li
HIGHLIGHT: In this paper, as the first work, we analyze and demonstrate the connections between knowledge transferability and another important phenomenon–adversarial transferability, \emph{i.e.}, adversarial examples generated against one model can be transferred to attack other models.

602, TITLE: Parallel Droplet Control in MEDA Biochips using Multi-Agent Reinforcement Learning
http://proceedings.mlr.press/v139/liang21c.html
AUTHORS: Tung-Che Liang, Jin Zhou, Yun-Sheng Chan, Tsung-Yi Ho, Krishnendu Chakrabarty, Cy Lee
HIGHLIGHT: To overcome these problems, we present a multi-agent reinforcement learning (MARL) droplet-routing solution that can be used for various sizes of MEDA biochips with integrated sensors, and we demonstrate the reliable execution of a serial-dilution bioassay with the MARL droplet router on a fabricated MEDA biochip.

603, TITLE: Information Obfuscation of Graph Neural Networks
http://proceedings.mlr.press/v139/liao21a.html
AUTHORS: Peiyuan Liao, Han Zhao, Keyulu Xu, Tommi Jaakkola, Geoffrey J. Gordon, Stefanie Jegelka, Ruslan Salakhutdinov
HIGHLIGHT: In this paper, we study the problem of protecting sensitive attributes by information obfuscation when learning with graph structured data.

604, TITLE: Guided Exploration with Proximal Policy Optimization using a Single Demonstration
http://proceedings.mlr.press/v139/libardi21a.html
AUTHORS: Gabriele Libardi, Gianni De Fabritiis, Sebastian Dittert
HIGHLIGHT: We train an agent on a combination of demonstrations and own experience to solve problems with variable initial conditions and we integrate it with proximal policy optimization (PPO).

605, TITLE: Debiasing a First-order Heuristic for Approximate Bi-level Optimization
http://proceedings.mlr.press/v139/likhosherstov21a.html
AUTHORS: Valerii Likhosherstov, Xingyou Song, Krzysztof Choromanski, Jared Q Davis, Adrian Weller
HIGHLIGHT: We contribute by theoretically characterizing FOM's gradient bias under mild assumptions.

606, TITLE: Making transport more robust and interpretable by moving data through a small number of anchor points
http://proceedings.mlr.press/v139/lin21a.html
AUTHORS: Chi-Heng Lin, Mehdi Azabou, Eva Dyer
HIGHLIGHT: Here, we introduce Latent Optimal Transport (LOT), a new approach for OT that simultaneously learns low-dimensional structure in data while leveraging this structure to solve the alignment task.

607, TITLE: Straight to the Gradient: Learning to Use Novel Tokens for Neural Text Generation
http://proceedings.mlr.press/v139/lin21b.html
AUTHORS: Xiang Lin, Simeng Han, Shafiq Joty
HIGHLIGHT: In this work, we introduce ScaleGrad, a modification straight to the gradient of the loss function, to remedy the degeneration issue of the standard MLE objective.

608, TITLE: Quasi-global Momentum: Accelerating Decentralized Deep Learning on Heterogeneous Data
http://proceedings.mlr.press/v139/lin21c.html
AUTHORS: Tao Lin, Sai Praneeth Karimireddy, Sebastian Stich, Martin Jaggi
HIGHLIGHT: In this paper, we investigate and identify the limitation of several decentralized optimization algorithms for different degrees of data heterogeneity.

609, TITLE: Generative Causal Explanations for Graph Neural Networks
http://proceedings.mlr.press/v139/lin21d.html
AUTHORS: Wanyu Lin, Hao Lan, Baochun Li
HIGHLIGHT: This paper presents {\em Gem}, a model-agnostic approach for providing interpretable explanations for any GNNs on various graph learning tasks.

610, TITLE: Tractable structured natural-gradient descent using local parameterizations
http://proceedings.mlr.press/v139/lin21e.html
AUTHORS: Wu Lin, Frank Nielsen, Khan Mohammad Emtiyaz, Mark Schmidt
HIGHLIGHT: We address this issue by using \emph{local-parameter coordinates} to obtain a flexible and efficient NGD method that works well for a wide-variety of structured parameterizations.

611, TITLE: Active Learning of Continuous-time Bayesian Networks through Interventions
http://proceedings.mlr.press/v139/linzner21a.html
AUTHORS: Dominik Linzner, Heinz Koeppl
HIGHLIGHT: We propose a novel criterion for experimental design based on a variational approximation of the expected information gain.

612, TITLE: Phase Transitions, Distance Functions, and Implicit Neural Representations
http://proceedings.mlr.press/v139/lipman21a.html
AUTHORS: Yaron Lipman
HIGHLIGHT: In this paper we draw inspiration from the theory of phase transitions of fluids and suggest a loss for training INRs that learns a density function that converges to a proper occupancy function, while its log transform converges to a distance function.

613, TITLE: The Earth Mover's Pinball Loss: Quantiles for Histogram-Valued Regression
http://proceedings.mlr.press/v139/list21a.html
AUTHORS: Florian List
HIGHLIGHT: We present a dedicated method for Deep Learning-based histogram regression, which incorporates cross-bin information and yields distributions over possible histograms, expressed by τ-quantiles of the cumulative histogram in each bin.

614, TITLE: Understanding Instance-Level Label Noise: Disparate Impacts and Treatments
http://proceedings.mlr.press/v139/liu21a.html
AUTHORS: Yang Liu
HIGHLIGHT: This paper aims to provide understandings for the effect of an over-parameterized model, e.g. a deep neural network, memorizing instance-dependent noisy labels.

615, TITLE: APS: Active Pretraining with Successor Features
http://proceedings.mlr.press/v139/liu21b.html
AUTHORS: Hao Liu, Pieter Abbeel
HIGHLIGHT: We introduce a new unsupervised pretraining objective for reinforcement learning.

616, TITLE: Learning by Turning: Neural Architecture Aware Optimisation
http://proceedings.mlr.press/v139/liu21c.html
AUTHORS: Yang Liu, Jeremy Bernstein, Markus Meister, Yisong Yue
HIGHLIGHT: To address this problem, this paper conducts a combined study of neural architecture and optimisation, leading to a new optimiser called Nero: the neuronal rotator.

617, TITLE: Dynamic Game Theoretic Neural Optimizer
http://proceedings.mlr.press/v139/liu21d.html
AUTHORS: Guan-Horng Liu, Tianrong Chen, Evangelos Theodorou
HIGHLIGHT: In this work, we propose a novel dynamic game perspective by viewing each layer as a player in a dynamic game characterized by the DNN itself.

618, TITLE: Besov Function Approximation and Binary Classification on Low-Dimensional Manifolds Using Convolutional Residual Networks
http://proceedings.mlr.press/v139/liu21e.html
AUTHORS: Hao Liu, Minshuo Chen, Tuo Zhao, Wenjing Liao
HIGHLIGHT: To bridge this gap, we propose to exploit the low-dimensional structures of the real world datasets and establish theoretical guarantees of convolutional residual networks (ConvResNet) in terms of function approximation and statistical recovery for binary classification problem.

619, TITLE: Just Train Twice: Improving Group Robustness without Training Group Information
http://proceedings.mlr.press/v139/liu21f.html
AUTHORS: Evan Z Liu, Behzad Haghgoo, Annie S Chen, Aditi Raghunathan, Pang Wei Koh, Shiori Sagawa, Percy Liang, Chelsea Finn
HIGHLIGHT: In this paper, we propose a simple two-stage approach, JTT, that achieves comparable performance to group DRO while only requiring group annotations on a significantly smaller validation set.

620, TITLE: Event Outlier Detection in Continuous Time
http://proceedings.mlr.press/v139/liu21g.html
AUTHORS: Siqi Liu, Milos Hauskrecht
HIGHLIGHT: In this work, we study and develop methods for detecting outliers in continuous-time event sequences, including unexpected absence and unexpected occurrences of events.

621, TITLE: Heterogeneous Risk Minimization
http://proceedings.mlr.press/v139/liu21h.html
AUTHORS: Jiashuo Liu, Zheyuan Hu, Peng Cui, Bo Li, Zheyan Shen
HIGHLIGHT: In this paper, we propose Heterogeneous Risk Minimization (HRM) framework to achieve joint learning of latent heterogeneity among the data and invariant relationship, which leads to stable prediction despite distributional shifts.

622, TITLE: Stochastic Iterative Graph Matching
http://proceedings.mlr.press/v139/liu21i.html
AUTHORS: Linfeng Liu, Michael C Hughes, Soha Hassoun, Liping Liu
HIGHLIGHT: Considering that model outputs are complex matchings, we devise several techniques to improve the learning of GNNs and obtain a new model, Stochastic Iterative Graph MAtching (SIGMA).

623, TITLE: Cooperative Exploration for Multi-Agent Deep Reinforcement Learning
http://proceedings.mlr.press/v139/liu21j.html
AUTHORS: Iou-Jen Liu, Unnat Jain, Raymond A Yeh, Alexander Schwing
HIGHLIGHT: To address this shortcoming, in this paper, we propose cooperative multi-agent exploration (CMAE): agents share a common goal while exploring.

624, TITLE: Elastic Graph Neural Networks
http://proceedings.mlr.press/v139/liu21k.html
AUTHORS: Xiaorui Liu, Wei Jin, Yao Ma, Yaxin Li, Hua Liu, Yiqi Wang, Ming Yan, Jiliang Tang
HIGHLIGHT: In particular, we propose a novel and general message passing scheme into GNNs.

625, TITLE: One Pass Late Fusion Multi-view Clustering
http://proceedings.mlr.press/v139/liu21l.html
AUTHORS: Xinwang Liu, Li Liu, Qing Liao, Siwei Wang, Yi Zhang, Wenxuan Tu, Chang Tang, Jiyuan Liu, En Zhu
HIGHLIGHT: To address this issue, we propose to unify the aforementioned two learning procedures into a single optimization, in which the consensus partition matrix can better serve for the generation of cluster labels, and the latter is able to guide the learning of the former.

626, TITLE: Coach-Player Multi-agent Reinforcement Learning for Dynamic Team Composition
http://proceedings.mlr.press/v139/liu21m.html
AUTHORS: Bo Liu, Qiang Liu, Peter Stone, Animesh Garg, Yuke Zhu, Anima Anandkumar
HIGHLIGHT: Coordinating teams with such dynamic composition is challenging: the optimal team strategy varies with the composition. We propose COPA, a coach-player framework to tackle this problem.

627, TITLE: From Local to Global Norm Emergence: Dissolving Self-reinforcing Substructures with Incremental Social Instruments

http://proceedings.mlr.press/v139/liu21n.html
AUTHORS: Yiwei Liu, Jiamou Liu, Kaibin Wan, Zhan Qin, Zijian Zhang, Bakhadyr Khoussainov, Liehuang Zhu
HIGHLIGHT: We propose incremental social instruments (ISI) to dissolve these SRSs by creating ties between agents.

628, TITLE: A Value-Function-based Interior-point Method for Non-convex Bi-level Optimization
http://proceedings.mlr.press/v139/liu21o.html
AUTHORS: Risheng Liu, Xuan Liu, Xiaoming Yuan, Shangzhi Zeng, Jin Zhang
HIGHLIGHT: In this work, we propose a new gradient-based solution scheme, namely, the Bi-level Value-Function-based Interior-point Method (BVFIM).

629, TITLE: Selfish Sparse RNN Training
http://proceedings.mlr.press/v139/liu21p.html
AUTHORS: Shiwei Liu, Decebal Constantin Mocanu, Yulong Pei, Mykola Pechenizkiy
HIGHLIGHT: In this paper, we propose an approach to train intrinsically sparse RNNs with a fixed parameter count in one single run, without compromising performance.

630, TITLE: Temporal Difference Learning as Gradient Splitting
http://proceedings.mlr.press/v139/liu21q.html
AUTHORS: Rui Liu, Alex Olshevsky
HIGHLIGHT: We provide an interpretation of this method in terms of a splitting of the gradient of an appropriately chosen function.

631, TITLE: On Robust Mean Estimation under Coordinate-level Corruption
http://proceedings.mlr.press/v139/liu21r.html
AUTHORS: Zifan Liu, Jong Ho Park, Theodoros Rekatsinas, Christos Tzamos
HIGHLIGHT: We study the problem of robust mean estimation and introduce a novel Hamming distance-based measure of distribution shift for coordinate-level corruptions.

632, TITLE: Decoupling Exploration and Exploitation for Meta-Reinforcement Learning without Sacrifices
http://proceedings.mlr.press/v139/liu21s.html
AUTHORS: Evan Z Liu, Aditi Raghunathan, Percy Liang, Chelsea Finn
HIGHLIGHT: We alleviate both concerns by constructing an exploitation objective that automatically identifies task-relevant information and an exploration objective to recover only this information.

633, TITLE: How Do Adam and Training Strategies Help BNNs Optimization
http://proceedings.mlr.press/v139/liu21t.html
AUTHORS: Zechun Liu, Zhiqiang Shen, Shichao Li, Koen Helwegen, Dong Huang, Kwang-Ting Cheng
HIGHLIGHT: To address this, in this paper we first investigate the trajectories of gradients and weights in BNNs during the training process.

634, TITLE: SagaNet: A Small Sample Gated Network for Pediatric Cancer Diagnosis
http://proceedings.mlr.press/v139/liu21u.html
AUTHORS: Yuhan Liu, Shiliang Sun
HIGHLIGHT: In this work, we propose a novel model to solve the diagnosis task of small round blue cell tumors (SRBCTs).

635, TITLE: Learning Deep Neural Networks under Agnostic Corrupted Supervision
http://proceedings.mlr.press/v139/liu21v.html
AUTHORS: Boyang Liu, Mengying Sun, Ding Wang, Pang-Ning Tan, Jiayu Zhou
HIGHLIGHT: To alleviate this problem, we present an efficient robust algorithm that achieves strong guarantees without any assumption on the type of corruption and provides a unified framework for both classification and regression problems.

636, TITLE: Leveraging Public Data for Practical Private Query Release
http://proceedings.mlr.press/v139/liu21w.html
AUTHORS: Terrance Liu, Giuseppe Vietri, Thomas Steinke, Jonathan Ullman, Steven Wu
HIGHLIGHT: With the goal of releasing statistics about a private dataset, we present PMW^Pub, which—unlike existing baselines—leverages public data drawn from a related distribution as prior information.

637, TITLE: Watermarking Deep Neural Networks with Greedy Residuals
http://proceedings.mlr.press/v139/liu21x.html
AUTHORS: Hanwen Liu, Zhenyu Weng, Yuesheng Zhu

HIGHLIGHT: In this paper, we propose a novel watermark-based ownership protection method by using the residuals of important parameters.

638, TITLE: Do We Actually Need Dense Over-Parameterization? In-Time Over-Parameterization in Sparse Training
http://proceedings.mlr.press/v139/liu21y.html
AUTHORS: Shiwei Liu, Lu Yin, Decebal Constantin Mocanu, Mykola Pechenizkiy
HIGHLIGHT: In this paper, we introduce a new perspective on training deep neural networks capable of state-of-the-art performance without the need for the expensive over-parameterization by proposing the concept of In-Time Over-Parameterization (ITOP) in sparse training.

639, TITLE: A Sharp Analysis of Model-based Reinforcement Learning with Self-Play
http://proceedings.mlr.press/v139/liu21z.html
AUTHORS: Qinghua Liu, Tiancheng Yu, Yu Bai, Chi Jin
HIGHLIGHT: In this paper, we present a sharp analysis of model-based self-play algorithms for multi-agent Markov games.

640, TITLE: Lottery Ticket Preserves Weight Correlation: Is It Desirable or Not?
http://proceedings.mlr.press/v139/liu21aa.html
AUTHORS: Ning Liu, Geng Yuan, Zhengping Che, Xuan Shen, Xiaolong Ma, Qing Jin, Jian Ren, Jian Tang, Sijia Liu, Yanzhi Wang
HIGHLIGHT: In this work, we investigate the underlying condition and rationale behind the winning property, and find that the underlying reason is largely attributed to the correlation between initialized weights and final-trained weights when the learning rate is not sufficiently large.

641, TITLE: Group Fisher Pruning for Practical Network Compression
http://proceedings.mlr.press/v139/liu21ab.html
AUTHORS: Liyang Liu, Shilong Zhang, Zhanghui Kuang, Aojun Zhou, Jing-Hao Xue, Xinjiang Wang, Yimin Chen, Wenming Yang, Qingmin Liao, Wayne Zhang
HIGHLIGHT: In this paper, we present a general channel pruning approach that can be applied to various complicated structures.

642, TITLE: Infinite-Dimensional Optimization for Zero-Sum Games via Variational Transport
http://proceedings.mlr.press/v139/liu21ac.html
AUTHORS: Lewis Liu, Yufeng Zhang, Zhuoran Yang, Reza Babanezhad, Zhaoran Wang
HIGHLIGHT: In this paper, we consider infinite-dimensional zero-sum games by a min-max distributional optimization problem over a space of probability measures defined on a continuous variable set, which is inspired by finding a mixed NE for finite-dimensional zero-sum games.

643, TITLE: Noise and Fluctuation of Finite Learning Rate Stochastic Gradient Descent
http://proceedings.mlr.press/v139/liu21ad.html
AUTHORS: Kangqiao Liu, Liu Ziyin, Masahito Ueda
HIGHLIGHT: In this work, we propose to study the basic properties of SGD and its variants in the non-vanishing learning rate regime.

644, TITLE: Multi-layered Network Exploration via Random Walks: From Offline Optimization to Online Learning
http://proceedings.mlr.press/v139/liu21ae.html
AUTHORS: Xutong Liu, Jinhang Zuo, Xiaowei Chen, Wei Chen, John C. S. Lui
HIGHLIGHT: The MuLaNE task is to allocate total random walk budget B into each network layer so that the total weights of the unique nodes visited by random walks are maximized. We systematically study this problem from offline optimization to online learning.

645, TITLE: Relative Positional Encoding for Transformers with Linear Complexity
http://proceedings.mlr.press/v139/liutkus21a.html
AUTHORS: Antoine Liutkus, Ondrej Ci?fka, Shih-Lun Wu, Umut Simsekli, Yi-Hsuan Yang, Gael Richard
HIGHLIGHT: In this paper, we bridge this gap and present Stochastic Positional Encoding as a way to generate PE that can be used as a replacement to the classical additive (sinusoidal) PE and provably behaves like RPE.

646, TITLE: Joint Online Learning and Decision-making via Dual Mirror Descent
http://proceedings.mlr.press/v139/lobos21a.html
AUTHORS: Alfonso Lobos, Paul Grigas, Zheng Wen
HIGHLIGHT: We propose a novel offline benchmark and a new algorithm that mixes an online dual mirror descent scheme with a generic parameter learning process.

647, TITLE: Symmetric Spaces for Graph Embeddings: A Finsler-Riemannian Approach
http://proceedings.mlr.press/v139/lopez21a.html
AUTHORS: Federico Lopez, Beatrice Pozzetti, Steve Trettel, Michael Strube, Anna Wienhard
HIGHLIGHT: This enables us to introduce a new method, the use of Finsler metrics integrated in a Riemannian optimization scheme, that better adapts to dissimilar structures in the graph.

648, TITLE: HEMET: A Homomorphic-Encryption-Friendly Privacy-Preserving Mobile Neural Network Architecture
http://proceedings.mlr.press/v139/lou21a.html
AUTHORS: Qian Lou, Lei Jiang
HIGHLIGHT: In this paper, we propose a \textbf{HE}-friendly privacy-preserving \textbf{M}obile neural n\textbf{ET}work architecture, \textbf{HEMET}.

649, TITLE: Optimal Complexity in Decentralized Training
http://proceedings.mlr.press/v139/lu21a.html
AUTHORS: Yucheng Lu, Christopher De Sa
HIGHLIGHT: In this paper, we provide a tight lower bound on the iteration complexity for such methods in a stochastic non-convex setting.

650, TITLE: DANCE: Enhancing saliency maps using decoys
http://proceedings.mlr.press/v139/lu21b.html
AUTHORS: Yang Young Lu, Wenbo Guo, Xinyu Xing, William Stafford Noble
HIGHLIGHT: To address these issues, we propose a framework, DANCE, which improves the robustness of saliency methods by following a two-step procedure.

651, TITLE: Binary Classification from Multiple Unlabeled Datasets via Surrogate Set Classification
http://proceedings.mlr.press/v139/lu21c.html
AUTHORS: Nan Lu, Shida Lei, Gang Niu, Issei Sato, Masashi Sugiyama
HIGHLIGHT: In this paper, we propose a new approach for binary classification from m U-sets for $m\ge2$.

652, TITLE: Variance Reduced Training with Stratified Sampling for Forecasting Models
http://proceedings.mlr.press/v139/lu21d.html
AUTHORS: Yucheng Lu, Youngsuk Park, Lifan Chen, Yuyang Wang, Christopher De Sa, Dean Foster
HIGHLIGHT: In this paper, we provably show under such heterogeneity, training a forecasting model with commonly used stochastic optimizers (e.g. SGD) potentially suffers large variance on gradient estimation, and thus incurs long-time training.

653, TITLE: ACE: Explaining cluster from an adversarial perspective
http://proceedings.mlr.press/v139/lu21e.html
AUTHORS: Yang Young Lu, Timothy C Yu, Giancarlo Bonora, William Stafford Noble
HIGHLIGHT: Here we propose an integrated deep learning framework, Adversarial Clustering Explanation (ACE), that bundles all three steps into a single workflow.

654, TITLE: On Monotonic Linear Interpolation of Neural Network Parameters
http://proceedings.mlr.press/v139/lucas21a.html
AUTHORS: James R Lucas, Juhan Bae, Michael R Zhang, Stanislav Fort, Richard Zemel, Roger B Grosse
HIGHLIGHT: This Monotonic Linear Interpolation (MLI) property, first observed by Goodfellow et al. 2014, persists in spite of the non-convex objectives and highly non-linear training dynamics of neural networks. Extending this work, we evaluate several hypotheses for this property that, to our knowledge, have not yet been explored.

655, TITLE: Improving Breadth-Wise Backpropagation in Graph Neural Networks Helps Learning Long-Range Dependencies.
http://proceedings.mlr.press/v139/lukovnikov21a.html
AUTHORS: Denis Lukovnikov, Asja Fischer
HIGHLIGHT: In this work, we focus on the ability of graph neural networks (GNNs) to learn long-range patterns in graphs with edge features.

656, TITLE: GraphDF: A Discrete Flow Model for Molecular Graph Generation
http://proceedings.mlr.press/v139/luo21a.html
AUTHORS: Youzhi Luo, Keqiang Yan, Shuiwang Ji

HIGHLIGHT: In this work, we propose GraphDF, a novel discrete latent variable model for molecular graph generation based on normalizing flow methods.

657, TITLE: Trajectory Diversity for Zero-Shot Coordination
http://proceedings.mlr.press/v139/lupu21a.html
AUTHORS: Andrei Lupu, Brandon Cui, Hengyuan Hu, Jakob Foerster
HIGHLIGHT: To this end, we introduce \textit{Trajectory Diversity} (TrajeDi) – a differentiable objective for generating diverse reinforcement learning policies.

658, TITLE: HyperHyperNetwork for the Design of Antenna Arrays
http://proceedings.mlr.press/v139/lutati21a.html
AUTHORS: Shahar Lutati, Lior Wolf
HIGHLIGHT: We present deep learning methods for the design of arrays and single instances of small antennas.

659, TITLE: Value Iteration in Continuous Actions, States and Time
http://proceedings.mlr.press/v139/lutter21a.html
AUTHORS: Michael Lutter, Shie Mannor, Jan Peters, Dieter Fox, Animesh Garg
HIGHLIGHT: In this paper, we propose continuous fitted value iteration (cFVI).

660, TITLE: Meta-Cal: Well-controlled Post-hoc Calibration by Ranking
http://proceedings.mlr.press/v139/ma21a.html
AUTHORS: Xingchen Ma, Matthew B. Blaschko
HIGHLIGHT: In this paper, we introduce two constraints that are worth consideration in designing a calibration map for post-hoc calibration.

661, TITLE: Neural-Pull: Learning Signed Distance Function from Point clouds by Learning to Pull Space onto Surface
http://proceedings.mlr.press/v139/ma21b.html
AUTHORS: Baorui Ma, Zhizhong Han, Yu-Shen Liu, Matthias Zwicker
HIGHLIGHT: In this paper, we introduce Neural-Pull, a new approach that is simple and leads to high quality SDFs.

662, TITLE: Learning Stochastic Behaviour from Aggregate Data
http://proceedings.mlr.press/v139/ma21c.html
AUTHORS: Shaojun Ma, Shu Liu, Hongyuan Zha, Haomin Zhou
HIGHLIGHT: We propose a novel method using the weak form of Fokker Planck Equation (FPE) — a partial differential equation — to describe the density evolution of data in a sampled form, which is then combined with Wasserstein generative adversarial network (WGAN) in the training process.

663, TITLE: Local Algorithms for Finding Densely Connected Clusters
http://proceedings.mlr.press/v139/macgregor21a.html
AUTHORS: Peter Macgregor, He Sun
HIGHLIGHT: Following this line of research, in this work we study local algorithms for finding a pair of vertex sets defined with respect to their inter-connection and their relationship with the rest of the graph.

664, TITLE: Learning to Generate Noise for Multi-Attack Robustness
http://proceedings.mlr.press/v139/madaan21a.html
AUTHORS: Divyam Madaan, Jinwoo Shin, Sung Ju Hwang
HIGHLIGHT: To address these challenges, we propose a novel meta-learning framework that explicitly learns to generate noise to improve the model's robustness against multiple types of attacks.

665, TITLE: Learning Interaction Kernels for Agent Systems on Riemannian Manifolds
http://proceedings.mlr.press/v139/maggioni21a.html
AUTHORS: Mauro Maggioni, Jason J Miller, Hongda Qiu, Ming Zhong
HIGHLIGHT: We consider the problem of learning interaction kernels in these dynamical systems constrained to evolve on Riemannian manifolds from given trajectory data.

666, TITLE: Tesseract: Tensorised Actors for Multi-Agent Reinforcement Learning
http://proceedings.mlr.press/v139/mahajan21a.html
AUTHORS: Anuj Mahajan, Mikayel Samvelyan, Lei Mao, Viktor Makoviychuk, Animesh Garg, Jean Kossaifi, Shimon Whiteson, Yuke Zhu, Animashree Anandkumar

HIGHLIGHT: In this work, we focus on the fundamental hurdle affecting both value-based and policy-gradient approaches: an exponential blowup of the action space with the number of agents.

667, TITLE: Domain Generalization using Causal Matching
http://proceedings.mlr.press/v139/mahajan21b.html
AUTHORS: Divyat Mahajan, Shruti Tople, Amit Sharma
HIGHLIGHT: Based on this objective, we propose matching-based algorithms when base objects are observed (e.g., through data augmentation) and approximate the objective when objects are not observed (MatchDG).

668, TITLE: Stability and Convergence of Stochastic Gradient Clipping: Beyond Lipschitz Continuity and Smoothness
http://proceedings.mlr.press/v139/mai21a.html
AUTHORS: Vien V. Mai, Mikael Johansson
HIGHLIGHT: This paper establishes both qualitative and quantitative convergence results of the clipped stochastic (sub)gradient method (SGD) for non-smooth convex functions with rapidly growing subgradients.

669, TITLE: Nonparametric Hamiltonian Monte Carlo
http://proceedings.mlr.press/v139/mak21a.html
AUTHORS: Carol Mak, Fabian Zaiser, Luke Ong
HIGHLIGHT: This paper introduces the Nonparametric Hamiltonian Monte Carlo (NP-HMC) algorithm which generalises HMC to nonparametric models.

670, TITLE: Exploiting structured data for learning contagious diseases under incomplete testing
http://proceedings.mlr.press/v139/makar21a.html
AUTHORS: Maggie Makar, Lauren West, David Hooper, Eric Horvitz, Erica Shenoy, John Guttag
HIGHLIGHT: In this work we ask: can we build reliable infection prediction models when the observed data is collected under limited, and biased testing that prioritizes testing symptomatic individuals?

671, TITLE: Near-Optimal Algorithms for Explainable k-Medians and k-Means
http://proceedings.mlr.press/v139/makarychev21a.html
AUTHORS: Konstantin Makarychev, Liren Shan
HIGHLIGHT: We propose a new algorithm for this problem which is $\tilde O(\log k)$ competitive with k-medians with ℓ_1 norm and $\tilde O(k)$ competitive with k-means.

672, TITLE: KO codes: inventing nonlinear encoding and decoding for reliable wireless communication via deep-learning
http://proceedings.mlr.press/v139/makkuva21a.html
AUTHORS: Ashok V Makkuva, Xiyang Liu, Mohammad Vahid Jamali, Hessam Mahdavifar, Sewoong Oh, Pramod Viswanath
HIGHLIGHT: In this paper, we construct KO codes, a computationally efficient family of deep-learning driven (encoder, decoder) pairs that outperform the state-of-the-art reliability performance on the standardized AWGN channel.

673, TITLE: Quantifying the Benefit of Using Differentiable Learning over Tangent Kernels
http://proceedings.mlr.press/v139/malach21a.html
AUTHORS: Eran Malach, Pritish Kamath, Emmanuel Abbe, Nathan Srebro
HIGHLIGHT: We study the relative power of learning with gradient descent on differentiable models, such as neural networks, versus using the corresponding tangent kernels.

674, TITLE: Inverse Constrained Reinforcement Learning
http://proceedings.mlr.press/v139/malik21a.html
AUTHORS: Shehryar Malik, Usman Anwar, Alireza Aghasi, Ali Ahmed
HIGHLIGHT: In this work, we consider the problem of learning constraints from demonstrations of a constraint-abiding agent’s behavior.

675, TITLE: A Sampling-Based Method for Tensor Ring Decomposition
http://proceedings.mlr.press/v139/malik21b.html
AUTHORS: Osman Asif Malik, Stephen Becker
HIGHLIGHT: We propose a sampling-based method for computing the tensor ring (TR) decomposition of a data tensor.

676, TITLE: Sample Efficient Reinforcement Learning In Continuous State Spaces: A Perspective Beyond Linearity
http://proceedings.mlr.press/v139/malik21c.html
AUTHORS: Dhruv Malik, Aldo Pacchiano, Vishwak Srinivasan, Yuanzhi Li

HIGHLIGHT: To resolve this discrepancy between theory and practice, we introduce the Effective Planning Window (EPW) condition, a structural condition on MDPs that makes no linearity assumptions.

677, TITLE: Beyond the Pareto Efficient Frontier: Constraint Active Search for Multiobjective Experimental Design
http://proceedings.mlr.press/v139/malkomes21a.html
AUTHORS: Gustavo Malkomes, Bolong Cheng, Eric H Lee, Mike Mccourt
HIGHLIGHT: We introduce an active search algorithm called Expected Coverage Improvement (ECI) to efficiently discover the region of satisfaction and simultaneously sample diverse acceptable configurations.

678, TITLE: Consistent Nonparametric Methods for Network Assisted Covariate Estimation
http://proceedings.mlr.press/v139/mao21a.html
AUTHORS: Xueyu Mao, Deepayan Chakrabarti, Purnamrita Sarkar
HIGHLIGHT: In this paper we propose a new similarity measure between two nodes based on the patterns of their 2-hop neighborhoods.

679, TITLE: Near-Optimal Model-Free Reinforcement Learning in Non-Stationary Episodic MDPs
http://proceedings.mlr.press/v139/mao21b.html
AUTHORS: Weichao Mao, Kaiqing Zhang, Ruihao Zhu, David Simchi-Levi, Tamer Basar
HIGHLIGHT: We propose Restarted Q-Learning with Upper Confidence Bounds (RestartQ-UCB), the first model-free algorithm for non-stationary RL, and show that it outperforms existing solutions in terms of dynamic regret.

680, TITLE: Adaptive Sampling for Best Policy Identification in Markov Decision Processes
http://proceedings.mlr.press/v139/marjani21a.html
AUTHORS: Aymen Al Marjani, Alexandre Proutiere
HIGHLIGHT: We investigate the problem of best-policy identification in discounted Markov Decision Processes (MDPs) when the learner has access to a generative model.

681, TITLE: Explanations for Monotonic Classifiers.
http://proceedings.mlr.press/v139/marques-silva21a.html
AUTHORS: Joao Marques-Silva, Thomas Gerspacher, Martin C Cooper, Alexey Ignatiev, Nina Narodytska
HIGHLIGHT: This paper describes novel algorithms for the computation of one formal explanation of a (black-box) monotonic classifier.

682, TITLE: Multi-Agent Training beyond Zero-Sum with Correlated Equilibrium Meta-Solvers
http://proceedings.mlr.press/v139/marris21a.html
AUTHORS: Luke Marris, Paul Muller, Marc Lanctot, Karl Tuyls, Thore Graepel
HIGHLIGHT: We propose Joint Policy-Space Response Oracles (JPSRO), an algorithm for training agents in n-player, general-sum extensive form games, which provably converges to an equilibrium.

683, TITLE: Blind Pareto Fairness and Subgroup Robustness
http://proceedings.mlr.press/v139/martinez21a.html
AUTHORS: Natalia L Martinez, Martin A Bertran, Afroditi Papadaki, Miguel Rodrigues, Guillermo Sapiro
HIGHLIGHT: In this work we analyze the space of solutions for worst-case fairness beyond demographics, and propose Blind Pareto Fairness (BPF), a method that leverages no-regret dynamics to recover a fair minimax classifier that reduces worst-case risk of any potential subgroup of sufficient size, and guarantees that the remaining population receives the best possible level of service.

684, TITLE: Necessary and sufficient conditions for causal feature selection in time series with latent common causes
http://proceedings.mlr.press/v139/mastakouri21a.html
AUTHORS: Atalanti A Mastakouri, Bernhard Sch?lkopf, Dominik Janzing
HIGHLIGHT: We study the identification of direct and indirect causes on time series with latent variables, and provide a constrained-based causal feature selection method, which we prove that is both sound and complete under some graph constraints.

685, TITLE: Proximal Causal Learning with Kernels: Two-Stage Estimation and Moment Restriction
http://proceedings.mlr.press/v139/mastouri21a.html
AUTHORS: Afsaneh Mastouri, Yuchen Zhu, Limor Gultchin, Anna Korba, Ricardo Silva, Matt Kusner, Arthur Gretton, Krikamol Muandet
HIGHLIGHT: We propose two kernel-based meth-ods for nonlinear causal effect estimation in thissetting: (a) a two-stage regression approach, and(b) a maximum moment restriction approach.

686, TITLE: Robust Unsupervised Learning via L-statistic Minimization

http://proceedings.mlr.press/v139/maurer21a.html
AUTHORS: Andreas Maurer, Daniela Angela Parletta, Andrea Paudice, Massimiliano Pontil
HIGHLIGHT: We present a general approach to this problem focusing on unsupervised learning.

687, TITLE: Adversarial Multi Class Learning under Weak Supervision with Performance Guarantees
http://proceedings.mlr.press/v139/mazzetto21a.html
AUTHORS: Alessio Mazzetto, Cyrus Cousins, Dylan Sam, Stephen H Bach, Eli Upfal
HIGHLIGHT: We develop a rigorous approach for using a set of arbitrarily correlated weak supervision sources in order to solve a multiclass classification task when only a very small set of labeled data is available.

688, TITLE: Fundamental Tradeoffs in Distributionally Adversarial Training
http://proceedings.mlr.press/v139/mehrabi21a.html
AUTHORS: Mohammad Mehrabi, Adel Javanmard, Ryan A. Rossi, Anup Rao, Tung Mai
HIGHLIGHT: In this paper, we focus on \emph{distribution perturbing} adversary framework wherein the adversary can change the test distribution within a neighborhood of the training data distribution.

689, TITLE: Leveraging Non-uniformity in First-order Non-convex Optimization
http://proceedings.mlr.press/v139/mei21a.html
AUTHORS: Jincheng Mei, Yue Gao, Bo Dai, Csaba Szepesvari, Dale Schuurmans
HIGHLIGHT: Motivated by properties of objective functions that arise in machine learning, we propose a non-uniform refinement of these notions, leading to \emph{Non-uniform Smoothness} (NS) and \emph{Non-uniform ?{}ojasiewicz inequality} (N?{}).

690, TITLE: Controlling Graph Dynamics with Reinforcement Learning and Graph Neural Networks
http://proceedings.mlr.press/v139/meirom21a.html
AUTHORS: Eli Meirom, Haggai Maron, Shie Mannor, Gal Chechik
HIGHLIGHT: We consider the problem of controlling a partially-observed dynamic process on a graph by a limited number of interventions.

691, TITLE: A theory of high dimensional regression with arbitrary correlations between input features and target functions: sample complexity, multiple descent curves and a hierarchy of phase transitions
http://proceedings.mlr.press/v139/mel21a.html
AUTHORS: Gabriel Mel, Surya Ganguli
HIGHLIGHT: To understand this better we revisit ridge regression in high dimensions, which corresponds to an exceedingly simple architecture and loss function, but we analyze its performance under arbitrary correlations between input features and the target function.

692, TITLE: Neural Architecture Search without Training
http://proceedings.mlr.press/v139/mellor21a.html
AUTHORS: Joe Mellor, Jack Turner, Amos Storkey, Elliot J Crowley
HIGHLIGHT: In this work, we examine the overlap of activations between datapoints in untrained networks and motivate how this can give a measure which is usefully indicative of a network’s trained performance.

693, TITLE: Fast active learning for pure exploration in reinforcement learning
http://proceedings.mlr.press/v139/menard21a.html
AUTHORS: Pierre Menard, Omar Darwiche Domingues, Anders Jonsson, Emilie Kaufmann, Edouard Leurent, Michal Valko
HIGHLIGHT: We show that, surprisingly, for a pure-exploration objective of \emph{reward-free exploration}, bonuses that scale with $1/n$ bring faster learning rates, improving the known upper bounds with respect to the dependence on the horizon H.

694, TITLE: UCB Momentum Q-learning: Correcting the bias without forgetting
http://proceedings.mlr.press/v139/menard21b.html
AUTHORS: Pierre Menard, Omar Darwiche Domingues, Xuedong Shang, Michal Valko
HIGHLIGHT: We propose UCBMQ, Upper Confidence Bound Momentum Q-learning, a new algorithm for reinforcement learning in tabular and possibly stage-dependent, episodic Markov decision process.

695, TITLE: An Integer Linear Programming Framework for Mining Constraints from Data
http://proceedings.mlr.press/v139/meng21a.html
AUTHORS: Tao Meng, Kai-Wei Chang
HIGHLIGHT: In this paper, we present a general framework for mining constraints from data.

696, TITLE: A statistical perspective on distillation
http://proceedings.mlr.press/v139/menon21a.html
AUTHORS: Aditya K Menon, Ankit Singh Rawat, Sashank Reddi, Seungyeon Kim, Sanjiv Kumar
HIGHLIGHT: In this paper, we present a statistical perspective on distillation which provides an answer to these questions.

697, TITLE: Learn2Hop: Learned Optimization on Rough Landscapes
http://proceedings.mlr.press/v139/merchant21a.html
AUTHORS: Amil Merchant, Luke Metz, Samuel S Schoenholz, Ekin D Cubuk
HIGHLIGHT: In this work, we propose adapting recent developments in meta-learning to these many-minima problems by learning the optimization algorithm for various loss landscapes.

698, TITLE: Counterfactual Credit Assignment in Model-Free Reinforcement Learning
http://proceedings.mlr.press/v139/mesnard21a.html
AUTHORS: Thomas Mesnard, Theophane Weber, Fabio Viola, Shantanu Thakoor, Alaa Saade, Anna Harutyunyan, Will Dabney, Thomas S Stepleton, Nicolas Heess, Arthur Guez, Eric Moulines, Marcus Hutter, Lars Buesing, Remi Munos
HIGHLIGHT: We formulate a family of policy gradient algorithms that use these future-conditional value functions as baselines or critics, and show that they are provably low variance.

699, TITLE: Provably Efficient Learning of Transferable Rewards
http://proceedings.mlr.press/v139/metelli21a.html
AUTHORS: Alberto Maria Metelli, Giorgia Ramponi, Alessandro Concetti, Marcello Restelli
HIGHLIGHT: In this paper, we study the theoretical properties of the class of reward functions that are compatible with the expert’s behavior.

700, TITLE: Mixed Nash Equilibria in the Adversarial Examples Game
http://proceedings.mlr.press/v139/meunier21a.html
AUTHORS: Laurent Meunier, Meyer Scetbon, Rafael B Pinot, Jamal Atif, Yann Chevaleyre
HIGHLIGHT: This paper tackles the problem of adversarial examples from a game theoretic point of view.

701, TITLE: Learning in Nonzero-Sum Stochastic Games with Potentials
http://proceedings.mlr.press/v139/mguni21a.html
AUTHORS: David H Mguni, Yutong Wu, Yali Du, Yaodong Yang, Ziyi Wang, Minne Li, Ying Wen, Joel Jennings, Jun Wang
HIGHLIGHT: In this paper, we introduce a new generation of MARL learners that can handle \textit{nonzero-sum} payoff structures and continuous settings.

702, TITLE: EfficientTTS: An Efficient and High-Quality Text-to-Speech Architecture
http://proceedings.mlr.press/v139/miao21a.html
AUTHORS: Chenfeng Miao, Liang Shuang, Zhengchen Liu, Chen Minchuan, Jun Ma, Shaojun Wang, Jing Xiao
HIGHLIGHT: In this work, we address the Text-to-Speech (TTS) task by proposing a non-autoregressive architecture called EfficientTTS.

703, TITLE: Outside the Echo Chamber: Optimizing the Performative Risk
http://proceedings.mlr.press/v139/miller21a.html
AUTHORS: John P Miller, Juan C Perdomo, Tijana Zrnic
HIGHLIGHT: In this paper, we shift attention beyond performative stability and focus on optimizing the performative risk directly.

704, TITLE: Accuracy on the Line: on the Strong Correlation Between Out-of-Distribution and In-Distribution Generalization
http://proceedings.mlr.press/v139/miller21b.html
AUTHORS: John P Miller, Rohan Taori, Aditi Raghunathan, Shiori Sagawa, Pang Wei Koh, Vaishaal Shankar, Percy Liang, Yair Carmon, Ludwig Schmidt
HIGHLIGHT: In this paper, we empirically show that out-of-distribution performance is strongly correlated with in-distribution performance for a wide range of models and distribution shifts.

705, TITLE: Signatured Deep Fictitious Play for Mean Field Games with Common Noise
http://proceedings.mlr.press/v139/min21a.html
AUTHORS: Ming Min, Ruimeng Hu

HIGHLIGHT: In this paper, based on the rough path theory, we propose a novel single-loop algorithm, named signatured deep fictitious play (Sig-DFP), by which we can work with the unfixed common noise setup to avoid the nested loop structure and reduce the computational complexity significantly.

706, TITLE: Meta-StyleSpeech : Multi-Speaker Adaptive Text-to-Speech Generation
http://proceedings.mlr.press/v139/min21b.html
AUTHORS: Dongchan Min, Dong Bok Lee, Eunho Yang, Sung Ju Hwang
HIGHLIGHT: In this work, we propose StyleSpeech, a new TTS model which not only synthesizes high-quality speech but also effectively adapts to new speakers.

707, TITLE: On the Explicit Role of Initialization on the Convergence and Implicit Bias of Overparametrized Linear Networks
http://proceedings.mlr.press/v139/min21c.html
AUTHORS: Hancheng Min, Salma Tarmoun, Rene Vidal, Enrique Mallada
HIGHLIGHT: In this paper, we present a novel analysis of single-hidden-layer linear networks trained under gradient flow, which connects initialization, optimization, and overparametrization.

708, TITLE: An Identifiable Double VAE For Disentangled Representations
http://proceedings.mlr.press/v139/mita21a.html
AUTHORS: Graziano Mita, Maurizio Filippone, Pietro Michiardi
HIGHLIGHT: Working along this line, we propose a novel VAE-based generative model with theoretical guarantees on identifiability.

709, TITLE: Offline Meta-Reinforcement Learning with Advantage Weighting
http://proceedings.mlr.press/v139/mitchell21a.html
AUTHORS: Eric Mitchell, Rafael Rafailov, Xue Bin Peng, Sergey Levine, Chelsea Finn
HIGHLIGHT: This paper introduces the offline meta-reinforcement learning (offline meta-RL) problem setting and proposes an algorithm that performs well in this setting.

710, TITLE: The Power of Log-Sum-Exp: Sequential Density Ratio Matrix Estimation for Speed-Accuracy Optimization
http://proceedings.mlr.press/v139/miyagawa21a.html
AUTHORS: Taiki Miyagawa, Akinori F Ebihara
HIGHLIGHT: We propose a model for multiclass classification of time series to make a prediction as early and as accurate as possible.

711, TITLE: PODS: Policy Optimization via Differentiable Simulation
http://proceedings.mlr.press/v139/mora21a.html
AUTHORS: Miguel Angel Zamora Mora, Momchil P Peychev, Sehoon Ha, Martin Vechev, Stelian Coros
HIGHLIGHT: In this paper, with the goal of improving the performance exhibited by RL algorithms, we explore a systematic way of leveraging the additional information provided by an emerging class of differentiable simulators.

712, TITLE: Efficient Deviation Types and Learning for Hindsight Rationality in Extensive-Form Games
http://proceedings.mlr.press/v139/morrill21a.html
AUTHORS: Dustin Morrill, Ryan D?Orazio, Marc Lanctot, James R Wright, Michael Bowling, Amy R Greenwald
HIGHLIGHT: Integrating the idea of time selection into counterfactual regret minimization (CFR), we introduce the extensive-form regret minimization (EFR) algorithm that achieves hindsight rationality for any given set of behavioral deviations with computation that scales closely with the complexity of the set.

713, TITLE: Neural Rough Differential Equations for Long Time Series
http://proceedings.mlr.press/v139/morrill21b.html
AUTHORS: James Morrill, Cristopher Salvi, Patrick Kidger, James Foster
HIGHLIGHT: Existing methods for computing the forward pass of a Neural CDE involve embedding the incoming time series into path space, often via interpolation, and using evaluations of this path to drive the hidden state. Here, we use rough path theory to extend this formulation.

714, TITLE: Connecting Interpretability and Robustness in Decision Trees through Separation
http://proceedings.mlr.press/v139/moshkovitz21a.html
AUTHORS: Michal Moshkovitz, Yao-Yuan Yang, Kamalika Chaudhuri
HIGHLIGHT: Curiously, a connection between robustness and interpretability was empirically observed, but the theoretical reasoning behind it remained elusive. In this paper, we rigorously investigate this connection.

715, TITLE: Outlier-Robust Optimal Transport
http://proceedings.mlr.press/v139/mukherjee21a.html
AUTHORS: Debarghya Mukherjee, Aritra Guha, Justin M Solomon, Yuekai Sun, Mikhail Yurochkin
HIGHLIGHT: To address this issue, we propose an outlier-robust formulation of OT.

716, TITLE: Oblivious Sketching for Logistic Regression
http://proceedings.mlr.press/v139/munteanu21a.html
AUTHORS: Alexander Munteanu, Simon Omlor, David Woodruff
HIGHLIGHT: To answer this question, we present the first data oblivious sketch for logistic regression.

717, TITLE: Bias-Variance Reduced Local SGD for Less Heterogeneous Federated Learning
http://proceedings.mlr.press/v139/murata21a.html
AUTHORS: Tomoya Murata, Taiji Suzuki
HIGHLIGHT: In this paper, we study a new local algorithm called Bias-Variance Reduced Local SGD (BVR-L-SGD) for nonconvex distributed optimization.

718, TITLE: Implicit-PDF: Non-Parametric Representation of Probability Distributions on the Rotation Manifold
http://proceedings.mlr.press/v139/murphy21a.html
AUTHORS: Kieran A Murphy, Carlos Esteves, Varun Jampani, Srikumar Ramalingam, Ameesh Makadia
HIGHLIGHT: Our key idea is to represent the distributions implicitly, with a neural network that estimates the probability density, given the input image and a candidate pose.

719, TITLE: No-regret Algorithms for Capturing Events in Poisson Point Processes
http://proceedings.mlr.press/v139/mutny21a.html
AUTHORS: Mojmir Mutny, Andreas Krause
HIGHLIGHT: By partitioning the domain into separate small regions, and using heteroscedastic linear regression, we propose a tractable estimator of Poisson process rates for two feedback models: \emph{count-record}, where exact locations of events are observed, and \emph{histogram} feedback, where only counts of events are observed.

720, TITLE: Online Limited Memory Neural-Linear Bandits with Likelihood Matching
http://proceedings.mlr.press/v139/nabati21a.html
AUTHORS: Ofir Nabati, Tom Zahavy, Shie Mannor
HIGHLIGHT: To alleviate this, we propose a likelihood matching algorithm that is resilient to catastrophic forgetting and is completely online.

721, TITLE: Quantitative Understanding of VAE as a Non-linearly Scaled Isometric Embedding
http://proceedings.mlr.press/v139/nakagawa21a.html
AUTHORS: Akira Nakagawa, Keizo Kato, Taiji Suzuki
HIGHLIGHT: This paper provides a quantitative understanding of VAE property through the differential geometric and information-theoretic interpretations of VAE.

722, TITLE: GMAC: A Distributional Perspective on Actor-Critic Framework
http://proceedings.mlr.press/v139/nam21a.html
AUTHORS: Daniel W Nam, Younghoon Kim, Chan Y Park
HIGHLIGHT: In this paper, we devise a distributional framework on actor-critic as a solution to distributional instability, action type restriction, and conflation between samples and statistics.

723, TITLE: Memory-Efficient Pipeline-Parallel DNN Training
http://proceedings.mlr.press/v139/narayanan21a.html
AUTHORS: Deepak Narayanan, Amar Phanishayee, Kaiyu Shi, Xie Chen, Matei Zaharia
HIGHLIGHT: In this work, we propose PipeDream-2BW, a system that supports memory-efficient pipeline parallelism.

724, TITLE: Randomized Dimensionality Reduction for Facility Location and Single-Linkage Clustering
http://proceedings.mlr.press/v139/narayanan21b.html
AUTHORS: Shyam Narayanan, Sandeep Silwal, Piotr Indyk, Or Zamir
HIGHLIGHT: Random dimensionality reduction is a versatile tool for speeding up algorithms for high-dimensional problems. We study its application to two clustering problems: the facility location problem, and the single-linkage hierarchical clustering problem, which is equivalent to computing the minimum spanning tree.

725, TITLE: Generating images with sparse representations
http://proceedings.mlr.press/v139/nash21a.html
AUTHORS: Charlie Nash, Jacob Menick, Sander Dieleman, Peter Battaglia
HIGHLIGHT: We present an alternative approach, inspired by common image compression methods like JPEG, and convert images to quantized discrete cosine transform (DCT) blocks, which are represented sparsely as a sequence of DCT channel, spatial location, and DCT coefficient triples.

726, TITLE: Geometric convergence of elliptical slice sampling
http://proceedings.mlr.press/v139/natarovskii21a.html
AUTHORS: Viacheslav Natarovskii, Daniel Rudolf, Bj?rn Sprungk
HIGHLIGHT: Under weak regularity assumptions on the posterior density we show that the corresponding Markov chain is geometrically ergodic and therefore yield qualitative convergence guarantees.

727, TITLE: HardCoRe-NAS: Hard Constrained diffeRentiable Neural Architecture Search
http://proceedings.mlr.press/v139/nayman21a.html
AUTHORS: Niv Nayman, Yonathan Aflalo, Asaf Noy, Lihi Zelnik
HIGHLIGHT: In this work we resolve this by introducing Hard Constrained diffeRentiable NAS (HardCoRe-NAS), that is based on an accurate formulation of the expected resource requirement and a scalable search method that satisfies the hard constraint throughout the search.

728, TITLE: Emergent Social Learning via Multi-agent Reinforcement Learning
http://proceedings.mlr.press/v139/ndousse21a.html
AUTHORS: Kamal K Ndousse, Douglas Eck, Sergey Levine, Natasha Jaques
HIGHLIGHT: This paper investigates whether independent reinforcement learning (RL) agents in a multi-agent environment can learn to use social learning to improve their performance.

729, TITLE: Bayesian Algorithm Execution: Estimating Computable Properties of Black-box Functions Using Mutual Information
http://proceedings.mlr.press/v139/neiswanger21a.html
AUTHORS: Willie Neiswanger, Ke Alexander Wang, Stefano Ermon
HIGHLIGHT: To tackle this problem, we present a procedure, InfoBAX, that sequentially chooses queries that maximize mutual information with respect to the algorithm’s output.

730, TITLE: Continuous Coordination As a Realistic Scenario for Lifelong Learning
http://proceedings.mlr.press/v139/nekoei21a.html
AUTHORS: Hadi Nekoei, Akilesh Badrinaaraayanan, Aaron Courville, Sarath Chandar
HIGHLIGHT: In this work, we introduce a multi-agent lifelong learning testbed that supports both zero-shot and few-shot settings.

731, TITLE: Policy Caches with Successor Features
http://proceedings.mlr.press/v139/nemecek21a.html
AUTHORS: Mark W Nemecek, Ron Parr
HIGHLIGHT: We present new bounds for the performance of optimal policies in a new task, as well as an approach to use these bounds to decide, when presented with a new task, whether to use cached policies or learn a new policy.

732, TITLE: Causality-aware counterfactual confounding adjustment as an alternative to linear residualization in anticausal prediction tasks based on linear learners
http://proceedings.mlr.press/v139/neto21a.html
AUTHORS: Elias Chaibub Neto
HIGHLIGHT: In this paper, we compare the linear residualization approach against the causality-aware confounding adjustment in anticausal prediction tasks.

733, TITLE: Incentivizing Compliance with Algorithmic Instruments
http://proceedings.mlr.press/v139/ngo21a.html
AUTHORS: Dung Daniel T Ngo, Logan Stapleton, Vasilis Syrgkanis, Steven Wu
HIGHLIGHT: We develop a novel recommendation mechanism that views the planner’s recommendation as a form of instrumental variable (IV) that only affects an agents’ action selection, but not the observed rewards.

734, TITLE: On the Proof of Global Convergence of Gradient Descent for Deep ReLU Networks with Linear Widths
http://proceedings.mlr.press/v139/nguyen21a.html
AUTHORS: Quynh Nguyen

HIGHLIGHT: We give a simple proof for the global convergence of gradient descent in training deep ReLU networks with the standard square loss, and show some of its improvements over the state-of-the-art.

735, TITLE: Value-at-Risk Optimization with Gaussian Processes
http://proceedings.mlr.press/v139/nguyen21b.html
AUTHORS: Quoc Phong Nguyen, Zhongxiang Dai, Bryan Kian Hsiang Low, Patrick Jaillet
HIGHLIGHT: This paper presents a novel VaR upper confidence bound (V-UCB) algorithm for maximizing the VaR of a black-box objective function with the first no-regret guarantee.

736, TITLE: Cross-model Back-translated Distillation for Unsupervised Machine Translation
http://proceedings.mlr.press/v139/nguyen21c.html
AUTHORS: Xuan-Phi Nguyen, Shafiq Joty, Thanh-Tung Nguyen, Kui Wu, Ai Ti Aw
HIGHLIGHT: We introduce a novel component to the standard UMT framework called Cross-model Back-translated Distillation (CBD), that is aimed to induce another level of data diversification that existing principles lack.

737, TITLE: Optimal Transport Kernels for Sequential and Parallel Neural Architecture Search
http://proceedings.mlr.press/v139/nguyen21d.html
AUTHORS: Vu Nguyen, Tam Le, Makoto Yamada, Michael A. Osborne
HIGHLIGHT: Building upon tree-Wasserstein (TW), which is a negative definite variant of OT, we develop a novel discrepancy for neural architectures, and demonstrate it within a Gaussian process surrogate model for the sequential NAS settings.

738, TITLE: Interactive Learning from Activity Description
http://proceedings.mlr.press/v139/nguyen21e.html
AUTHORS: Khanh X Nguyen, Dipendra Misra, Robert Schapire, Miroslav Dudik, Patrick Shafto
HIGHLIGHT: We present a novel interactive learning protocol that enables training request-fulfilling agents by verbally describing their activities.

739, TITLE: Nonmyopic Multifidelity Acitve Search
http://proceedings.mlr.press/v139/nguyen21f.html
AUTHORS: Quan Nguyen, Arghavan Modiri, Roman Garnett
HIGHLIGHT: We propose a model of multifidelity active search, as well as a novel, computationally efficient policy for this setting that is motivated by state-of-the-art classical policies.

740, TITLE: Tight Bounds on the Smallest Eigenvalue of the Neural Tangent Kernel for Deep ReLU Networks
http://proceedings.mlr.press/v139/nguyen21g.html
AUTHORS: Quynh Nguyen, Marco Mondelli, Guido F Montufar
HIGHLIGHT: In this paper, we provide tight bounds on the smallest eigenvalue of NTK matrices for deep ReLU nets, both in the limiting case of infinite widths and for finite widths.

741, TITLE: Temporal Predictive Coding For Model-Based Planning In Latent Space
http://proceedings.mlr.press/v139/nguyen21h.html
AUTHORS: Tung D Nguyen, Rui Shu, Tuan Pham, Hung Bui, Stefano Ermon
HIGHLIGHT: In this work, we present an information-theoretic approach that employs temporal predictive coding to encode elements in the environment that can be predicted across time.

742, TITLE: Differentially Private Densest Subgraph Detection
http://proceedings.mlr.press/v139/nguyen21i.html
AUTHORS: Dung Nguyen, Anil Vullikanti
HIGHLIGHT: We study the densest subgraph problem in the edge privacy model, in which the edges of the graph are private. We present the first sequential and parallel differentially private algorithms for this problem.

743, TITLE: Data Augmentation for Meta-Learning
http://proceedings.mlr.press/v139/ni21a.html
AUTHORS: Renkun Ni, Micah Goldblum, Amr Sharaf, Kezhi Kong, Tom Goldstein
HIGHLIGHT: We systematically dissect the meta-learning pipeline and investigate the distinct ways in which data augmentation can be integrated at both the image and class levels.

744, TITLE: Improved Denoising Diffusion Probabilistic Models
http://proceedings.mlr.press/v139/nichol21a.html
AUTHORS: Alexander Quinn Nichol, Prafulla Dhariwal

HIGHLIGHT: We show that with a few simple modifications, DDPMs can also achieve competitive log-likelihoods while maintaining high sample quality.

745, TITLE: Smooth p-Wasserstein Distance: Structure, Empirical Approximation, and Statistical Applications
http://proceedings.mlr.press/v139/nietert21a.html
AUTHORS: Sloan Nietert, Ziv Goldfeld, Kengo Kato
HIGHLIGHT: Motivated by the scalability of this framework to high dimensions, we investigate the structural and statistical behavior of the Gaussian-smoothed p-Wasserstein distance W(s)p, for arbitrary p=1.

746, TITLE: AdaXpert: Adapting Neural Architecture for Growing Data
http://proceedings.mlr.press/v139/niu21a.html
AUTHORS: Shuaicheng Niu, Jiaxiang Wu, Guanghui Xu, Yifan Zhang, Yong Guo, Peilin Zhao, Peng Wang, Mingkui Tan
HIGHLIGHT: To address this, we present a neural architecture adaptation method, namely Adaptation eXpert (AdaXpert), to efficiently adjust previous architectures on the growing data.

747, TITLE: Asynchronous Decentralized Optimization With Implicit Stochastic Variance Reduction
http://proceedings.mlr.press/v139/niwa21a.html
AUTHORS: Kenta Niwa, Guoqiang Zhang, W. Bastiaan Kleijn, Noboru Harada, Hiroshi Sawada, Akinori Fujino
HIGHLIGHT: In this paper, we reformulate the update procedure of ECL such that it implicitly includes the gradient modification of SVR by optimally selecting a constraint-strength control parameter.

748, TITLE: WGAN with an Infinitely Wide Generator Has No Spurious Stationary Points
http://proceedings.mlr.press/v139/no21a.html
AUTHORS: Albert No, Taeho Yoon, Kwon Sehyun, Ernest K Ryu
HIGHLIGHT: In this work, we show that GANs with a 2-layer infinite-width generator and a 2-layer finite-width discriminator trained with stochastic gradient ascent-descent have no spurious stationary points.

749, TITLE: The Impact of Record Linkage on Learning from Feature Partitioned Data
http://proceedings.mlr.press/v139/nock21a.html
AUTHORS: Richard Nock, Stephen Hardy, Wilko Henecka, Hamish Ivey-Law, Jakub Nabaglo, Giorgio Patrini, Guillaume Smith, Brian Thorne
HIGHLIGHT: In this paper, we provide the first assessment of the problem for supervised learning.

750, TITLE: Accuracy, Interpretability, and Differential Privacy via Explainable Boosting
http://proceedings.mlr.press/v139/nori21a.html
AUTHORS: Harsha Nori, Rich Caruana, Zhiqi Bu, Judy Hanwen Shen, Janardhan Kulkarni
HIGHLIGHT: We show that adding differential privacy to Explainable Boosting Machines (EBMs), a recent method for training interpretable ML models, yields state-of-the-art accuracy while protecting privacy.

751, TITLE: Posterior Value Functions: Hindsight Baselines for Policy Gradient Methods
http://proceedings.mlr.press/v139/nota21a.html
AUTHORS: Chris Nota, Philip Thomas, Bruno C. Da Silva
HIGHLIGHT: In this paper, we exploit the idea of hindsight and introduce posterior value functions.

752, TITLE: Global inducing point variational posteriors for Bayesian neural networks and deep Gaussian processes
http://proceedings.mlr.press/v139/ober21a.html
AUTHORS: Sebastian W Ober, Laurence Aitchison
HIGHLIGHT: We consider the optimal approximate posterior over the top-layer weights in a Bayesian neural network for regression, and show that it exhibits strong dependencies on the lower-layer weights.

753, TITLE: Regularizing towards Causal Invariance: Linear Models with Proxies
http://proceedings.mlr.press/v139/oberst21a.html
AUTHORS: Michael Oberst, Nikolaj Thams, Jonas Peters, David Sontag
HIGHLIGHT: We propose a method for learning linear models whose predictive performance is robust to causal interventions on unobserved variables, when noisy proxies of those variables are available.

754, TITLE: Sparsity-Agnostic Lasso Bandit
http://proceedings.mlr.press/v139/oh21a.html
AUTHORS: Min-Hwan Oh, Garud Iyengar, Assaf Zeevi

HIGHLIGHT: The main contribution of this paper is to propose an algorithm that does not require prior knowledge of the sparsity index s_0 and establish tight regret bounds on its performance under mild conditions.

755, TITLE: Autoencoder Image Interpolation by Shaping the Latent Space
http://proceedings.mlr.press/v139/oring21a.html
AUTHORS: Alon Oring, Zohar Yakhini, Yacov Hel-Or
HIGHLIGHT: In this paper, we propose a regularization technique that shapes the latent representation to follow a manifold that is consistent with the training images and that forces the manifold to be smooth and locally convex.

756, TITLE: Generalization Guarantees for Neural Architecture Search with Train-Validation Split
http://proceedings.mlr.press/v139/oymak21a.html
AUTHORS: Samet Oymak, Mingchen Li, Mahdi Soltanolkotabi
HIGHLIGHT: NAS methods commonly use bilevel optimization where one optimizes the weights over the training data (lower-level problem) and hyperparameters - such as the architecture - over the validation data (upper-level problem). This paper explores the statistical aspects of such problems with train-validation splits.

757, TITLE: Vector Quantized Models for Planning
http://proceedings.mlr.press/v139/ozair21a.html
AUTHORS: Sherjil Ozair, Yazhe Li, Ali Razavi, Ioannis Antonoglou, Aaron Van Den Oord, Oriol Vinyals
HIGHLIGHT: We present a new approach that handles stochastic and partially-observable environments.

758, TITLE: Training Adversarially Robust Sparse Networks via Bayesian Connectivity Sampling
http://proceedings.mlr.press/v139/ozdenizci21a.html
AUTHORS: Ozan ?zdenizci, Robert Legenstein
HIGHLIGHT: Motivated by the efficient and stable computational function of the brain in the presence of a highly dynamic synaptic connectivity structure, we propose an intrinsically sparse rewiring approach to train neural networks with state-of-the-art robust learning objectives under high sparsity.

759, TITLE: Opening the Blackbox: Accelerating Neural Differential Equations by Regularizing Internal Solver Heuristics
http://proceedings.mlr.press/v139/pal21a.html
AUTHORS: Avik Pal, Yingbo Ma, Viral Shah, Christopher V Rackauckas
HIGHLIGHT: We describe a novel regularization method that uses the internal cost heuristics of adaptive differential equation solvers combined with discrete adjoint sensitivities to guide the training process towards learning NDEs that are easier to solve.

760, TITLE: RNN with Particle Flow for Probabilistic Spatio-temporal Forecasting
http://proceedings.mlr.press/v139/pal21b.html
AUTHORS: Soumyasundar Pal, Liheng Ma, Yingxue Zhang, Mark Coates
HIGHLIGHT: In this work, we consider the time-series data as a random realization from a nonlinear state-space model and target Bayesian inference of the hidden states for probabilistic forecasting.

761, TITLE: Inference for Network Regression Models with Community Structure
http://proceedings.mlr.press/v139/pan21a.html
AUTHORS: Mengjie Pan, Tyler Mccormick, Bailey Fosdick
HIGHLIGHT: In this work, we present a novel regression modeling framework that models the errors as resulting from a community-based dependence structure and exploits the subsequent exchangeability properties of the error distribution to obtain parsimonious standard errors for regression parameters.

762, TITLE: Latent Space Energy-Based Model of Symbol-Vector Coupling for Text Generation and Classification
http://proceedings.mlr.press/v139/pang21a.html
AUTHORS: Bo Pang, Ying Nian Wu
HIGHLIGHT: We propose a latent space energy-based prior model for text generation and classification.

763, TITLE: Leveraging Good Representations in Linear Contextual Bandits
http://proceedings.mlr.press/v139/papini21a.html
AUTHORS: Matteo Papini, Andrea Tirinzoni, Marcello Restelli, Alessandro Lazaric, Matteo Pirotta
HIGHLIGHT: In this paper, we first provide a systematic analysis of the different definitions of “good” representations proposed in the literature. We then propose a novel selection algorithm able to adapt to the best representation in a set of M candidates.

764, TITLE: Wasserstein Distributional Normalization For Robust Distributional Certification of Noisy Labeled Data

http://proceedings.mlr.press/v139/park21a.html
AUTHORS: Sung Woo Park, Junseok Kwon
HIGHLIGHT: We propose a novel Wasserstein distributional normalization method that can classify noisy labeled data accurately.

765, TITLE: Unsupervised Representation Learning via Neural Activation Coding
http://proceedings.mlr.press/v139/park21b.html
AUTHORS: Yookoon Park, Sangho Lee, Gunhee Kim, David Blei
HIGHLIGHT: We present neural activation coding (NAC) as a novel approach for learning deep representations from unlabeled data for downstream applications.

766, TITLE: Conditional Distributional Treatment Effect with Kernel Conditional Mean Embeddings and U-Statistic Regression
http://proceedings.mlr.press/v139/park21c.html
AUTHORS: Junhyung Park, Uri Shalit, Bernhard Sch?lkopf, Krikamol Muandet
HIGHLIGHT: We propose to analyse the conditional distributional treatment effect (CoDiTE), which, in contrast to the more common conditional average treatment effect (CATE), is designed to encode a treatment's distributional aspects beyond the mean.

767, TITLE: Generative Adversarial Networks for Markovian Temporal Dynamics: Stochastic Continuous Data Generation
http://proceedings.mlr.press/v139/park21d.html
AUTHORS: Sung Woo Park, Dong Wook Shu, Junseok Kwon
HIGHLIGHT: In this paper, we present a novel generative adversarial network (GAN) that can describe Markovian temporal dynamics.

768, TITLE: Optimal Counterfactual Explanations in Tree Ensembles
http://proceedings.mlr.press/v139/parmentier21a.html
AUTHORS: Axel Parmentier, Thibaut Vidal
HIGHLIGHT: In this paper, we take a disciplined approach towards counterfactual explanations for tree ensembles.

769, TITLE: PHEW : Constructing Sparse Networks that Learn Fast and Generalize Well without Training Data
http://proceedings.mlr.press/v139/patil21a.html
AUTHORS: Shreyas Malakarjun Patil, Constantine Dovrolis
HIGHLIGHT: We first show that even though Synflow-L2 is optimal in terms of convergence, for a given network density, it results in sub-networks with "bottleneck" (narrow) layers {-} leading to poor performance as compared to other data-agnostic methods that use the same number of parameters. Then we propose a new method to construct sparse networks, without any training data, referred to as Paths with Higher-Edge Weights (PHEW).

770, TITLE: CombOptNet: Fit the Right NP-Hard Problem by Learning Integer Programming Constraints
http://proceedings.mlr.press/v139/paulus21a.html
AUTHORS: Anselm Paulus, Michal Rolinek, Vit Musil, Brandon Amos, Georg Martius
HIGHLIGHT: In this work, we aim to integrate integer programming solvers into neural network architectures as layers capable of learning both the cost terms and the constraints.

771, TITLE: Ensemble Bootstrapping for Q-Learning
http://proceedings.mlr.press/v139/peer21a.html
AUTHORS: Oren Peer, Chen Tessler, Nadav Merlis, Ron Meir
HIGHLIGHT: In this work, we introduce a new bias-reduced algorithm called Ensemble Bootstrapped Q-Learning (EBQL), a natural extension of Double-Q-learning to ensembles.

772, TITLE: Homomorphic Sensing: Sparsity and Noise
http://proceedings.mlr.press/v139/peng21a.html
AUTHORS: Liangzu Peng, Boshi Wang, Manolis Tsakiris
HIGHLIGHT: In this paper we present tighter and simpler conditions for the homomorphic sensing problem to admit a unique solution.

773, TITLE: How could Neural Networks understand Programs?
http://proceedings.mlr.press/v139/peng21b.html
AUTHORS: Dinglan Peng, Shuxin Zheng, Yatao Li, Guolin Ke, Di He, Tie-Yan Liu
HIGHLIGHT: Inspired by this, we propose a novel program semantics learning paradigm, that the model should learn from information composed of (1) the representations which align well with the fundamental operations in operational semantics, and (2) the information of environment transition, which is indispensable for program understanding.

774, TITLE: Privacy-Preserving Video Classification with Convolutional Neural Networks
http://proceedings.mlr.press/v139/pentyala21a.html
AUTHORS: Sikha Pentyala, Rafael Dowsley, Martine De Cock
HIGHLIGHT: We propose a privacy-preserving implementation of single-frame method based video classification with convolutional neural networks that allows a party to infer a label from a video without necessitating the video owner to disclose their video to other entities in an unencrypted manner.

775, TITLE: Rissanen Data Analysis: Examining Dataset Characteristics via Description Length
http://proceedings.mlr.press/v139/perez21a.html
AUTHORS: Ethan Perez, Douwe Kiela, Kyunghyun Cho
HIGHLIGHT: We introduce a method to determine if a certain capability helps to achieve an accurate model of given data.

776, TITLE: Modelling Behavioural Diversity for Learning in Open-Ended Games
http://proceedings.mlr.press/v139/perez-nieves21a.html
AUTHORS: Nicolas Perez-Nieves, Yaodong Yang, Oliver Slumbers, David H Mguni, Ying Wen, Jun Wang
HIGHLIGHT: In this work, we offer a geometric interpretation of behavioural diversity in games and introduce a novel diversity metric based on \emph{determinantal point processes} (DPP).

777, TITLE: From Poincar? Recurrence to Convergence in Imperfect Information Games: Finding Equilibrium via Regularization
http://proceedings.mlr.press/v139/perolat21a.html
AUTHORS: Julien Perolat, Remi Munos, Jean-Baptiste Lespiau, Shayegan Omidshafiei, Mark Rowland, Pedro Ortega, Neil Burch, Thomas Anthony, David Balduzzi, Bart De Vylder, Georgios Piliouras, Marc Lanctot, Karl Tuyls
HIGHLIGHT: In this paper we investigate the Follow the Regularized Leader dynamics in sequential imperfect information games (IIG).

778, TITLE: Spectral Smoothing Unveils Phase Transitions in Hierarchical Variational Autoencoders
http://proceedings.mlr.press/v139/pervez21a.html
AUTHORS: Adeel Pervez, Efstratios Gavves
HIGHLIGHT: We suggest that the hierarchical VAE objective explicitly includes the variance of the function parameterizing the mean and variance of the latent Gaussian distribution which itself is often a high variance function.

779, TITLE: Differentiable Sorting Networks for Scalable Sorting and Ranking Supervision
http://proceedings.mlr.press/v139/petersen21a.html
AUTHORS: Felix Petersen, Christian Borgelt, Hilde Kuehne, Oliver Deussen
HIGHLIGHT: That is, the ground truth order of sets of samples is known, while their absolute values remain unsupervised. For that, we propose differentiable sorting networks by relaxing their pairwise conditional swap operations.

780, TITLE: Megaverse: Simulating Embodied Agents at One Million Experiences per Second
http://proceedings.mlr.press/v139/petrenko21a.html
AUTHORS: Aleksei Petrenko, Erik Wijmans, Brennan Shacklett, Vladlen Koltun
HIGHLIGHT: We present Megaverse, a new 3D simulation platform for reinforcement learning and embodied AI research.

781, TITLE: Towards Practical Mean Bounds for Small Samples
http://proceedings.mlr.press/v139/phan21a.html
AUTHORS: My Phan, Philip Thomas, Erik Learned-Miller
HIGHLIGHT: For the first time since then, we present a new family of bounds that compares favorably to Anderson’s.

782, TITLE: DG-LMC: A Turn-key and Scalable Synchronous Distributed MCMC Algorithm via Langevin Monte Carlo within Gibbs
http://proceedings.mlr.press/v139/plassier21a.html
AUTHORS: Vincent Plassier, Maxime Vono, Alain Durmus, Eric Moulines
HIGHLIGHT: In this paper, we propose to fill this gap in the case where the dataset is partitioned and stored on computing nodes within a cluster under a master/slaves architecture.

783, TITLE: GeomCA: Geometric Evaluation of Data Representations
http://proceedings.mlr.press/v139/poklukar21a.html
AUTHORS: Petra Poklukar, Anastasiia Varava, Danica Kragic

HIGHLIGHT: In this work, we present Geometric Component Analysis (GeomCA) algorithm that evaluates representation spaces based on their geometric and topological properties.

784, TITLE: Grad-TTS: A Diffusion Probabilistic Model for Text-to-Speech
http://proceedings.mlr.press/v139/popov21a.html
AUTHORS: Vadim Popov, Ivan Vovk, Vladimir Gogoryan, Tasnima Sadekova, Mikhail Kudinov
HIGHLIGHT: In this paper we introduce Grad-TTS, a novel text-to-speech model with score-based decoder producing mel-spectrograms by gradually transforming noise predicted by encoder and aligned with text input by means of Monotonic Alignment Search.

785, TITLE: Bias-Free Scalable Gaussian Processes via Randomized Truncations
http://proceedings.mlr.press/v139/potapczynski21a.html
AUTHORS: Andres Potapczynski, Luhuan Wu, Dan Biderman, Geoff Pleiss, John P Cunningham
HIGHLIGHT: We find that both methods introduce a systematic bias on the learned hyperparameters: CG tends to underfit while RFF tends to overfit. We address these issues using randomized truncation estimators that eliminate bias in exchange for increased variance.

786, TITLE: Dense for the Price of Sparse: Improved Performance of Sparsely Initialized Networks via a Subspace Offset
http://proceedings.mlr.press/v139/price21a.html
AUTHORS: Ilan Price, Jared Tanner
HIGHLIGHT: In this work, we introduce a new ‘DCT plus Sparse’ layer architecture, which maintains information propagation and trainability even with as little as 0.01% trainable parameters remaining.

787, TITLE: BANG: Bridging Autoregressive and Non-autoregressive Generation with Large Scale Pretraining
http://proceedings.mlr.press/v139/qi21a.html
AUTHORS: Weizhen Qi, Yeyun Gong, Jian Jiao, Yu Yan, Weizhu Chen, Dayiheng Liu, Kewen Tang, Houqiang Li, Jiusheng Chen, Ruofei Zhang, Ming Zhou, Nan Duan
HIGHLIGHT: In this paper, we propose BANG, a new pretraining model to Bridge the gap between Autoregressive (AR) and Non-autoregressive (NAR) Generation.

788, TITLE: A Probabilistic Approach to Neural Network Pruning
http://proceedings.mlr.press/v139/qian21a.html
AUTHORS: Xin Qian, Diego Klabjan
HIGHLIGHT: Given a target network, we provide a universal approach to bound the gap between a pruned and the target network in a probabilistic sense, which is the first study of this nature.

789, TITLE: Global Prosody Style Transfer Without Text Transcriptions
http://proceedings.mlr.press/v139/qian21b.html
AUTHORS: Kaizhi Qian, Yang Zhang, Shiyu Chang, Jinjun Xiong, Chuang Gan, David Cox, Mark Hasegawa-Johnson
HIGHLIGHT: In this paper, we propose AutoPST, which can disentangle global prosody style from speech without relying on any text transcriptions.

790, TITLE: Efficient Differentiable Simulation of Articulated Bodies
http://proceedings.mlr.press/v139/qiao21a.html
AUTHORS: Yi-Ling Qiao, Junbang Liang, Vladlen Koltun, Ming C Lin
HIGHLIGHT: We present a method for efficient differentiable simulation of articulated bodies.

791, TITLE: Oneshot Differentially Private Top-k Selection
http://proceedings.mlr.press/v139/qiao21b.html
AUTHORS: Gang Qiao, Weijie Su, Li Zhang
HIGHLIGHT: In this paper, we present the oneshot Laplace mechanism, which generalizes the well-known Report Noisy Max \cite{dwork2014algorithmic} mechanism to reporting noisy top-k elements.

792, TITLE: Density Constrained Reinforcement Learning
http://proceedings.mlr.press/v139/qin21a.html
AUTHORS: Zengyi Qin, Yuxiao Chen, Chuchu Fan
HIGHLIGHT: We study constrained reinforcement learning (CRL) from a novel perspective by setting constraints directly on state density functions, rather than the value functions considered by previous works.

793, TITLE: Budgeted Heterogeneous Treatment Effect Estimation

http://proceedings.mlr.press/v139/qin21b.html
AUTHORS: Tian Qin, Tian-Zuo Wang, Zhi-Hua Zhou
HIGHLIGHT: By deriving an informative generalization bound and connecting to active learning, we propose an effective and efficient method which is validated both theoretically and empirically.

794, TITLE: Neural Transformation Learning for Deep Anomaly Detection Beyond Images
http://proceedings.mlr.press/v139/qiu21a.html
AUTHORS: Chen Qiu, Timo Pfrommer, Marius Kloft, Stephan Mandt, Maja Rudolph
HIGHLIGHT: The key idea is to embed the transformed data into a semantic space such that the transformed data still resemble their untransformed form, while different transformations are easily distinguishable.

795, TITLE: Provably Efficient Fictitious Play Policy Optimization for Zero-Sum Markov Games with Structured Transitions
http://proceedings.mlr.press/v139/qiu21b.html
AUTHORS: Shuang Qiu, Xiaohan Wei, Jieping Ye, Zhaoran Wang, Zhuoran Yang
HIGHLIGHT: We take steps forward by proposing and analyzing new fictitious play policy optimization algorithms for two-player zero-sum Markov games with structured but unknown transitions.

796, TITLE: Optimization Planning for 3D ConvNets
http://proceedings.mlr.press/v139/qiu21c.html
AUTHORS: Zhaofan Qiu, Ting Yao, Chong-Wah Ngo, Tao Mei
HIGHLIGHT: In this paper, we decompose the path into a series of training "states" and specify the hyper-parameters, e.g., learning rate and the length of input clips, in each state.

797, TITLE: On Reward-Free RL with Kernel and Neural Function Approximations: Single-Agent MDP and Markov Game
http://proceedings.mlr.press/v139/qiu21d.html
AUTHORS: Shuang Qiu, Jieping Ye, Zhaoran Wang, Zhuoran Yang
HIGHLIGHT: Specifically, we propose to explore via an optimistic variant of the value-iteration algorithm incorporating kernel and neural function approximations, where we adopt the associated exploration bonus as the exploration reward.

798, TITLE: Learning Transferable Visual Models From Natural Language Supervision
http://proceedings.mlr.press/v139/radford21a.html
AUTHORS: Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, Gretchen Krueger, Ilya Sutskever
HIGHLIGHT: We demonstrate that the simple pre-training task of predicting which caption goes with which image is an efficient and scalable way to learn SOTA image representations from scratch on a dataset of 400 million (image, text) pairs collected from the internet.

799, TITLE: A General Framework For Detecting Anomalous Inputs to DNN Classifiers
http://proceedings.mlr.press/v139/raghuram21a.html
AUTHORS: Jayaram Raghuram, Varun Chandrasekaran, Somesh Jha, Suman Banerjee
HIGHLIGHT: We propose an unsupervised anomaly detection framework based on the internal DNN layer representations in the form of a meta-algorithm with configurable components.

800, TITLE: Towards Open Ad Hoc Teamwork Using Graph-based Policy Learning
http://proceedings.mlr.press/v139/rahman21a.html
AUTHORS: Muhammad A Rahman, Niklas Hopner, Filippos Christianos, Stefano V Albrecht
HIGHLIGHT: In this work, we consider open teams by allowing agents with different fixed policies to enter and leave the environment without prior notification.

801, TITLE: Decoupling Value and Policy for Generalization in Reinforcement Learning
http://proceedings.mlr.press/v139/raileanu21a.html
AUTHORS: Roberta Raileanu, Rob Fergus
HIGHLIGHT: To alleviate this problem, we propose two approaches which are combined to create IDAAC: Invariant Decoupled Advantage Actor-Critic.

802, TITLE: Hierarchical Clustering of Data Streams: Scalable Algorithms and Approximation Guarantees
http://proceedings.mlr.press/v139/rajagopalan21a.html
AUTHORS: Anand Rajagopalan, Fabio Vitale, Danny Vainstein, Gui Citovsky, Cecilia M Procopiuc, Claudio Gentile
HIGHLIGHT: We investigate the problem of hierarchically clustering data streams containing metric data in R^d.

803, TITLE: Differentially Private Sliced Wasserstein Distance
http://proceedings.mlr.press/v139/rakotomamonjy21a.html
AUTHORS: Alain Rakotomamonjy, Ralaivola Liva
HIGHLIGHT: Our main contribution is as follows: we analyze the property of adding a Gaussian perturbation to the intrinsic randomized mechanism of the Sliced Wasserstein Distance, and we establish the sensitivity of the resulting differentially private mechanism.

804, TITLE: Zero-Shot Text-to-Image Generation
http://proceedings.mlr.press/v139/ramesh21a.html
AUTHORS: Aditya Ramesh, Mikhail Pavlov, Gabriel Goh, Scott Gray, Chelsea Voss, Alec Radford, Mark Chen, Ilya Sutskever
HIGHLIGHT: We describe a simple approach for this task based on a transformer that autoregressively models the text and image tokens as a single stream of data.

805, TITLE: End-to-End Learning of Coherent Probabilistic Forecasts for Hierarchical Time Series
http://proceedings.mlr.press/v139/rangapuram21a.html
AUTHORS: Syama Sundar Rangapuram, Lucien D Werner, Konstantinos Benidis, Pedro Mercado, Jan Gasthaus, Tim Januschowski
HIGHLIGHT: This paper presents a novel approach for hierarchical time series forecasting that produces coherent, probabilistic forecasts without requiring any explicit post-processing reconciliation.

806, TITLE: MSA Transformer
http://proceedings.mlr.press/v139/rao21a.html
AUTHORS: Roshan M Rao, Jason Liu, Robert Verkuil, Joshua Meier, John Canny, Pieter Abbeel, Tom Sercu, Alexander Rives
HIGHLIGHT: We introduce a protein language model which takes as input a set of sequences in the form of a multiple sequence alignment.

807, TITLE: Autoregressive Denoising Diffusion Models for Multivariate Probabilistic Time Series Forecasting
http://proceedings.mlr.press/v139/rasul21a.html
AUTHORS: Kashif Rasul, Calvin Seward, Ingmar Schuster, Roland Vollgraf
HIGHLIGHT: In this work, we propose TimeGrad, an autoregressive model for multivariate probabilistic time series forecasting which samples from the data distribution at each time step by estimating its gradient.

808, TITLE: Generative Particle Variational Inference via Estimation of Functional Gradients
http://proceedings.mlr.press/v139/ratzlaff21a.html
AUTHORS: Neale Ratzlaff, Qinxun Bai, Li Fuxin, Wei Xu
HIGHLIGHT: This work proposes a new method for learning to approximately sample from the posterior distribution.

809, TITLE: Enhancing Robustness of Neural Networks through Fourier Stabilization
http://proceedings.mlr.press/v139/raviv21a.html
AUTHORS: Netanel Raviv, Aidan Kelley, Minzhe Guo, Yevgeniy Vorobeychik
HIGHLIGHT: We propose a novel approach, Fourier stabilization, for designing evasion-robust neural networks with binary inputs.

810, TITLE: Disentangling Sampling and Labeling Bias for Learning in Large-output Spaces
http://proceedings.mlr.press/v139/rawat21a.html
AUTHORS: Ankit Singh Rawat, Aditya K Menon, Wittawat Jitkrittum, Sadeep Jayasumana, Felix Yu, Sashank Reddi, Sanjiv Kumar
HIGHLIGHT: In this paper, we present a new connection between these schemes and loss modification techniques for countering label imbalance.

811, TITLE: Cross-domain Imitation from Observations
http://proceedings.mlr.press/v139/raychaudhuri21a.html
AUTHORS: Dripta S. Raychaudhuri, Sujoy Paul, Jeroen Vanbaar, Amit K. Roy-Chowdhury
HIGHLIGHT: In this paper, we study the problem of how to imitate tasks when discrepancies exist between the expert and agent MDP.

812, TITLE: Implicit Regularization in Tensor Factorization
http://proceedings.mlr.press/v139/razin21a.html

AUTHORS: Noam Razin, Asaf Maman, Nadav Cohen
HIGHLIGHT: As a step further towards practical deep learning, we provide the first theoretical analysis of implicit regularization in tensor factorization — tensor completion via certain type of non-linear neural network.

813, TITLE: Align, then memorise: the dynamics of learning with feedback alignment
http://proceedings.mlr.press/v139/refinetti21a.html
AUTHORS: Maria Refinetti, St?phane D?Ascoli, Ruben Ohana, Sebastian Goldt
HIGHLIGHT: Here, we propose a theory of feedback alignment algorithms.

814, TITLE: Classifying high-dimensional Gaussian mixtures: Where kernel methods fail and neural networks succeed
http://proceedings.mlr.press/v139/refinetti21b.html
AUTHORS: Maria Refinetti, Sebastian Goldt, Florent Krzakala, Lenka Zdeborova
HIGHLIGHT: Here, we show that two-layer neural networks with *only a few neurons* achieve near-optimal performance on high-dimensional Gaussian mixture classification while lazy training approaches such as random features and kernel methods do not.

815, TITLE: Sharf: Shape-conditioned Radiance Fields from a Single View
http://proceedings.mlr.press/v139/rematas21a.html
AUTHORS: Konstantinos Rematas, Ricardo Martin-Brualla, Vittorio Ferrari
HIGHLIGHT: We present a method for estimating neural scenes representations of objects given only a single image.

816, TITLE: LEGO: Latent Execution-Guided Reasoning for Multi-Hop Question Answering on Knowledge Graphs
http://proceedings.mlr.press/v139/ren21a.html
AUTHORS: Hongyu Ren, Hanjun Dai, Bo Dai, Xinyun Chen, Michihiro Yasunaga, Haitian Sun, Dale Schuurmans, Jure Leskovec, Denny Zhou
HIGHLIGHT: Here we present LEGO, a Latent Execution-Guided reasOning framework to handle this challenge in KGQA.

817, TITLE: Interpreting and Disentangling Feature Components of Various Complexity from DNNs
http://proceedings.mlr.press/v139/ren21b.html
AUTHORS: Jie Ren, Mingjie Li, Zexu Liu, Quanshi Zhang
HIGHLIGHT: This paper aims to define, visualize, and analyze the feature complexity that is learned by a DNN.

818, TITLE: Integrated Defense for Resilient Graph Matching
http://proceedings.mlr.press/v139/ren21c.html
AUTHORS: Jiaxiang Ren, Zijie Zhang, Jiayin Jin, Xin Zhao, Sixing Wu, Yang Zhou, Yelong Shen, Tianshi Che, Ruoming Jin, Dejing Dou
HIGHLIGHT: In this paper, we identify and study two types of unique topology attacks in graph matching: inter-graph dispersion and intra-graph assembly attacks.

819, TITLE: Solving high-dimensional parabolic PDEs using the tensor train format
http://proceedings.mlr.press/v139/richter21a.html
AUTHORS: Lorenz Richter, Leon Sallandt, Nikolas N?sken
HIGHLIGHT: In this paper, we argue that tensor trains provide an appealing approximation framework for parabolic PDEs: the combination of reformulations in terms of backward stochastic differential equations and regression-type methods in the tensor format holds the promise of leveraging latent low-rank structures enabling both compression and efficient computation.

820, TITLE: Best Arm Identification in Graphical Bilinear Bandits
http://proceedings.mlr.press/v139/rizk21a.html
AUTHORS: Geovani Rizk, Albert Thomas, Igor Colin, Rida Laraki, Yann Chevaleyre
HIGHLIGHT: We introduce a new graphical bilinear bandit problem where a learner (or a \emph{central entity}) allocates arms to the nodes of a graph and observes for each edge a noisy bilinear reward representing the interaction between the two end nodes.

821, TITLE: Principled Simplicial Neural Networks for Trajectory Prediction
http://proceedings.mlr.press/v139/roddenberry21a.html
AUTHORS: T. Mitchell Roddenberry, Nicholas Glaze, Santiago Segarra
HIGHLIGHT: Based on these properties, we propose a simple convolutional architecture, rooted in tools from algebraic topology, for the problem of trajectory prediction, and show that it obeys all three of these properties when an odd, nonlinear activation function is used.

822, TITLE: On Linear Identifiability of Learned Representations

http://proceedings.mlr.press/v139/roeder21a.html
AUTHORS: Geoffrey Roeder, Luke Metz, Durk Kingma
HIGHLIGHT: In this paper, building on recent advances in nonlinear Independent Components Analysis, we aim to rehabilitate identifiability by showing that a large family of discriminative models are in fact identifiable in function space, up to a linear indeterminacy.

823, TITLE: Representation Matters: Assessing the Importance of Subgroup Allocations in Training Data
http://proceedings.mlr.press/v139/rolf21a.html
AUTHORS: Esther Rolf, Theodora T Worledge, Benjamin Recht, Michael Jordan
HIGHLIGHT: Our analysis and experiments describe how dataset compositions influence performance and provide constructive results for using trends in existing data, alongside domain knowledge, to help guide intentional, objective-aware dataset design

824, TITLE: TeachMyAgent: a Benchmark for Automatic Curriculum Learning in Deep RL
http://proceedings.mlr.press/v139/romac21a.html
AUTHORS: Cl?ment Romac, R?my Portelas, Katja Hofmann, Pierre-Yves Oudeyer
HIGHLIGHT: In this work, we identify several key challenges faced by ACL algorithms.

825, TITLE: Discretization Drift in Two-Player Games
http://proceedings.mlr.press/v139/rosca21a.html
AUTHORS: Mihaela C Rosca, Yan Wu, Benoit Dherin, David Barrett
HIGHLIGHT: Using backward error analysis, we derive modified continuous dynamical systems that closely follow the discrete dynamics.

826, TITLE: On the Predictability of Pruning Across Scales
http://proceedings.mlr.press/v139/rosenfeld21a.html
AUTHORS: Jonathan S Rosenfeld, Jonathan Frankle, Michael Carbin, Nir Shavit
HIGHLIGHT: We show that the error of iteratively magnitude-pruned networks empirically follows a scaling law with interpretable coefficients that depend on the architecture and task.

827, TITLE: Benchmarks, Algorithms, and Metrics for Hierarchical Disentanglement
http://proceedings.mlr.press/v139/ross21a.html
AUTHORS: Andrew Ross, Finale Doshi-Velez
HIGHLIGHT: In this work, we develop benchmarks, algorithms, and metrics for learning such hierarchical representations.

828, TITLE: Simultaneous Similarity-based Self-Distillation for Deep Metric Learning
http://proceedings.mlr.press/v139/roth21a.html
AUTHORS: Karsten Roth, Timo Milbich, Bjorn Ommer, Joseph Paul Cohen, Marzyeh Ghassemi
HIGHLIGHT: To remedy this, we propose S2SD - Simultaneous Similarity-based Self-distillation.

829, TITLE: Multi-group Agnostic PAC Learnability
http://proceedings.mlr.press/v139/rothblum21a.html
AUTHORS: Guy N Rothblum, Gal Yona
HIGHLIGHT: Motivated by such fairness concerns, we study "multi-group agnostic PAC learnability": fixing a measure of loss, a benchmark class \H and a (potentially) rich collection of subgroups \G, the objective is to learn a single predictor such that the loss experienced by every group g?\G is not much larger than the best possible loss for this group within \H.

830, TITLE: PACOH: Bayes-Optimal Meta-Learning with PAC-Guarantees
http://proceedings.mlr.press/v139/rothfuss21a.html
AUTHORS: Jonas Rothfuss, Vincent Fortuin, Martin Josifoski, Andreas Krause
HIGHLIGHT: We provide a theoretical analysis using the PAC-Bayesian framework and derive novel generalization bounds for meta-learning.

831, TITLE: An Algorithm for Stochastic and Adversarial Bandits with Switching Costs
http://proceedings.mlr.press/v139/rouyer21a.html
AUTHORS: Chlo? Rouyer, Yevgeny Seldin, Nicol? Cesa-Bianchi
HIGHLIGHT: We propose an algorithm for stochastic and adversarial multiarmed bandits with switching costs, where the algorithm pays a price λ every time it switches the arm being played.

832, TITLE: Improving Lossless Compression Rates via Monte Carlo Bits-Back Coding

http://proceedings.mlr.press/v139/ruan21a.html
AUTHORS: Yangjun Ruan, Karen Ullrich, Daniel S Severo, James Townsend, Ashish Khisti, Arnaud Doucet, Alireza Makhzani, Chris Maddison
HIGHLIGHT: In this paper, we show how to remove this gap asymptotically by deriving bits-back coding algorithms from tighter variational bounds.

833, TITLE: On Signal-to-Noise Ratio Issues in Variational Inference for Deep Gaussian Processes
http://proceedings.mlr.press/v139/rudner21a.html
AUTHORS: Tim G. J. Rudner, Oscar Key, Yarin Gal, Tom Rainforth
HIGHLIGHT: We show that the gradient estimates used in training Deep Gaussian Processes (DGPs) with importance-weighted variational inference are susceptible to signal-to-noise ratio (SNR) issues.

834, TITLE: Tilting the playing field: Dynamical loss functions for machine learning
http://proceedings.mlr.press/v139/ruiz-garcia21a.html
AUTHORS: Miguel Ruiz-Garcia, Ge Zhang, Samuel S Schoenholz, Andrea J. Liu
HIGHLIGHT: We show that learning can be improved by using loss functions that evolve cyclically during training to emphasize one class at a time.

835, TITLE: UnICORNN: A recurrent model for learning very long time dependencies
http://proceedings.mlr.press/v139/rusch21a.html
AUTHORS: T. Konstantin Rusch, Siddhartha Mishra
HIGHLIGHT: To overcome this, we propose a novel RNN architecture which is based on a structure preserving discretization of a Hamiltonian system of second-order ordinary differential equations that models networks of oscillators.

836, TITLE: Simple and Effective VAE Training with Calibrated Decoders
http://proceedings.mlr.press/v139/rybkin21a.html
AUTHORS: Oleh Rybkin, Kostas Daniilidis, Sergey Levine
HIGHLIGHT: We study the impact of calibrated decoders, which learn the uncertainty of the decoding distribution and can determine this amount of information automatically, on the VAE performance.

837, TITLE: Model-Based Reinforcement Learning via Latent-Space Collocation
http://proceedings.mlr.press/v139/rybkin21b.html
AUTHORS: Oleh Rybkin, Chuning Zhu, Anusha Nagabandi, Kostas Daniilidis, Igor Mordatch, Sergey Levine
HIGHLIGHT: In this work, we study how the long-horizon planning abilities can be improved with an algorithm that optimizes over sequences of states, rather than actions, which allows better credit assignment.

838, TITLE: Training Data Subset Selection for Regression with Controlled Generalization Error
http://proceedings.mlr.press/v139/s21a.html
AUTHORS: Durga S, Rishabh Iyer, Ganesh Ramakrishnan, Abir De
HIGHLIGHT: In this paper, our goal is to design an algorithm for selecting a subset of the training data, so that the model can be trained quickly, without significantly sacrificing on accuracy.

839, TITLE: Unsupervised Part Representation by Flow Capsules
http://proceedings.mlr.press/v139/sabour21a.html
AUTHORS: Sara Sabour, Andrea Tagliasacchi, Soroosh Yazdani, Geoffrey Hinton, David J Fleet
HIGHLIGHT: To address this issue we propose a way to learn primary capsule encoders that detect atomic parts from a single image.

840, TITLE: Stochastic Sign Descent Methods: New Algorithms and Better Theory
http://proceedings.mlr.press/v139/safaryan21a.html
AUTHORS: Mher Safaryan, Peter Richtarik
HIGHLIGHT: In this paper, we analyze sign-based methods for non-convex optimization in three key settings: (i) standard single node, (ii) parallel with shared data and (iii) distributed with partitioned data.

841, TITLE: Adversarial Dueling Bandits
http://proceedings.mlr.press/v139/saha21a.html
AUTHORS: Aadirupa Saha, Tomer Koren, Yishay Mansour
HIGHLIGHT: We introduce the problem of regret minimization in Adversarial Dueling Bandits.

842, TITLE: Dueling Convex Optimization

http://proceedings.mlr.press/v139/saha21b.html
AUTHORS: Aadirupa Saha, Tomer Koren, Yishay Mansour
HIGHLIGHT: We address the problem of convex optimization with preference (dueling) feedback.

843, TITLE: Optimal regret algorithm for Pseudo-1d Bandit Convex Optimization
http://proceedings.mlr.press/v139/saha21c.html
AUTHORS: Aadirupa Saha, Nagarajan Natarajan, Praneeth Netrapalli, Prateek Jain
HIGHLIGHT: We propose a new algorithm \sbcalg that combines randomized online gradient descent with a kernelized exponential weights method to exploit the pseudo-1d structure effectively, guaranteeing the {\em optimal} regret bound mentioned above, up to additional logarithmic factors.

844, TITLE: Asymptotics of Ridge Regression in Convolutional Models
http://proceedings.mlr.press/v139/sahraee-ardakan21a.html
AUTHORS: Mojtaba Sahraee-Ardakan, Tung Mai, Anup Rao, Ryan A. Rossi, Sundeep Rangan, Alyson K Fletcher
HIGHLIGHT: In this work, we analyze the asymptotics of estimation error in ridge estimators for convolutional linear models.

845, TITLE: Momentum Residual Neural Networks
http://proceedings.mlr.press/v139/sander21a.html
AUTHORS: Michael E. Sander, Pierre Ablin, Mathieu Blondel, Gabriel Peyr?
HIGHLIGHT: In this paper, we propose to change the forward rule of a ResNet by adding a momentum term.

846, TITLE: Meta-Learning Bidirectional Update Rules
http://proceedings.mlr.press/v139/sandler21a.html
AUTHORS: Mark Sandler, Max Vladymyrov, Andrey Zhmoginov, Nolan Miller, Tom Madams, Andrew Jackson, Blaise Ag?era Y Arcas
HIGHLIGHT: In this paper, we introduce a new type of generalized neural network where neurons and synapses maintain multiple states.

847, TITLE: Recomposing the Reinforcement Learning Building Blocks with Hypernetworks
http://proceedings.mlr.press/v139/sarafian21a.html
AUTHORS: Elad Sarafian, Shai Keynan, Sarit Kraus
HIGHLIGHT: Standard architectures tend to ignore these variables' underlying interpretations and simply concatenate their features into a single vector. In this work, we argue that this choice may lead to poor gradient estimation in actor-critic algorithms and high variance learning steps in Meta-RL algorithms.

848, TITLE: Towards Understanding Learning in Neural Networks with Linear Teachers
http://proceedings.mlr.press/v139/sarussi21a.html
AUTHORS: Roei Sarussi, Alon Brutzkus, Amir Globerson
HIGHLIGHT: Here we prove that SGD globally optimizes this learning problem for a two-layer network with Leaky ReLU activations.

849, TITLE: E(n) Equivariant Graph Neural Networks
http://proceedings.mlr.press/v139/satorras21a.html
AUTHORS: Vi?ctor Garcia Satorras, Emiel Hoogeboom, Max Welling
HIGHLIGHT: This paper introduces a new model to learn graph neural networks equivariant to rotations, translations, reflections and permutations called E(n)-Equivariant Graph Neural Networks (EGNNs).

850, TITLE: A Representation Learning Perspective on the Importance of Train-Validation Splitting in Meta-Learning
http://proceedings.mlr.press/v139/saunshi21a.html
AUTHORS: Nikunj Saunshi, Arushi Gupta, Wei Hu
HIGHLIGHT: We present theoretical results that formalize this idea for linear representation learning on a subspace meta-learning instance, and experimentally verify this practical benefit of splitting in simulations and on standard meta-learning benchmarks.

851, TITLE: Low-Rank Sinkhorn Factorization
http://proceedings.mlr.press/v139/scetbon21a.html
AUTHORS: Meyer Scetbon, Marco Cuturi, Gabriel Peyr?
HIGHLIGHT: Building on this, we introduce in this work a generic approach that aims at solving, in full generality, the OT problem under low-nonnegative rank constraints with arbitrary costs.

852, TITLE: Linear Transformers Are Secretly Fast Weight Programmers
http://proceedings.mlr.press/v139/schlag21a.html
AUTHORS: Imanol Schlag, Kazuki Irie, J?rgen Schmidhuber
HIGHLIGHT: We show the formal equivalence of linearised self-attention mechanisms and fast weight controllers from the early ’90s, where a slow neural net learns by gradient descent to program the fast weights of another net through sequences of elementary programming instructions which are additive outer products of self-invented activation patterns (today called keys and values).

853, TITLE: Descending through a Crowded Valley - Benchmarking Deep Learning Optimizers
http://proceedings.mlr.press/v139/schmidt21a.html
AUTHORS: Robin M Schmidt, Frank Schneider, Philipp Hennig
HIGHLIGHT: In this work, we aim to replace these anecdotes, if not with a conclusive ranking, then at least with evidence-backed heuristics.

854, TITLE: Equivariant message passing for the prediction of tensorial properties and molecular spectra
http://proceedings.mlr.press/v139/schutt21a.html
AUTHORS: Kristof Sch?tt, Oliver Unke, Michael Gastegger
HIGHLIGHT: On this basis, we propose the polarizable atom interaction neural network (PaiNN) and improve on common molecule benchmarks over previous networks, while reducing model size and inference time.

855, TITLE: Just How Toxic is Data Poisoning? A Unified Benchmark for Backdoor and Data Poisoning Attacks
http://proceedings.mlr.press/v139/schwarzschild21a.html
AUTHORS: Avi Schwarzschild, Micah Goldblum, Arjun Gupta, John P Dickerson, Tom Goldstein
HIGHLIGHT: We observe that data poisoning and backdoor attacks are highly sensitive to variations in the testing setup.

856, TITLE: Connecting Sphere Manifolds Hierarchically for Regularization
http://proceedings.mlr.press/v139/scieur21a.html
AUTHORS: Damien Scieur, Youngsung Kim
HIGHLIGHT: This paper considers classification problems with hierarchically organized classes.

857, TITLE: Learning Intra-Batch Connections for Deep Metric Learning
http://proceedings.mlr.press/v139/seidenschwarz21a.html
AUTHORS: Jenny Denise Seidenschwarz, Ismail Elezi, Laura Leal-Taix?
HIGHLIGHT: To this end, we propose an approach based on message passing networks that takes all the relations in a mini-batch into account.

858, TITLE: Top-k eXtreme Contextual Bandits with Arm Hierarchy
http://proceedings.mlr.press/v139/sen21a.html
AUTHORS: Rajat Sen, Alexander Rakhlin, Lexing Ying, Rahul Kidambi, Dean Foster, Daniel N Hill, Inderjit S. Dhillon
HIGHLIGHT: Motivated by modern applications, such as online advertisement and recommender systems, we study the top-k extreme contextual bandits problem, where the total number of arms can be enormous, and the learner is allowed to select k arms and observe all or some of the rewards for the chosen arms.

859, TITLE: Pure Exploration and Regret Minimization in Matching Bandits
http://proceedings.mlr.press/v139/sentenac21a.html
AUTHORS: Flore Sentenac, Jialin Yi, Clement Calauzenes, Vianney Perchet, Milan Vojnovic
HIGHLIGHT: We prove that it is possible to leverage a rank-1 assumption on the adjacency matrix to reduce the sample complexity and the regret of off-the-shelf algorithms up to reaching a linear dependency in the number of vertices (up to to poly-log terms).

860, TITLE: State Entropy Maximization with Random Encoders for Efficient Exploration
http://proceedings.mlr.press/v139/seo21a.html
AUTHORS: Younggyo Seo, Lili Chen, Jinwoo Shin, Honglak Lee, Pieter Abbeel, Kimin Lee
HIGHLIGHT: This paper presents Random Encoders for Efficient Exploration (RE3), an exploration method that utilizes state entropy as an intrinsic reward.

861, TITLE: Online Submodular Resource Allocation with Applications to Rebalancing Shared Mobility Systems
http://proceedings.mlr.press/v139/sessa21a.html
AUTHORS: Pier Giuseppe Sessa, Ilija Bogunovic, Andreas Krause, Maryam Kamgarpour
HIGHLIGHT: We propose a distributed scheme to maximize the cumulative welfare by designing a repeated game among the agents, who learn to act via regret minimization.

862, TITLE: RRL: Resnet as representation for Reinforcement Learning
http://proceedings.mlr.press/v139/shah21a.html
AUTHORS: Rutav M Shah, Vikash Kumar
HIGHLIGHT: We propose RRL: Resnet as representation for Reinforcement Learning {–} a straightforward yet effective approach that can learn complex behaviors directly from proprioceptive inputs.

863, TITLE: Equivariant Networks for Pixelized Spheres
http://proceedings.mlr.press/v139/shakerinava21a.html
AUTHORS: Mehran Shakerinava, Siamak Ravanbakhsh
HIGHLIGHT: We show how to model this interplay using ideas from group theory, identify the equivariant linear maps, and introduce equivariant padding that respects these symmetries.

864, TITLE: Personalized Federated Learning using Hypernetworks
http://proceedings.mlr.press/v139/shamsian21a.html
AUTHORS: Aviv Shamsian, Aviv Navon, Ethan Fetaya, Gal Chechik
HIGHLIGHT: We propose a novel approach to this problem using hypernetworks, termed pFedHN for personalized Federated HyperNetworks.

865, TITLE: On the Power of Localized Perceptron for Label-Optimal Learning of Halfspaces with Adversarial Noise
http://proceedings.mlr.press/v139/shen21a.html
AUTHORS: Jie Shen
HIGHLIGHT: Our main contribution is a Perceptron-like online active learning algorithm that runs in polynomial time, and under the conditions that the marginal distribution is isotropic log-concave and $\nu = \Omega(\epsilon)$, where $\epsilon \in (0, 1)$ is the target error rate, our algorithm PAC learns the underlying halfspace with near-optimal label complexity of $\tilde{O}\big(d \cdot \polylog(\frac{1}{\epsilon})\big)$ and sample complexity of $\tilde{O}\big(\frac{d}{\epsilon} \big)$.

866, TITLE: Sample-Optimal PAC Learning of Halfspaces with Malicious Noise
http://proceedings.mlr.press/v139/shen21b.html
AUTHORS: Jie Shen
HIGHLIGHT: In this work, we present a new analysis for the algorithm of Awasthi et al. (2017) and show that it essentially achieves the near-optimal sample complexity bound of $\tilde{O}(d)$, improving the best known result of $\tilde{O}(d^2)$.

867, TITLE: Backdoor Scanning for Deep Neural Networks through K-Arm Optimization
http://proceedings.mlr.press/v139/shen21c.html
AUTHORS: Guangyu Shen, Yingqi Liu, Guanhong Tao, Shengwei An, Qiuling Xu, Siyuan Cheng, Shiqing Ma, Xiangyu Zhang
HIGHLIGHT: Inspired by Multi-Arm Bandit in Reinforcement Learning, we propose a K-Arm optimization method for backdoor detection.

868, TITLE: State Relevance for Off-Policy Evaluation
http://proceedings.mlr.press/v139/shen21d.html
AUTHORS: Simon P Shen, Yecheng Ma, Omer Gottesman, Finale Doshi-Velez
HIGHLIGHT: In this work, we introduce Omitting-States-Irrelevant-to-Return Importance Sampling (OSIRIS), an estimator which reduces variance by strategically omitting likelihood ratios associated with certain states.

869, TITLE: SparseBERT: Rethinking the Importance Analysis in Self-attention
http://proceedings.mlr.press/v139/shi21a.html
AUTHORS: Han Shi, Jiahui Gao, Xiaozhe Ren, Hang Xu, Xiaodan Liang, Zhenguo Li, James Tin-Yau Kwok
HIGHLIGHT: To rethink the importance analysis in self-attention, we study the significance of different positions in attention matrix during pre-training.

870, TITLE: Learning Gradient Fields for Molecular Conformation Generation
http://proceedings.mlr.press/v139/shi21b.html
AUTHORS: Chence Shi, Shitong Luo, Minkai Xu, Jian Tang
HIGHLIGHT: Inspired by the traditional force field methods for molecular dynamics simulation, in this paper, we propose a novel approach called ConfGF by directly estimating the gradient fields of the log density of atomic coordinates.

871, TITLE: Segmenting Hybrid Trajectories using Latent ODEs
http://proceedings.mlr.press/v139/shi21c.html

AUTHORS: Ruian Shi, Quaid Morris
HIGHLIGHT: Here, we propose the Latent Segmented ODE (LatSegODE), which uses Latent ODEs to perform reconstruction and changepoint detection within hybrid trajectories featuring jump discontinuities and switching dynamical modes.

872, TITLE: Deeply-Debiased Off-Policy Interval Estimation
http://proceedings.mlr.press/v139/shi21d.html
AUTHORS: Chengchun Shi, Runzhe Wan, Victor Chernozhukov, Rui Song
HIGHLIGHT: In this paper, we propose a novel procedure to construct an efficient, robust, and flexible CI on a target policy's value.

873, TITLE: GANMEX: One-vs-One Attributions using GAN-based Model Explainability
http://proceedings.mlr.press/v139/shih21a.html
AUTHORS: Sheng-Min Shih, Pin-Ju Tien, Zohar Karnin
HIGHLIGHT: In this paper, we present GANMEX, a novel approach applying Generative Adversarial Networks (GAN) by incorporating the to-be-explained classifier as part of the adversarial networks.

874, TITLE: Large-Scale Meta-Learning with Continual Trajectory Shifting
http://proceedings.mlr.press/v139/shin21a.html
AUTHORS: Jaewoong Shin, Hae Beom Lee, Boqing Gong, Sung Ju Hwang
HIGHLIGHT: In this paper, we first show that allowing the meta-learners to take a larger number of inner gradient steps better captures the structure of heterogeneous and large-scale task distributions, thus results in obtaining better initialization points. Further, in order to increase the frequency of meta-updates even with the excessively long inner-optimization trajectories, we propose to estimate the required shift of the task-specific parameters with respect to the change of the initialization parameters.

875, TITLE: AGENT: A Benchmark for Core Psychological Reasoning
http://proceedings.mlr.press/v139/shu21a.html
AUTHORS: Tianmin Shu, Abhishek Bhandwaldar, Chuang Gan, Kevin Smith, Shari Liu, Dan Gutfreund, Elizabeth Spelke, Joshua Tenenbaum, Tomer Ullman
HIGHLIGHT: Inspired by cognitive development studies on intuitive psychology, we present a benchmark consisting of a large dataset of procedurally generated 3D animations, AGENT (Action, Goal, Efficiency, coNstraint, uTility), structured around four scenarios (goal preferences, action efficiency, unobserved constraints, and cost-reward trade-offs) that probe key concepts of core intuitive psychology.

876, TITLE: Zoo-Tuning: Adaptive Transfer from A Zoo of Models
http://proceedings.mlr.press/v139/shu21b.html
AUTHORS: Yang Shu, Zhi Kou, Zhangjie Cao, Jianmin Wang, Mingsheng Long
HIGHLIGHT: We propose \emph{Zoo-Tuning} to address these challenges, which learns to adaptively transfer the parameters of pretrained models to the target task.

877, TITLE: Aggregating From Multiple Target-Shifted Sources
http://proceedings.mlr.press/v139/shui21a.html
AUTHORS: Changjian Shui, Zijian Li, Jiaqi Li, Christian Gagn?, Charles X Ling, Boyu Wang
HIGHLIGHT: In this paper, we analyzed the problem for aggregating source domains with different label distributions, where most recent source selection approaches fail.

878, TITLE: Testing Group Fairness via Optimal Transport Projections
http://proceedings.mlr.press/v139/si21a.html
AUTHORS: Nian Si, Karthyek Murthy, Jose Blanchet, Viet Anh Nguyen
HIGHLIGHT: We have developed a statistical testing framework to detect if a given machine learning classifier fails to satisfy a wide range of group fairness notions.

879, TITLE: On Characterizing GAN Convergence Through Proximal Duality Gap
http://proceedings.mlr.press/v139/sidheekh21a.html
AUTHORS: Sahil Sidheekh, Aroof Aimen, Narayanan C Krishnan
HIGHLIGHT: In this work, we extend the notion of duality gap to proximal duality gap that is applicable to the general context of training GANs where Nash equilibria may not exist.

880, TITLE: A Precise Performance Analysis of Support Vector Regression
http://proceedings.mlr.press/v139/sifaou21a.html
AUTHORS: Houssem Sifaou, Abla Kammoun, Mohamed-Slim Alouini

HIGHLIGHT: In this paper, we study the hard and soft support vector regression techniques applied to a set of n linear measurements of the form $y_i=\boldsymbol{\beta}_\star^{T}{\bf x}_i +n_i$ where $\boldsymbol{\beta}_\star$ is an unknown vector, $\left\{{\bf x}_i\right\}_{i=1}^n$ are the feature vectors and $\left\{{n}_i\right\}_{i=1}^n$ model the noise.

881, TITLE: Directed Graph Embeddings in Pseudo-Riemannian Manifolds
http://proceedings.mlr.press/v139/sim21a.html
AUTHORS: Aaron Sim, Maciej L Wiatrak, Angus Brayne, Paidi Creed, Saee Paliwal
HIGHLIGHT: In this paper, we show that general directed graphs can be effectively represented by an embedding model that combines three components: a pseudo-Riemannian metric structure, a non-trivial global topology, and a unique likelihood function that explicitly incorporates a preferred direction in embedding space.

882, TITLE: Collaborative Bayesian Optimization with Fair Regret
http://proceedings.mlr.press/v139/sim21b.html
AUTHORS: Rachael Hwee Ling Sim, Yehong Zhang, Bryan Kian Hsiang Low, Patrick Jaillet
HIGHLIGHT: Inspired by social welfare concepts from economics, we propose a new notion of regret capturing these properties and a collaborative BO algorithm whose convergence rate can be theoretically guaranteed by bounding the new regret, both of which share an adjustable parameter for trading off between fairness vs. efficiency.

883, TITLE: Dynamic Planning and Learning under Recovering Rewards
http://proceedings.mlr.press/v139/simchi-levi21a.html
AUTHORS: David Simchi-Levi, Zeyu Zheng, Feng Zhu
HIGHLIGHT: With the objective of maximizing expected cumulative rewards over T time periods, we propose, construct and prove performance guarantees for a class of “Purely Periodic Policies”.

884, TITLE: PopSkipJump: Decision-Based Attack for Probabilistic Classifiers
http://proceedings.mlr.press/v139/simon-gabriel21a.html
AUTHORS: Carl-Johann Simon-Gabriel, Noman Ahmed Sheikh, Andreas Krause
HIGHLIGHT: We therefore propose a new adversarial decision-based attack specifically designed for classifiers with probabilistic outputs.

885, TITLE: Geometry of the Loss Landscape in Overparameterized Neural Networks: Symmetries and Invariances
http://proceedings.mlr.press/v139/simsek21a.html
AUTHORS: Berfin Simsek, Fran?ois Ged, Arthur Jacot, Francesco Spadaro, Clement Hongler, Wulfram Gerstner, Johanni Brea
HIGHLIGHT: We study how permutation symmetries in overparameterized multi-layer neural networks generate ‘symmetry-induced’ critical points.

886, TITLE: Flow-based Attribution in Graphical Models: A Recursive Shapley Approach
http://proceedings.mlr.press/v139/singal21a.html
AUTHORS: Raghav Singal, George Michailidis, Hoiyi Ng
HIGHLIGHT: We study the attribution problem in a graphical model, wherein the objective is to quantify how the effect of changes at the source nodes propagates through the graph.

887, TITLE: Structured World Belief for Reinforcement Learning in POMDP
http://proceedings.mlr.press/v139/singh21a.html
AUTHORS: Gautam Singh, Skand Peri, Junghyun Kim, Hyunseok Kim, Sungjin Ahn
HIGHLIGHT: In this paper, we propose Structured World Belief, a model for learning and inference of object-centric belief states.

888, TITLE: Skew Orthogonal Convolutions
http://proceedings.mlr.press/v139/singla21a.html
AUTHORS: Sahil Singla, Soheil Feizi
HIGHLIGHT: In this work, we propose a GNP convolution layer called \textbf{S}kew \textbf{O}rthogonal \textbf{C}onvolution (SOC) that uses the following mathematical property: when a matrix is {\it Skew-Symmetric}, its exponential function is an {\it orthogonal} matrix.

889, TITLE: Multi-Task Reinforcement Learning with Context-based Representations
http://proceedings.mlr.press/v139/sodhani21a.html
AUTHORS: Shagun Sodhani, Amy Zhang, Joelle Pineau
HIGHLIGHT: In this framework, metadata can help to learn interpretable representations and provide the context to inform which representations to compose and how to compose them.

890, TITLE: Shortest-Path Constrained Reinforcement Learning for Sparse Reward Tasks
http://proceedings.mlr.press/v139/sohn21a.html
AUTHORS: Sungryull Sohn, Sungtae Lee, Jongwook Choi, Harm H Van Seijen, Mehdi Fatemi, Honglak Lee
HIGHLIGHT: We propose the k-Shortest-Path (k-SP) constraint: a novel constraint on the agent's trajectory that improves the sample efficiency in sparse-reward MDPs.

891, TITLE: Accelerating Feedforward Computation via Parallel Nonlinear Equation Solving
http://proceedings.mlr.press/v139/song21a.html
AUTHORS: Yang Song, Chenlin Meng, Renjie Liao, Stefano Ermon
HIGHLIGHT: To enable parallelization, we frame the task of feedforward computation as solving a system of nonlinear equations.

892, TITLE: PC-MLP: Model-based Reinforcement Learning with Policy Cover Guided Exploration
http://proceedings.mlr.press/v139/song21b.html
AUTHORS: Yuda Song, Wen Sun
HIGHLIGHT: This work studies a computationally and statistically efficient model-based algorithm for both Kernelized Nonlinear Regulators (KNR) and linear Markov Decision Processes (MDPs).

893, TITLE: Fast Sketching of Polynomial Kernels of Polynomial Degree
http://proceedings.mlr.press/v139/song21c.html
AUTHORS: Zhao Song, David Woodruff, Zheng Yu, Lichen Zhang
HIGHLIGHT: Combined with a novel sampling scheme, we give the fastest algorithms for approximating a large family of slow-growing kernels.

894, TITLE: Variance Reduction via Primal-Dual Accelerated Dual Averaging for Nonsmooth Convex Finite-Sums
http://proceedings.mlr.press/v139/song21d.html
AUTHORS: Chaobing Song, Stephen J Wright, Jelena Diakonikolas
HIGHLIGHT: For the primal-dual formulation of this problem, we propose a novel algorithm called \emph{Variance Reduction via Primal-Dual Accelerated Dual Averaging (\vrpda)}.

895, TITLE: Oblivious Sketching-based Central Path Method for Linear Programming
http://proceedings.mlr.press/v139/song21e.html
AUTHORS: Zhao Song, Zheng Yu
HIGHLIGHT: In this work, we propose a sketching-based central path method for solving linear programmings, whose running time matches the state of the art results [Cohen, Lee, Song STOC 19; Lee, Song, Zhang COLT 19].

896, TITLE: Causal Curiosity: RL Agents Discovering Self-supervised Experiments for Causal Representation Learning
http://proceedings.mlr.press/v139/sontakke21a.html
AUTHORS: Sumedh A Sontakke, Arash Mehrjou, Laurent Itti, Bernhard Sch?lkopf
HIGHLIGHT: We introduce a novel intrinsic reward, called causal curiosity, and show that it allows our agents to learn optimal sequences of actions, and to discover causal factors in the dynamics.

897, TITLE: Decomposed Mutual Information Estimation for Contrastive Representation Learning
http://proceedings.mlr.press/v139/sordoni21a.html
AUTHORS: Alessandro Sordoni, Nouha Dziri, Hannes Schulz, Geoff Gordon, Philip Bachman, Remi Tachet Des Combes
HIGHLIGHT: We propose decomposing the full MI estimation problem into a sum of smaller estimation problems by splitting one of the views into progressively more informed subviews and by applying the chain rule on MI between the decomposed views.

898, TITLE: Decoupling Representation Learning from Reinforcement Learning
http://proceedings.mlr.press/v139/stooke21a.html
AUTHORS: Adam Stooke, Kimin Lee, Pieter Abbeel, Michael Laskin
HIGHLIGHT: In an effort to overcome limitations of reward-driven feature learning in deep reinforcement learning (RL) from images, we propose decoupling representation learning from policy learning.

899, TITLE: K-shot NAS: Learnable Weight-Sharing for NAS with K-shot Supernets
http://proceedings.mlr.press/v139/su21a.html
AUTHORS: Xiu Su, Shan You, Mingkai Zheng, Fei Wang, Chen Qian, Changshui Zhang, Chang Xu
HIGHLIGHT: In this paper, instead of counting on a single supernet, we introduce K-shot supernets and take their weights for each operation as a dictionary.

900, TITLE: More Powerful and General Selective Inference for Stepwise Feature Selection using Homotopy Method
http://proceedings.mlr.press/v139/sugiyama21a.html
AUTHORS: Kazuya Sugiyama, Vo Nguyen Le Duy, Ichiro Takeuchi
HIGHLIGHT: In this study, we develop a more powerful and general conditional SI method for SFS using the homotopy method which enables us to overcome this limitation.

901, TITLE: Not All Memories are Created Equal: Learning to Forget by Expiring
http://proceedings.mlr.press/v139/sukhbaatar21a.html
AUTHORS: Sainbayar Sukhbaatar, Da Ju, Spencer Poff, Stephen Roller, Arthur Szlam, Jason Weston, Angela Fan
HIGHLIGHT: We propose Expire-Span, a method that learns to retain the most important information and expire the irrelevant information.

902, TITLE: Nondeterminism and Instability in Neural Network Optimization
http://proceedings.mlr.press/v139/summers21a.html
AUTHORS: Cecilia Summers, Michael J. Dinneen
HIGHLIGHT: In this work, we establish an experimental protocol for understanding the effect of optimization nondeterminism on model diversity, allowing us to isolate the effects of a variety of sources of nondeterminism.

903, TITLE: AutoSampling: Search for Effective Data Sampling Schedules
http://proceedings.mlr.press/v139/sun21a.html
AUTHORS: Ming Sun, Haoxuan Dou, Baopu Li, Junjie Yan, Wanli Ouyang, Lei Cui
HIGHLIGHT: In this paper, we propose an AutoSampling method to automatically learn sampling schedules for model training, which consists of the multi-exploitation step aiming for optimal local sampling schedules and the exploration step for the ideal sampling distribution.

904, TITLE: What Makes for End-to-End Object Detection?
http://proceedings.mlr.press/v139/sun21b.html
AUTHORS: Peize Sun, Yi Jiang, Enze Xie, Wenqi Shao, Zehuan Yuan, Changhu Wang, Ping Luo
HIGHLIGHT: In this paper, we first point out that one-to-one positive sample assignment is the key factor, while, one-to-many assignment in previous detectors causes redundant predictions in inference.

905, TITLE: DFAC Framework: Factorizing the Value Function via Quantile Mixture for Multi-Agent Distributional Q-Learning
http://proceedings.mlr.press/v139/sun21c.html
AUTHORS: Wei-Fang Sun, Cheng-Kuang Lee, Chun-Yi Lee
HIGHLIGHT: To address the above issues, we integrate distributional RL and value function factorization methods by proposing a Distributional Value Function Factorization (DFAC) framework to generalize expected value function factorization methods to their distributional variants.

906, TITLE: Scalable Variational Gaussian Processes via Harmonic Kernel Decomposition
http://proceedings.mlr.press/v139/sun21d.html
AUTHORS: Shengyang Sun, Jiaxin Shi, Andrew Gordon Gordon Wilson, Roger B Grosse
HIGHLIGHT: We introduce a new scalable variational Gaussian process approximation which provides a high fidelity approximation while retaining general applicability.

907, TITLE: Reasoning Over Virtual Knowledge Bases With Open Predicate Relations
http://proceedings.mlr.press/v139/sun21e.html
AUTHORS: Haitian Sun, Patrick Verga, Bhuwan Dhingra, Ruslan Salakhutdinov, William W Cohen
HIGHLIGHT: We present the Open Predicate Query Language (OPQL); a method for constructing a virtual KB (VKB) trained entirely from text.

908, TITLE: PAC-Learning for Strategic Classification
http://proceedings.mlr.press/v139/sundaram21a.html
AUTHORS: Ravi Sundaram, Anil Vullikanti, Haifeng Xu, Fan Yao
HIGHLIGHT: In this paper, we generalize both of these through a unified framework for strategic classification and introduce the notion of strategic VC-dimension (SVC) to capture the PAC-learnability in our general strategic setup.

909, TITLE: Reinforcement Learning for Cost-Aware Markov Decision Processes
http://proceedings.mlr.press/v139/suttle21a.html

AUTHORS: Wesley Suttle, Kaiqing Zhang, Zhuoran Yang, Ji Liu, David Kraemer
HIGHLIGHT: This paper addresses this deficiency by introducing two new, model-free RL algorithms for solving cost-aware Markov decision processes, where the goal is to maximize the ratio of long-run average reward to long-run average cost.

910, TITLE: Model-Targeted Poisoning Attacks with Provable Convergence
http://proceedings.mlr.press/v139/suya21a.html
AUTHORS: Fnu Suya, Saeed Mahloujifar, Anshuman Suri, David Evans, Yuan Tian
HIGHLIGHT: We consider poisoning attacks against convex machine learning models and propose an efficient poisoning attack designed to induce a model specified by the adversary.

911, TITLE: Generalization Error Bound for Hyperbolic Ordinal Embedding
http://proceedings.mlr.press/v139/suzuki21a.html
AUTHORS: Atsushi Suzuki, Atsushi Nitanda, Jing Wang, Linchuan Xu, Kenji Yamanishi, Marc Cavazza
HIGHLIGHT: In this paper, through our novel characterization of HOE with decomposed Lorentz Gramian matrices, we provide a generalization error bound of HOE for the first time, which is at most exponential with respect to the embedding space's radius.

912, TITLE: Of Moments and Matching: A Game-Theoretic Framework for Closing the Imitation Gap
http://proceedings.mlr.press/v139/swamy21a.html
AUTHORS: Gokul Swamy, Sanjiban Choudhury, J. Andrew Bagnell, Steven Wu
HIGHLIGHT: We provide a unifying view of a large family of previous imitation learning algorithms through the lens of moment matching.

913, TITLE: Parallel tempering on optimized paths
http://proceedings.mlr.press/v139/syed21a.html
AUTHORS: Saifuddin Syed, Vittorio Romaniello, Trevor Campbell, Alexandre Bouchard-Cote
HIGHLIGHT: To address this issue, we expand the framework of PT to general families of paths, formulate the choice of path as an optimization problem that admits tractable gradient estimates, and propose a flexible new family of spline interpolation paths for use in practice.

914, TITLE: Robust Representation Learning via Perceptual Similarity Metrics
http://proceedings.mlr.press/v139/taghanaki21a.html
AUTHORS: Saeid A Taghanaki, Kristy Choi, Amir Hosein Khasahmadi, Anirudh Goyal
HIGHLIGHT: In this work, we propose Contrastive Input Morphing (CIM), a representation learning framework that learns input-space transformations of the data to mitigate the effect of irrelevant input features on downstream performance.

915, TITLE: DriftSurf: Stable-State / Reactive-State Learning under Concept Drift
http://proceedings.mlr.press/v139/tahmasbi21a.html
AUTHORS: Ashraf Tahmasbi, Ellango Jothimurugesan, Srikanta Tirthapura, Phillip B Gibbons
HIGHLIGHT: We present an adaptive learning algorithm that extends previous drift-detection-based methods by incorporating drift detection into a broader stable-state/reactive-state process.

916, TITLE: Sinkhorn Label Allocation: Semi-Supervised Classification via Annealed Self-Training
http://proceedings.mlr.press/v139/tai21a.html
AUTHORS: Kai Sheng Tai, Peter D Bailis, Gregory Valiant
HIGHLIGHT: In this paper, we reinterpret this label assignment process as an optimal transportation problem between examples and classes, wherein the cost of assigning an example to a class is mediated by the current predictions of the classifier.

917, TITLE: Approximation Theory Based Methods for RKHS Bandits
http://proceedings.mlr.press/v139/takemori21a.html
AUTHORS: Sho Takemori, Masahiro Sato
HIGHLIGHT: Using an approximation method, we propose efficient algorithms for the stochastic RKHS bandit problem and the first general algorithm for the adversarial RKHS bandit problem.

918, TITLE: Supervised Tree-Wasserstein Distance
http://proceedings.mlr.press/v139/takezawa21a.html
AUTHORS: Yuki Takezawa, Ryoma Sato, Makoto Yamada
HIGHLIGHT: In this work, we propose the Supervised Tree-Wasserstein (STW) distance, a fast, supervised metric learning method based on the tree metric.

919, TITLE: EfficientNetV2: Smaller Models and Faster Training
http://proceedings.mlr.press/v139/tan21a.html
AUTHORS: Mingxing Tan, Quoc Le
HIGHLIGHT: This paper introduces EfficientNetV2, a new family of convolutional networks that have faster training speed and better parameter efficiency than previous models.

920, TITLE: SGA: A Robust Algorithm for Partial Recovery of Tree-Structured Graphical Models with Noisy Samples
http://proceedings.mlr.press/v139/tandon21a.html
AUTHORS: Anshoo Tandon, Aldric Han, Vincent Tan
HIGHLIGHT: We consider learning Ising tree models when the observations from the nodes are corrupted by independent but non-identically distributed noise with unknown statistics.

921, TITLE: 1-bit Adam: Communication Efficient Large-Scale Training with Adam?s Convergence Speed
http://proceedings.mlr.press/v139/tang21a.html
AUTHORS: Hanlin Tang, Shaoduo Gan, Ammar Ahmad Awan, Samyam Rajbhandari, Conglong Li, Xiangru Lian, Ji Liu, Ce Zhang, Yuxiong He
HIGHLIGHT: In this paper, we propose 1-bit Adam that reduces the communication volume by up to 5x, offers much better scalability, and provides the same convergence speed as uncompressed Adam.

922, TITLE: Taylor Expansion of Discount Factors
http://proceedings.mlr.press/v139/tang21b.html
AUTHORS: Yunhao Tang, Mark Rowland, Remi Munos, Michal Valko
HIGHLIGHT: In this work, we study the effect that this discrepancy of discount factors has during learning, and discover a family of objectives that interpolate value functions of two distinct discount factors.

923, TITLE: REPAINT: Knowledge Transfer in Deep Reinforcement Learning
http://proceedings.mlr.press/v139/tao21a.html
AUTHORS: Yunzhe Tao, Sahika Genc, Jonathan Chung, Tao Sun, Sunil Mallya
HIGHLIGHT: This work proposes REPresentation And INstance Transfer (REPAINT) algorithm for knowledge transfer in deep reinforcement learning.

924, TITLE: Understanding the Dynamics of Gradient Flow in Overparameterized Linear models
http://proceedings.mlr.press/v139/tarmoun21a.html
AUTHORS: Salma Tarmoun, Guilherme Franca, Benjamin D Haeffele, Rene Vidal
HIGHLIGHT: We provide a detailed analysis of the dynamics ofthe gradient flow in overparameterized two-layerlinear models.

925, TITLE: Sequential Domain Adaptation by Synthesizing Distributionally Robust Experts
http://proceedings.mlr.press/v139/taskesen21a.html
AUTHORS: Bahar Taskesen, Man-Chung Yue, Jose Blanchet, Daniel Kuhn, Viet Anh Nguyen
HIGHLIGHT: Given available data, we investigate novel strategies to synthesize a family of least squares estimator experts that are robust with regard to moment conditions.

926, TITLE: A Language for Counterfactual Generative Models
http://proceedings.mlr.press/v139/tavares21a.html
AUTHORS: Zenna Tavares, James Koppel, Xin Zhang, Ria Das, Armando Solar-Lezama
HIGHLIGHT: We present Omega, a probabilistic programming language with support for counterfactual inference.

927, TITLE: Synthesizer: Rethinking Self-Attention for Transformer Models
http://proceedings.mlr.press/v139/tay21a.html
AUTHORS: Yi Tay, Dara Bahri, Donald Metzler, Da-Cheng Juan, Zhe Zhao, Che Zheng
HIGHLIGHT: To this end, we propose \textsc{Synthesizer}, a model that learns synthetic attention weights without token-token interactions.

928, TITLE: OmniNet: Omnidirectional Representations from Transformers
http://proceedings.mlr.press/v139/tay21b.html
AUTHORS: Yi Tay, Mostafa Dehghani, Vamsi Aribandi, Jai Gupta, Philip M Pham, Zhen Qin, Dara Bahri, Da-Cheng Juan, Donald Metzler
HIGHLIGHT: This paper proposes Omnidirectional Representations from Transformers (OMNINET).

929, TITLE: T-SCI: A Two-Stage Conformal Inference Algorithm with Guaranteed Coverage for Cox-MLP
http://proceedings.mlr.press/v139/teng21a.html
AUTHORS: Jiaye Teng, Zeren Tan, Yang Yuan
HIGHLIGHT: To recover the guaranteed coverage without linear assumption, we propose two algorithms based on conformal inference.

930, TITLE: Moreau-Yosida f-divergences
http://proceedings.mlr.press/v139/terjek21a.html
AUTHORS: D?vid Terj?k
HIGHLIGHT: Inspired by this, we define the Moreau-Yosida approximation of f-divergences with respect to the Wasserstein-1 metric.

931, TITLE: Understanding Invariance via Feedforward Inversion of Discriminatively Trained Classifiers
http://proceedings.mlr.press/v139/teterwak21a.html
AUTHORS: Piotr Teterwak, Chiyuan Zhang, Dilip Krishnan, Michael C Mozer
HIGHLIGHT: We explore this phenomenon further using a novel synthesis of methods, yielding a feedforward inversion model that produces remarkably high fidelity reconstructions, qualitatively superior to those of past efforts.

932, TITLE: Resource Allocation in Multi-armed Bandit Exploration: Overcoming Sublinear Scaling with Adaptive Parallelism
http://proceedings.mlr.press/v139/thananjeyan21a.html
AUTHORS: Brijen Thananjeyan, Kirthevasan Kandasamy, Ion Stoica, Michael Jordan, Ken Goldberg, Joseph Gonzalez
HIGHLIGHT: We study exploration in stochastic multi-armed bandits when we have access to a divisible resource that can be allocated in varying amounts to arm pulls.

933, TITLE: Monte Carlo Variational Auto-Encoders
http://proceedings.mlr.press/v139/thin21a.html
AUTHORS: Achille Thin, Nikita Kotelevskii, Arnaud Doucet, Alain Durmus, Eric Moulines, Maxim Panov
HIGHLIGHT: In this paper, we address both issues and demonstrate the performance of the resulting Monte Carlo VAEs on a variety of applications.

934, TITLE: Efficient Generative Modelling of Protein Structure Fragments using a Deep Markov Model
http://proceedings.mlr.press/v139/thygesen21a.html
AUTHORS: Christian B Thygesen, Christian Skj?dt Steenmans, Ahmad Salim Al-Sibahi, Lys Sanz Moreta, Anders Bundg?rd S?rensen, Thomas Hamelryck
HIGHLIGHT: To address these issues, we developed BIFROST, a novel take on the fragment library problem based on a Deep Markov Model architecture combined with directional statistics for angular degrees of freedom, implemented in the deep probabilistic programming language Pyro.

935, TITLE: Understanding self-supervised learning dynamics without contrastive pairs
http://proceedings.mlr.press/v139/tian21a.html
AUTHORS: Yuandong Tian, Xinlei Chen, Surya Ganguli
HIGHLIGHT: In this paper, we answer this question via a simple theoretical study and propose a novel approach, \ourmethod{}, that \emph{directly} sets the linear predictor based on the statistics of its inputs, rather than trained with gradient update.

936, TITLE: Online Learning in Unknown Markov Games
http://proceedings.mlr.press/v139/tian21b.html
AUTHORS: Yi Tian, Yuanhao Wang, Tiancheng Yu, Suvrit Sra
HIGHLIGHT: We study online learning in unknown Markov games, a problem that arises in episodic multi-agent reinforcement learning where the actions of the opponents are unobservable.

937, TITLE: BORE: Bayesian Optimization by Density-Ratio Estimation
http://proceedings.mlr.press/v139/tiao21a.html
AUTHORS: Louis C Tiao, Aaron Klein, Matthias W Seeger, Edwin V. Bonilla, Cedric Archambeau, Fabio Ramos
HIGHLIGHT: In this paper, we cast the computation of EI as a binary classification problem, building on the link between class-probability estimation and density-ratio estimation, and the lesser-known link between density-ratios and EI.

938, TITLE: Nonparametric Decomposition of Sparse Tensors
http://proceedings.mlr.press/v139/tillinghast21a.html
AUTHORS: Conor Tillinghast, Shandian Zhe

HIGHLIGHT: To address this model misspecification and to exploit the sparse tensor structures, we propose Nonparametric dEcomposition of Sparse Tensors (\ours), which can capture both the sparse structure properties and complex relationships between the tensor nodes to enhance the embedding estimation.

939, TITLE: Probabilistic Programs with Stochastic Conditioning
http://proceedings.mlr.press/v139/tolpin21a.html
AUTHORS: David Tolpin, Yuan Zhou, Tom Rainforth, Hongseok Yang
HIGHLIGHT: We propose a generalization of deterministic conditioning to stochastic conditioning, that is, conditioning on the marginal distribution of a variable taking a particular form.

940, TITLE: Deep Continuous Networks
http://proceedings.mlr.press/v139/tomen21a.html
AUTHORS: Nergis Tomen, Silvia-Laura Pintea, Jan Van Gemert
HIGHLIGHT: Here we propose deep continuous networks (DCNs), which combine spatially continuous filters, with the continuous depth framework of neural ODEs.

941, TITLE: Diffusion Earth Mover's Distance and Distribution Embeddings
http://proceedings.mlr.press/v139/tong21a.html
AUTHORS: Alexander Y Tong, Guillaume Huguet, Amine Natik, Kincaid Macdonald, Manik Kuchroo, Ronald Coifman, Guy Wolf, Smita Krishnaswamy
HIGHLIGHT: We propose a new fast method of measuring distances between large numbers of related high dimensional datasets called the Diffusion Earth Mover’s Distance (EMD).

942, TITLE: Training data-efficient image transformers & distillation through attention
http://proceedings.mlr.press/v139/touvron21a.html
AUTHORS: Hugo Touvron, Matthieu Cord, Matthijs Douze, Francisco Massa, Alexandre Sablayrolles, Herve Jegou
HIGHLIGHT: In this work, we produce competitive convolution-free transformers trained on ImageNet only using a single computer in less than 3 days.

943, TITLE: Conservative Objective Models for Effective Offline Model-Based Optimization
http://proceedings.mlr.press/v139/trabucco21a.html
AUTHORS: Brandon Trabucco, Aviral Kumar, Xinyang Geng, Sergey Levine
HIGHLIGHT: In this paper, we aim to solve data-driven model-based optimization (MBO) problems, where the goal is to find a design input that maximizes an unknown objective function provided access to only a static dataset of inputs and their corresponding objective values.

944, TITLE: Sparse within Sparse Gaussian Processes using Neighbor Information
http://proceedings.mlr.press/v139/tran21a.html
AUTHORS: Gia-Lac Tran, Dimitrios Milios, Pietro Michiardi, Maurizio Filippone
HIGHLIGHT: In particular, we introduce a novel hierarchical prior, which imposes sparsity on the set of inducing variables.

945, TITLE: SMG: A Shuffling Gradient-Based Method with Momentum
http://proceedings.mlr.press/v139/tran21b.html
AUTHORS: Trang H Tran, Lam M Nguyen, Quoc Tran-Dinh
HIGHLIGHT: We combine two advanced ideas widely used in optimization for machine learning: \textit{shuffling} strategy and \textit{momentum} technique to develop a novel shuffling gradient-based method with momentum, coined \textbf{S}huffling \textbf{M}omentum \textbf{G}radient (SMG), for non-convex finite-sum optimization problems.

946, TITLE: Bayesian Optimistic Optimisation with Exponentially Decaying Regret
http://proceedings.mlr.press/v139/tran-the21a.html
AUTHORS: Hung Tran-The, Sunil Gupta, Santu Rana, Svetha Venkatesh
HIGHLIGHT: We propose the BOO algorithm, a first practical approach which can achieve an exponential regret bound with order $\mathcal O(N^{-\sqrt{N}})$ under the assumption that the objective function is sampled from a Gaussian process with a Matérn kernel with smoothness parameter $\nu > 4 +\frac{D}{2}$, where D is the number of dimensions.

947, TITLE: On Disentangled Representations Learned from Correlated Data
http://proceedings.mlr.press/v139/trauble21a.html
AUTHORS: Frederik Tr?uble, Elliot Creager, Niki Kilbertus, Francesco Locatello, Andrea Dittadi, Anirudh Goyal, Bernhard Sch?lkopf, Stefan Bauer
HIGHLIGHT: In this work, we bridge the gap to real-world scenarios by analyzing the behavior of the most prominent disentanglement approaches on correlated data in a large-scale empirical study (including 4260 models).

948, TITLE: A New Formalism, Method and Open Issues for Zero-Shot Coordination
http://proceedings.mlr.press/v139/treutlein21a.html
AUTHORS: Johannes Treutlein, Michael Dennis, Caspar Oesterheld, Jakob Foerster
HIGHLIGHT: However, until now, this "label-free" problem has only been informally defined. We formalize this setting as the label-free coordination (LFC) problem by defining the label-free coordination game.

949, TITLE: Learning a Universal Template for Few-shot Dataset Generalization
http://proceedings.mlr.press/v139/triantafillou21a.html
AUTHORS: Eleni Triantafillou, Hugo Larochelle, Richard Zemel, Vincent Dumoulin
HIGHLIGHT: To this end, we propose to utilize the diverse training set to construct a \emph{universal template}: a partial model that can define a wide array of dataset-specialized models, by plugging in appropriate components.

950, TITLE: Provable Meta-Learning of Linear Representations
http://proceedings.mlr.press/v139/tripuraneni21a.html
AUTHORS: Nilesh Tripuraneni, Chi Jin, Michael Jordan
HIGHLIGHT: In this paper, we focus on the problem of multi-task linear regression—in which multiple linear regression models share a common, low-dimensional linear representation.

951, TITLE: Cumulants of Hawkes Processes are Robust to Observation Noise
http://proceedings.mlr.press/v139/trouleau21a.html
AUTHORS: William Trouleau, Jalal Etesami, Matthias Grossglauser, Negar Kiyavash, Patrick Thiran
HIGHLIGHT: In this work, we address the problem of learning the causal structure of MHPs when the observed timestamps of events are subject to random and unknown shifts, also known as random translations.

952, TITLE: PixelTransformer: Sample Conditioned Signal Generation
http://proceedings.mlr.press/v139/tulsiani21a.html
AUTHORS: Shubham Tulsiani, Abhinav Gupta
HIGHLIGHT: We propose a generative model that can infer a distribution for the underlying spatial signal conditioned on sparse samples e.g. plausible images given a few observed pixels.

953, TITLE: A Framework for Private Matrix Analysis in Sliding Window Model
http://proceedings.mlr.press/v139/upadhyay21a.html
AUTHORS: Jalaj Upadhyay, Sarvagya Upadhyay
HIGHLIGHT: We perform a rigorous study of private matrix analysis when only the last W updates to matrices are considered useful for analysis.

954, TITLE: Fast Projection Onto Convex Smooth Constraints
http://proceedings.mlr.press/v139/usmanova21a.html
AUTHORS: Ilnura Usmanova, Maryam Kamgarpour, Andreas Krause, Kfir Levy
HIGHLIGHT: In this work, we focus on projection problems where the constraints are smooth and the number of constraints is significantly smaller than the dimension.

955, TITLE: SGLB: Stochastic Gradient Langevin Boosting
http://proceedings.mlr.press/v139/ustimenko21a.html
AUTHORS: Aleksei Ustimenko, Liudmila Prokhorenkova
HIGHLIGHT: This paper introduces Stochastic Gradient Langevin Boosting (SGLB) - a powerful and efficient machine learning framework that may deal with a wide range of loss functions and has provable generalization guarantees.

956, TITLE: LTL2Action: Generalizing LTL Instructions for Multi-Task RL
http://proceedings.mlr.press/v139/vaezipoor21a.html
AUTHORS: Pashootan Vaezipoor, Andrew C Li, Rodrigo A Toro Icarte, Sheila A. Mcilraith
HIGHLIGHT: To reduce the overhead of learning LTL semantics, we introduce an environment-agnostic LTL pretraining scheme which improves sample-efficiency in downstream environments.

957, TITLE: Active Deep Probabilistic Subsampling
http://proceedings.mlr.press/v139/van-gorp21a.html
AUTHORS: Hans Van Gorp, Iris Huijben, Bastiaan S Veeling, Nicola Pezzotti, Ruud J. G. Van Sloun
HIGHLIGHT: We generalize DPS to a sequential method that actively picks the next sample based on the information acquired so far; dubbed Active-DPS (A-DPS).

958, TITLE: CURI: A Benchmark for Productive Concept Learning Under Uncertainty
http://proceedings.mlr.press/v139/vedantam21a.html
AUTHORS: Ramakrishna Vedantam, Arthur Szlam, Maximillian Nickel, Ari Morcos, Brenden M Lake
HIGHLIGHT: We introduce a new benchmark, Compositional Reasoning Under Uncertainty (CURI) that instantiates a series of few-shot, meta-learning tasks in a productive concept space to evaluate different aspects of systematic generalization under uncertainty, including splits that test abstract understandings of disentangling, productive generalization, learning boolean operations, variable binding, etc.

959, TITLE: Towards Domain-Agnostic Contrastive Learning
http://proceedings.mlr.press/v139/verma21a.html
AUTHORS: Vikas Verma, Thang Luong, Kenji Kawaguchi, Hieu Pham, Quoc Le
HIGHLIGHT: To overcome such limitation, we propose a domain-agnostic approach to contrastive learning, named DACL, that is applicable to problems where domain-specific data augmentations are not readily available.

960, TITLE: Sparsifying Networks via Subdifferential Inclusion
http://proceedings.mlr.press/v139/verma21b.html
AUTHORS: Sagar Verma, Jean-Christophe Pesquet
HIGHLIGHT: In this article, we propose a new formulation of the problem of generating sparse weights for a pre-trained neural network.

961, TITLE: Unbiased Gradient Estimation in Unrolled Computation Graphs with Persistent Evolution Strategies
http://proceedings.mlr.press/v139/vicol21a.html
AUTHORS: Paul Vicol, Luke Metz, Jascha Sohl-Dickstein
HIGHLIGHT: We introduce a method called Persistent Evolution Strategies (PES), which divides the computation graph into a series of truncated unrolls, and performs an evolution strategies-based update step after each unroll.

962, TITLE: Online Graph Dictionary Learning
http://proceedings.mlr.press/v139/vincent-cuaz21a.html
AUTHORS: C?dric Vincent-Cuaz, Titouan Vayer, R?mi Flamary, Marco Corneli, Nicolas Courty
HIGHLIGHT: We fill this gap by proposing a new online Graph Dictionary Learning approach, which uses the Gromov Wasserstein divergence for the data fitting term.

963, TITLE: Neuro-algorithmic Policies Enable Fast Combinatorial Generalization
http://proceedings.mlr.press/v139/vlastelica21a.html
AUTHORS: Marin Vlastelica, Michal Rolinek, Georg Martius
HIGHLIGHT: We show that, for a certain subclass of the MDP framework, this can be alleviated by a neuro-algorithmic policy architecture that embeds a time-dependent shortest path solver in a deep neural network.

964, TITLE: Efficient Training of Robust Decision Trees Against Adversarial Examples
http://proceedings.mlr.press/v139/vos21a.html
AUTHORS: Dani?l Vos, Sicco Verwer
HIGHLIGHT: We present GROOT, an efficient algorithm for training robust decision trees and random forests that runs in a matter of seconds to minutes.

965, TITLE: Object Segmentation Without Labels with Large-Scale Generative Models
http://proceedings.mlr.press/v139/voynov21a.html
AUTHORS: Andrey Voynov, Stanislav Morozov, Artem Babenko
HIGHLIGHT: This work demonstrates that large-scale unsupervised models can also perform a more challenging object segmentation task, requiring neither pixel-level nor image-level labeling.

966, TITLE: Principal Component Hierarchy for Sparse Quadratic Programs
http://proceedings.mlr.press/v139/vreugdenhil21a.html
AUTHORS: Robbie Vreugdenhil, Viet Anh Nguyen, Armin Eftekhari, Peyman Mohajerin Esfahani
HIGHLIGHT: Exploiting this property, we propose two scalable optimization algorithms, coined as the "best response" and the "dual program", that can efficiently screen the potential indices of the nonzero elements of the original program.

967, TITLE: Whitening and Second Order Optimization Both Make Information in the Dataset Unusable During Training, and Can Reduce or Prevent Generalization
http://proceedings.mlr.press/v139/wadia21a.html

AUTHORS: Neha Wadia, Daniel Duckworth, Samuel S Schoenholz, Ethan Dyer, Jascha Sohl-Dickstein
HIGHLIGHT: We show that both data whitening and second order optimization can harm or entirely prevent generalization.

968, TITLE: Safe Reinforcement Learning Using Advantage-Based Intervention
http://proceedings.mlr.press/v139/wagener21a.html
AUTHORS: Nolan C Wagener, Byron Boots, Ching-An Cheng
HIGHLIGHT: We propose a new algorithm, SAILR, that uses an intervention mechanism based on advantage functions to keep the agent safe throughout training and optimizes the agent's policy using off-the-shelf RL algorithms designed for unconstrained MDPs.

969, TITLE: Task-Optimal Exploration in Linear Dynamical Systems
http://proceedings.mlr.press/v139/wagenmaker21a.html
AUTHORS: Andrew J Wagenmaker, Max Simchowitz, Kevin Jamieson
HIGHLIGHT: In this work, we study task-guided exploration and determine what precisely an agent must learn about their environment in order to complete a particular task.

970, TITLE: Learning and Planning in Average-Reward Markov Decision Processes
http://proceedings.mlr.press/v139/wan21a.html
AUTHORS: Yi Wan, Abhishek Naik, Richard S Sutton
HIGHLIGHT: We introduce learning and planning algorithms for average-reward MDPs, including 1) the first general proven-convergent off-policy model-free control algorithm without reference states, 2) the first proven-convergent off-policy model-free prediction algorithm, and 3) the first off-policy learning algorithm that converges to the actual value function rather than to the value function plus an offset.

971, TITLE: Think Global and Act Local: Bayesian Optimisation over High-Dimensional Categorical and Mixed Search Spaces
http://proceedings.mlr.press/v139/wan21b.html
AUTHORS: Xingchen Wan, Vu Nguyen, Huong Ha, Binxin Ru, Cong Lu, Michael A. Osborne
HIGHLIGHT: We propose a novel solution—we combine local optimisation with a tailored kernel design, effectively handling high-dimensional categorical and mixed search spaces, whilst retaining sample efficiency.

972, TITLE: Zero-Shot Knowledge Distillation from a Decision-Based Black-Box Model
http://proceedings.mlr.press/v139/wang21a.html
AUTHORS: Zi Wang
HIGHLIGHT: We propose to generate pseudo samples that are distinguished by the decision boundaries of the DB3 teacher to the largest extent and construct soft labels for these samples, which are used as the transfer set.

973, TITLE: Fairness of Exposure in Stochastic Bandits
http://proceedings.mlr.press/v139/wang21b.html
AUTHORS: Lequn Wang, Yiwei Bai, Wen Sun, Thorsten Joachims
HIGHLIGHT: To remedy this problem, we propose a new bandit objective that guarantees merit-based fairness of exposure to the items while optimizing utility to the users.

974, TITLE: A Proxy Variable View of Shared Confounding
http://proceedings.mlr.press/v139/wang21c.html
AUTHORS: Yixin Wang, David Blei
HIGHLIGHT: In this paper, we focus on the setting where there are many treatments with shared confounding, and we study under what conditions is causal identification possible.

975, TITLE: Fast Algorithms for Stackelberg Prediction Game with Least Squares Loss
http://proceedings.mlr.press/v139/wang21d.html
AUTHORS: Jiali Wang, He Chen, Rujun Jiang, Xudong Li, Zihao Li
HIGHLIGHT: In contrast, we propose a novel approach that reformulates a SPG-LS as a single SDP of a similar form and the same dimension as those solved in the bisection method.

976, TITLE: Accelerate CNNs from Three Dimensions: A Comprehensive Pruning Framework
http://proceedings.mlr.press/v139/wang21e.html
AUTHORS: Wenxiao Wang, Minghao Chen, Shuai Zhao, Long Chen, Jinming Hu, Haifeng Liu, Deng Cai, Xiaofei He, Wei Liu

HIGHLIGHT: In this framework, since collecting too much data for training the regression is very time-costly, we propose two approaches to lower the cost: 1) specializing the polynomial to ensure an accurate regression even with less training data; 2) employing iterative pruning and fine-tuning to collect the data faster.

977, TITLE: Explainable Automated Graph Representation Learning with Hyperparameter Importance
http://proceedings.mlr.press/v139/wang21f.html
AUTHORS: Xin Wang, Shuyi Fan, Kun Kuang, Wenwu Zhu
HIGHLIGHT: We propose an explainable AutoML approach for graph representation (e-AutoGR) which utilizes explainable graph features during performance estimation and learns decorrelated importance weights for different hyperparameters in affecting the model performance through a non-linear decorrelated weighting regression.

978, TITLE: Self-Tuning for Data-Efficient Deep Learning
http://proceedings.mlr.press/v139/wang21g.html
AUTHORS: Ximei Wang, Jinghan Gao, Mingsheng Long, Jianmin Wang
HIGHLIGHT: To escape from this dilemma, we present Self-Tuning to enable data-efficient deep learning by unifying the exploration of labeled and unlabeled data and the transfer of a pre-trained model, as well as a Pseudo Group Contrast (PGC) mechanism to mitigate the reliance on pseudo-labels and boost the tolerance to false labels.

979, TITLE: Label Distribution Learning Machine
http://proceedings.mlr.press/v139/wang21h.html
AUTHORS: Jing Wang, Xin Geng
HIGHLIGHT: Specifically, we extend the margin theory to LDL and propose a new LDL method called \textbf{L}abel \textbf{D}istribution \textbf{L}earning \textbf{M}achine (LDLM).

980, TITLE: AlphaNet: Improved Training of Supernets with Alpha-Divergence
http://proceedings.mlr.press/v139/wang21i.html
AUTHORS: Dilin Wang, Chengyue Gong, Meng Li, Qiang Liu, Vikas Chandra
HIGHLIGHT: In this work, we propose to improve the supernet training with a more generalized alpha-divergence.

981, TITLE: Global Convergence of Policy Gradient for Linear-Quadratic Mean-Field Control/Game in Continuous Time
http://proceedings.mlr.press/v139/wang21j.html
AUTHORS: Weichen Wang, Jiequn Han, Zhuoran Yang, Zhaoran Wang
HIGHLIGHT: In this paper, we study the policy gradient (PG) method for the linear-quadratic mean-field control and game, where we assume each agent has identical linear state transitions and quadratic cost functions.

982, TITLE: SG-PALM: a Fast Physically Interpretable Tensor Graphical Model
http://proceedings.mlr.press/v139/wang21k.html
AUTHORS: Yu Wang, Alfred Hero
HIGHLIGHT: We propose a new graphical model inference procedure, called SG-PALM, for learning conditional dependency structure of high-dimensional tensor-variate data.

983, TITLE: Deep Generative Learning via Schr?dinger Bridge
http://proceedings.mlr.press/v139/wang21l.html
AUTHORS: Gefei Wang, Yuling Jiao, Qian Xu, Yang Wang, Can Yang
HIGHLIGHT: We propose to learn a generative model via entropy interpolation with a Schr{ö}dinger Bridge.

984, TITLE: Robust Inference for High-Dimensional Linear Models via Residual Randomization
http://proceedings.mlr.press/v139/wang21m.html
AUTHORS: Y. Samuel Wang, Si Kai Lee, Panos Toulis, Mladen Kolar
HIGHLIGHT: We propose a residual randomization procedure designed for robust inference using Lasso estimates in the high-dimensional setting.

985, TITLE: A Modular Analysis of Provable Acceleration via Polyak's Momentum: Training a Wide ReLU Network and a Deep Linear Network
http://proceedings.mlr.press/v139/wang21n.html
AUTHORS: Jun-Kun Wang, Chi-Heng Lin, Jacob D Abernethy
HIGHLIGHT: This work establishes that momentum does indeed speed up neural net training.

986, TITLE: Optimal Non-Convex Exact Recovery in Stochastic Block Model via Projected Power Method
http://proceedings.mlr.press/v139/wang21o.html

AUTHORS: Peng Wang, Huikang Liu, Zirui Zhou, Anthony Man-Cho So
HIGHLIGHT: In this paper, we study the problem of exact community recovery in the symmetric stochastic block model, where a graph of n vertices is randomly generated by partitioning the vertices into $K \ge 2$ equal-sized communities and then connecting each pair of vertices with probability that depends on their community memberships.

987, TITLE: ConvexVST: A Convex Optimization Approach to Variance-stabilizing Transformation
http://proceedings.mlr.press/v139/wang21p.html
AUTHORS: Mengfan Wang, Boyu Lyu, Guoqiang Yu
HIGHLIGHT: In this paper, we converted the VST problem into a convex optimization problem, which can always be efficiently solved, identified the specific structure of the convex problem, which further improved the efficiency of the proposed algorithm, and showed that any finite discrete distributions and the discretized version of any continuous distributions from real data can be variance-stabilized in an easy and nonparametric way.

988, TITLE: The Implicit Bias for Adaptive Optimization Algorithms on Homogeneous Neural Networks
http://proceedings.mlr.press/v139/wang21q.html
AUTHORS: Bohan Wang, Qi Meng, Wei Chen, Tie-Yan Liu
HIGHLIGHT: In this paper, we study the implicit bias of adaptive optimization algorithms on homogeneous neural networks.

989, TITLE: Robust Learning for Data Poisoning Attacks
http://proceedings.mlr.press/v139/wang21r.html
AUTHORS: Yunjuan Wang, Poorya Mianjy, Raman Arora
HIGHLIGHT: We investigate the robustness of stochastic approximation approaches against data poisoning attacks.

990, TITLE: SketchEmbedNet: Learning Novel Concepts by Imitating Drawings
http://proceedings.mlr.press/v139/wang21s.html
AUTHORS: Alexander Wang, Mengye Ren, Richard Zemel
HIGHLIGHT: While earlier approaches focus on generation quality or retrieval, we explore properties of image representations learned by training a model to produce sketches of images.

991, TITLE: Directional Bias Amplification
http://proceedings.mlr.press/v139/wang21t.html
AUTHORS: Angelina Wang, Olga Russakovsky
HIGHLIGHT: In this work, we focus on one aspect of the problem, namely bias amplification: the tendency of models to amplify the biases present in the data they are trained on.

992, TITLE: An exact solver for the Weston-Watkins SVM subproblem
http://proceedings.mlr.press/v139/wang21u.html
AUTHORS: Yutong Wang, Clayton Scott
HIGHLIGHT: In this work, we propose an algorithm that solves the subproblem exactly using a novel reparametrization of the Weston-Watkins dual problem.

993, TITLE: SCC: an efficient deep reinforcement learning agent mastering the game of StarCraft II
http://proceedings.mlr.press/v139/wang21v.html
AUTHORS: Xiangjun Wang, Junxiao Song, Penghui Qi, Peng Peng, Zhenkun Tang, Wei Zhang, Weimin Li, Xiongjun Pi, Jujie He, Chao Gao, Haitao Long, Quan Yuan
HIGHLIGHT: In this paper, we'll share the key insights and optimizations on efficient imitation learning and reinforcement learning for StarCraft II full game.

994, TITLE: Quantum algorithms for reinforcement learning with a generative model
http://proceedings.mlr.press/v139/wang21w.html
AUTHORS: Daochen Wang, Aarthi Sundaram, Robin Kothari, Ashish Kapoor, Martin Roetteler
HIGHLIGHT: For such an MDP, we design quantum algorithms that approximate an optimal policy (π^*), the optimal value function (v^*), and the optimal Q-function (q^*), assuming the algorithms can access samples from the environment in quantum superposition.

995, TITLE: Matrix Completion with Model-free Weighting
http://proceedings.mlr.press/v139/wang21x.html
AUTHORS: Jiayi Wang, Raymond K. W. Wong, Xiaojun Mao, Kwun Chuen Gary Chan
HIGHLIGHT: In this paper, we propose a novel method for matrix completion under general non-uniform missing structures.

996, TITLE: UniSpeech: Unified Speech Representation Learning with Labeled and Unlabeled Data
http://proceedings.mlr.press/v139/wang21y.html
AUTHORS: Chengyi Wang, Yu Wu, Yao Qian, Kenichi Kumatani, Shujie Liu, Furu Wei, Michael Zeng, Xuedong Huang
HIGHLIGHT: In this paper, we propose a unified pre-training approach called UniSpeech to learn speech representations with both labeled and unlabeled data, in which supervised phonetic CTC learning and phonetically-aware contrastive self-supervised learning are conducted in a multi-task learning manner.

997, TITLE: Instabilities of Offline RL with Pre-Trained Neural Representation
http://proceedings.mlr.press/v139/wang21z.html
AUTHORS: Ruosong Wang, Yifan Wu, Ruslan Salakhutdinov, Sham Kakade
HIGHLIGHT: In particular, our methodology explores these ideas when using features from pre-trained neural networks, in the hope that these representations are powerful enough to permit sample efficient offline RL.

998, TITLE: Learning to Weight Imperfect Demonstrations
http://proceedings.mlr.press/v139/wang21aa.html
AUTHORS: Yunke Wang, Chang Xu, Bo Du, Honglak Lee
HIGHLIGHT: In contrast, this paper proposes a method of learning to weight imperfect demonstrations in GAIL without imposing extensive prior information.

999, TITLE: Evolving Attention with Residual Convolutions
http://proceedings.mlr.press/v139/wang21ab.html
AUTHORS: Yujing Wang, Yaming Yang, Jiangang Bai, Mingliang Zhang, Jing Bai, Jing Yu, Ce Zhang, Gao Huang, Yunhai Tong
HIGHLIGHT: In this paper, we propose a novel and generic mechanism based on evolving attention to improve the performance of transformers.

1000, TITLE: Guarantees for Tuning the Step Size using a Learning-to-Learn Approach
http://proceedings.mlr.press/v139/wang21ac.html
AUTHORS: Xiang Wang, Shuai Yuan, Chenwei Wu, Rong Ge
HIGHLIGHT: In this paper we give meta-optimization guarantees for the learning-to-learn approach on a simple problem of tuning the step size for quadratic loss.

1001, TITLE: Bridging Multi-Task Learning and Meta-Learning: Towards Efficient Training and Effective Adaptation
http://proceedings.mlr.press/v139/wang21ad.html
AUTHORS: Haoxiang Wang, Han Zhao, Bo Li
HIGHLIGHT: In this paper, we take one important step further to understand the close connection between these two learning paradigms, through both theoretical analysis and empirical investigation.

1002, TITLE: Towards Better Laplacian Representation in Reinforcement Learning with Generalized Graph Drawing
http://proceedings.mlr.press/v139/wang21ae.html
AUTHORS: Kaixin Wang, Kuangqi Zhou, Qixin Zhang, Jie Shao, Bryan Hooi, Jiashi Feng
HIGHLIGHT: To solve this problem, we reformulate the graph drawing objective into a generalized form and derive a new learning objective, which is proved to have eigenvectors as its unique global minimizer.

1003, TITLE: Robust Asymmetric Learning in POMDPs
http://proceedings.mlr.press/v139/warrington21a.html
AUTHORS: Andrew Warrington, Jonathan W Lavington, Adam Scibior, Mark Schmidt, Frank Wood
HIGHLIGHT: To address this issue, we derive an update which, when applied iteratively to an expert, maximizes the expected reward of the trainee’s policy.

1004, TITLE: A Unified Generative Adversarial Network Training via Self-Labeling and Self-Attention
http://proceedings.mlr.press/v139/watanabe21a.html
AUTHORS: Tomoki Watanabe, Paolo Favaro
HIGHLIGHT: We propose a novel GAN training scheme that can handle any level of labeling in a unified manner.

1005, TITLE: Decision-Making Under Selective Labels: Optimal Finite-Domain Policies and Beyond
http://proceedings.mlr.press/v139/wei21a.html
AUTHORS: Dennis Wei
HIGHLIGHT: This paper studies the learning of decision policies in the face of selective labels, in an online setting that balances learning costs against future utility.

1006, TITLE: Inferring serial correlation with dynamic backgrounds
http://proceedings.mlr.press/v139/wei21b.html
AUTHORS: Song Wei, Yao Xie, Dobromir Rahnev
HIGHLIGHT: We propose a Total Variation (TV) constrained least square estimator coupled with hypothesis tests to infer the serial correlation in the presence of unknown and unstructured dynamic background.

1007, TITLE: Meta-learning Hyperparameter Performance Prediction with Neural Processes
http://proceedings.mlr.press/v139/wei21c.html
AUTHORS: Ying Wei, Peilin Zhao, Junzhou Huang
HIGHLIGHT: We propose an end-to-end surrogate named as Transfer NeuralProcesses (TNP) that learns a comprehensive set of meta-knowledge, including the parameters of historical surrogates, historical trials, and initial configurations for other datasets.

1008, TITLE: A Structured Observation Distribution for Generative Biological Sequence Prediction and Forecasting
http://proceedings.mlr.press/v139/weinstein21a.html
AUTHORS: Eli N Weinstein, Debora Marks
HIGHLIGHT: To address these problems, we propose a principled drop-in alternative to MSA preprocessing in the form of a structured observation distribution (the "MuE" distribution).

1009, TITLE: Thinking Like Transformers
http://proceedings.mlr.press/v139/weiss21a.html
AUTHORS: Gail Weiss, Yoav Goldberg, Eran Yahav
HIGHLIGHT: In this paper we aim to change that, proposing a computational model for the transformer-encoder in the form of a programming language.

1010, TITLE: Leveraged Weighted Loss for Partial Label Learning
http://proceedings.mlr.press/v139/wen21a.html
AUTHORS: Hongwei Wen, Jingyi Cui, Hanyuan Hang, Jiabin Liu, Yisen Wang, Zhouchen Lin
HIGHLIGHT: In this paper, we propose a family of loss functions named \textit{Leveraged Weighted} (LW) loss, which for the first time introduces the leverage parameter β to consider the trade-off between losses on partial labels and non-partial ones.

1011, TITLE: Characterizing the Gap Between Actor-Critic and Policy Gradient
http://proceedings.mlr.press/v139/wen21b.html
AUTHORS: Junfeng Wen, Saurabh Kumar, Ramki Gummadi, Dale Schuurmans
HIGHLIGHT: In this paper, we explain the gap between AC and PG methods by identifying the exact adjustment to the AC objective/gradient that recovers the true policy gradient of the cumulative reward objective (PG).

1012, TITLE: Toward Understanding the Feature Learning Process of Self-supervised Contrastive Learning
http://proceedings.mlr.press/v139/wen21c.html
AUTHORS: Zixin Wen, Yuanzhi Li
HIGHLIGHT: We present an underlying principle called feature decoupling to explain the effects of augmentations, where we theoretically characterize how augmentations can reduce the correlations of dense features between positive samples while keeping the correlations of sparse features intact, thereby forcing the neural networks to learn from the self-supervision of sparse features.

1013, TITLE: Keyframe-Focused Visual Imitation Learning
http://proceedings.mlr.press/v139/wen21d.html
AUTHORS: Chuan Wen, Jierui Lin, Jianing Qian, Yang Gao, Dinesh Jayaraman
HIGHLIGHT: We propose a solution that outperforms these prior approaches by upweighting demonstration keyframes corresponding to expert action changepoints.

1014, TITLE: Learning de-identified representations of prosody from raw audio
http://proceedings.mlr.press/v139/weston21a.html
AUTHORS: Jack Weston, Raphael Lenain, Udeepa Meepegama, Emil Fristed
HIGHLIGHT: We propose a method for learning de-identified prosody representations from raw audio using a contrastive self-supervised signal.

1015, TITLE: Solving Inverse Problems with a Flow-based Noise Model
http://proceedings.mlr.press/v139/whang21a.html
AUTHORS: Jay Whang, Qi Lei, Alex Dimakis
HIGHLIGHT: We study image inverse problems with a normalizing flow prior.

1016, TITLE: Composing Normalizing Flows for Inverse Problems
http://proceedings.mlr.press/v139/whang21b.html
AUTHORS: Jay Whang, Erik Lindgren, Alex Dimakis
HIGHLIGHT: Motivated by this, we propose a framework for approximate inference that estimates the target conditional as a composition of two flow models.

1017, TITLE: Which transformer architecture fits my data? A vocabulary bottleneck in self-attention
http://proceedings.mlr.press/v139/wies21a.html
AUTHORS: Noam Wies, Yoav Levine, Daniel Jannai, Amnon Shashua
HIGHLIGHT: We theoretically predict the existence of an embedding rank bottleneck that limits the contribution of self-attention width to the Transformer expressivity.

1018, TITLE: Prediction-Centric Learning of Independent Cascade Dynamics from Partial Observations
http://proceedings.mlr.press/v139/wilinski21a.html
AUTHORS: Mateusz Wilinski, Andrey Lokhov
HIGHLIGHT: We introduce a computationally efficient algorithm, based on a scalable dynamic message-passing approach, which is able to learn parameters of the effective spreading model given only limited information on the activation times of nodes in the network.

1019, TITLE: Leveraging Language to Learn Program Abstractions and Search Heuristics
http://proceedings.mlr.press/v139/wong21a.html
AUTHORS: Catherine Wong, Kevin M Ellis, Joshua Tenenbaum, Jacob Andreas
HIGHLIGHT: We introduce LAPS (Language for Abstraction and Program Search), a technique for using natural language annotations to guide joint learning of libraries and neurally-guided search models for synthesis.

1020, TITLE: Leveraging Sparse Linear Layers for Debuggable Deep Networks
http://proceedings.mlr.press/v139/wong21b.html
AUTHORS: Eric Wong, Shibani Santurkar, Aleksander Madry
HIGHLIGHT: We show how fitting sparse linear models over learned deep feature representations can lead to more debuggable neural networks.

1021, TITLE: Learning Neural Network Subspaces
http://proceedings.mlr.press/v139/wortsman21a.html
AUTHORS: Mitchell Wortsman, Maxwell C Horton, Carlos Guestrin, Ali Farhadi, Mohammad Rastegari
HIGHLIGHT: With a similar computational cost as training one model, we learn lines, curves, and simplexes of high-accuracy neural networks.

1022, TITLE: Conjugate Energy-Based Models
http://proceedings.mlr.press/v139/wu21a.html
AUTHORS: Hao Wu, Babak Esmaeili, Michael Wick, Jean-Baptiste Tristan, Jan-Willem Van De Meent
HIGHLIGHT: In this paper, we propose conjugate energy-based models (CEBMs), a new class of energy-based models that define a joint density over data and latent variables.

1023, TITLE: Making Paper Reviewing Robust to Bid Manipulation Attacks
http://proceedings.mlr.press/v139/wu21b.html
AUTHORS: Ruihan Wu, Chuan Guo, Felix Wu, Rahul Kidambi, Laurens Van Der Maaten, Kilian Weinberger
HIGHLIGHT: In this paper, we study the efficacy of such bid manipulation attacks and find that, indeed, they can jeopardize the integrity of the review process.

1024, TITLE: LIME: Learning Inductive Bias for Primitives of Mathematical Reasoning
http://proceedings.mlr.press/v139/wu21c.html
AUTHORS: Yuhuai Wu, Markus N Rabe, Wenda Li, Jimmy Ba, Roger B Grosse, Christian Szegedy
HIGHLIGHT: We specifically design these tasks to be synthetic and devoid of mathematical knowledge to ensure that only the fundamental reasoning biases can be learned from these tasks. This defines a new pre-training methodology called "LIME" (Learning Inductive bias for Mathematical rEasoning).

1025, TITLE: ChaCha for Online AutoML
http://proceedings.mlr.press/v139/wu21d.html
AUTHORS: Qingyun Wu, Chi Wang, John Langford, Paul Mineiro, Marco Rossi

HIGHLIGHT: We propose the ChaCha (Champion-Challengers) algorithm for making an online choice of hyperparameters in online learning settings.

1026, TITLE: Temporally Correlated Task Scheduling for Sequence Learning
http://proceedings.mlr.press/v139/wu21e.html
AUTHORS: Xueqing Wu, Lewen Wang, Yingce Xia, Weiqing Liu, Lijun Wu, Shufang Xie, Tao Qin, Tie-Yan Liu
HIGHLIGHT: In this work, we introduce a learnable scheduler to sequence learning, which can adaptively select auxiliary tasks for training depending on the model status and the current training data.

1027, TITLE: Class2Simi: A Noise Reduction Perspective on Learning with Noisy Labels
http://proceedings.mlr.press/v139/wu21f.html
AUTHORS: Songhua Wu, Xiaobo Xia, Tongliang Liu, Bo Han, Mingming Gong, Nannan Wang, Haifeng Liu, Gang Niu
HIGHLIGHT: To give an affirmative answer, in this paper, we propose a framework called \emph{Class2Simi}: it transforms data points with noisy \emph{class labels} to data pairs with noisy \emph{similarity labels}, where a similarity label denotes whether a pair shares the class label or not.

1028, TITLE: On Reinforcement Learning with Adversarial Corruption and Its Application to Block MDP
http://proceedings.mlr.press/v139/wu21g.html
AUTHORS: Tianhao Wu, Yunchang Yang, Simon Du, Liwei Wang
HIGHLIGHT: When the total number of corrupted episodes is known, we propose an algorithm, Corruption Robust Monotonic Value Propagation (\textsf{CR-MVP}), which achieves a regret bound of $\tilde{O}\left(\left(\sqrt{SAK}+S^2A+CSA)\right)\polylog(H)\right)$, where S is the number of states, A is the number of actions, H is the planning horizon, K is the number of episodes, and C is the corruption level.

1029, TITLE: Generative Video Transformer: Can Objects be the Words?
http://proceedings.mlr.press/v139/wu21h.html
AUTHORS: Yi-Fu Wu, Jaesik Yoon, Sungjin Ahn
HIGHLIGHT: In this paper, we propose the ObjectCentric Video Transformer (OCVT) which utilizes an object-centric approach for decomposing scenes into tokens suitable for use in a generative video transformer.

1030, TITLE: Uncertainty Weighted Actor-Critic for Offline Reinforcement Learning
http://proceedings.mlr.press/v139/wu21i.html
AUTHORS: Yue Wu, Shuangfei Zhai, Nitish Srivastava, Joshua M Susskind, Jian Zhang, Ruslan Salakhutdinov, Hanlin Goh
HIGHLIGHT: We propose Uncertainty Weighted Actor-Critic (UWAC), an algorithm that detects OOD state-action pairs and down-weights their contribution in the training objectives accordingly.

1031, TITLE: Towards Open-World Recommendation: An Inductive Model-based Collaborative Filtering Approach
http://proceedings.mlr.press/v139/wu21j.html
AUTHORS: Qitian Wu, Hengrui Zhang, Xiaofeng Gao, Junchi Yan, Hongyuan Zha
HIGHLIGHT: In this paper, we propose an inductive collaborative filtering framework that contains two representation models.

1032, TITLE: Data-efficient Hindsight Off-policy Option Learning
http://proceedings.mlr.press/v139/wulfmeier21a.html
AUTHORS: Markus Wulfmeier, Dushyant Rao, Roland Hafner, Thomas Lampe, Abbas Abdolmaleki, Tim Hertweck, Michael Neunert, Dhruva Tirumala, Noah Siegel, Nicolas Heess, Martin Riedmiller
HIGHLIGHT: We introduce Hindsight Off-policy Options (HO2), a data-efficient option learning algorithm.

1033, TITLE: A Bit More Bayesian: Domain-Invariant Learning with Uncertainty
http://proceedings.mlr.press/v139/xiao21a.html
AUTHORS: Zehao Xiao, Jiayi Shen, Xiantong Zhen, Ling Shao, Cees Snoek
HIGHLIGHT: In this paper, we address both challenges with a probabilistic framework based on variational Bayesian inference, by incorporating uncertainty into neural network weights.

1034, TITLE: On the Optimality of Batch Policy Optimization Algorithms
http://proceedings.mlr.press/v139/xiao21b.html
AUTHORS: Chenjun Xiao, Yifan Wu, Jincheng Mei, Bo Dai, Tor Lattimore, Lihong Li, Csaba Szepesvari, Dale Schuurmans
HIGHLIGHT: Therefore, to establish a framework for distinguishing algorithms, we introduce a new weighted-minimax criterion that considers the inherent difficulty of optimal value prediction.

1035, TITLE: CRFL: Certifiably Robust Federated Learning against Backdoor Attacks
http://proceedings.mlr.press/v139/xie21a.html
AUTHORS: Chulin Xie, Minghao Chen, Pin-Yu Chen, Bo Li
HIGHLIGHT: This paper provides the first general framework, Certifiably Robust Federated Learning (CRFL), to train certifiably robust FL models against backdoors.

1036, TITLE: RNNRepair: Automatic RNN Repair via Model-based Analysis
http://proceedings.mlr.press/v139/xie21b.html
AUTHORS: Xiaofei Xie, Wenbo Guo, Lei Ma, Wei Le, Jian Wang, Lingjun Zhou, Yang Liu, Xinyu Xing
HIGHLIGHT: We propose a lightweight model-based approach (RNNRepair) to help understand and repair incorrect behaviors of an RNN.

1037, TITLE: Deep Reinforcement Learning amidst Continual Structured Non-Stationarity
http://proceedings.mlr.press/v139/xie21c.html
AUTHORS: Annie Xie, James Harrison, Chelsea Finn
HIGHLIGHT: In this work, we formalize this problem setting, and draw upon ideas from the online learning and probabilistic inference literature to derive an off-policy RL algorithm that can reason about and tackle such lifelong non-stationarity.

1038, TITLE: Batch Value-function Approximation with Only Realizability
http://proceedings.mlr.press/v139/xie21d.html
AUTHORS: Tengyang Xie, Nan Jiang
HIGHLIGHT: We make progress in a long-standing problem of batch reinforcement learning (RL): learning Q* from an exploratory and polynomial-sized dataset, using a realizable and otherwise arbitrary function class.

1039, TITLE: Interaction-Grounded Learning
http://proceedings.mlr.press/v139/xie21e.html
AUTHORS: Tengyang Xie, John Langford, Paul Mineiro, Ida Momennejad
HIGHLIGHT: We propose \emph{Interaction-Grounded Learning} for this novel setting, in which a learner’s goal is to interact with the environment with no grounding or explicit reward to optimize its policies.

1040, TITLE: Composed Fine-Tuning: Freezing Pre-Trained Denoising Autoencoders for Improved Generalization
http://proceedings.mlr.press/v139/xie21f.html
AUTHORS: Sang Michael Xie, Tengyu Ma, Percy Liang
HIGHLIGHT: We focus on prediction problems with structured outputs that are subject to output validity constraints, e.g. pseudocode-to-code translation where the code must compile.

1041, TITLE: Learning While Playing in Mean-Field Games: Convergence and Optimality
http://proceedings.mlr.press/v139/xie21g.html
AUTHORS: Qiaomin Xie, Zhuoran Yang, Zhaoran Wang, Andreea Minca
HIGHLIGHT: To bridge such a gap, we propose a fictitious play algorithm, which alternatively updates the policy (learning) and the mean-field state (playing) by one step of policy optimization and gradient descent, respectively.

1042, TITLE: Positive-Negative Momentum: Manipulating Stochastic Gradient Noise to Improve Generalization
http://proceedings.mlr.press/v139/xie21h.html
AUTHORS: Zeke Xie, Li Yuan, Zhanxing Zhu, Masashi Sugiyama
HIGHLIGHT: For simulating SGN at low computational costs and without changing the learning rate or batch size, we propose the Positive-Negative Momentum (PNM) approach that is a powerful alternative to conventional Momentum in classic optimizers.

1043, TITLE: A Hybrid Variance-Reduced Method for Decentralized Stochastic Non-Convex Optimization
http://proceedings.mlr.press/v139/xin21a.html
AUTHORS: Ran Xin, Usman Khan, Soummya Kar
HIGHLIGHT: In this context, we propose a novel single-loop decentralized hybrid variance-reduced stochastic gradient method, called GT-HSGD, that outperforms the existing approaches in terms of both the oracle complexity and practical implementation.

1044, TITLE: Explore Visual Concept Formation for Image Classification
http://proceedings.mlr.press/v139/xiong21a.html
AUTHORS: Shengzhou Xiong, Yihua Tan, Guoyou Wang

HIGHLIGHT: Inspired by this, we propose a learning strategy of visual concept formation (LSOVCF) based on the ConvNet, in which the two intertwined parts of concept formation, i.e. feature extraction and concept description, are learned together.

1045, TITLE: CRPO: A New Approach for Safe Reinforcement Learning with Convergence Guarantee
http://proceedings.mlr.press/v139/xu21a.html
AUTHORS: Tengyu Xu, Yingbin Liang, Guanghui Lan
HIGHLIGHT: In contrast, we propose a primal approach, called constraint-rectified policy optimization (CRPO), which updates the policy alternatingly between objective improvement and constraint satisfaction.

1046, TITLE: To be Robust or to be Fair: Towards Fairness in Adversarial Training
http://proceedings.mlr.press/v139/xu21b.html
AUTHORS: Han Xu, Xiaorui Liu, Yaxin Li, Anil Jain, Jiliang Tang
HIGHLIGHT: In this work, we empirically and theoretically show that this phenomenon can generally happen under adversarial training algorithms which minimize DNN models' robust errors.

1047, TITLE: Interpretable Stein Goodness-of-fit Tests on Riemannian Manifold
http://proceedings.mlr.press/v139/xu21c.html
AUTHORS: Wenkai Xu, Takeru Matsuda
HIGHLIGHT: In this study, we develop goodness-of-fit testing and interpretable model criticism methods for general distributions on Riemannian manifolds, including those with an intractable normalization constant.

1048, TITLE: Rethinking Neural vs. Matrix-Factorization Collaborative Filtering: the Theoretical Perspectives
http://proceedings.mlr.press/v139/xu21d.html
AUTHORS: Da Xu, Chuanwei Ruan, Evren Korpeoglu, Sushant Kumar, Kannan Achan
HIGHLIGHT: In this paper, we address the comparison rigorously by answering the following questions: 1. what is the limiting expressivity of each model; 2. under the practical gradient descent, to which solution does each optimization path converge; 3. how would the models generalize under the inductive and transductive learning setting.

1049, TITLE: Dash: Semi-Supervised Learning with Dynamic Thresholding
http://proceedings.mlr.press/v139/xu21e.html
AUTHORS: Yi Xu, Lei Shang, Jinxing Ye, Qi Qian, Yu-Feng Li, Baigui Sun, Hao Li, Rong Jin
HIGHLIGHT: In this work we develop a simple yet powerful framework, whose key idea is to select a subset of training examples from the unlabeled data when performing existing SSL methods so that only the unlabeled examples with pseudo labels related to the labeled data will be used to train models.

1050, TITLE: An End-to-End Framework for Molecular Conformation Generation via Bilevel Programming
http://proceedings.mlr.press/v139/xu21f.html
AUTHORS: Minkai Xu, Wujie Wang, Shitong Luo, Chence Shi, Yoshua Bengio, Rafael Gomez-Bombarelli, Jian Tang
HIGHLIGHT: In this paper, we propose an end-to-end solution for molecular conformation prediction called ConfVAE based on the conditional variational autoencoder framework.

1051, TITLE: Self-supervised Graph-level Representation Learning with Local and Global Structure
http://proceedings.mlr.press/v139/xu21g.html
AUTHORS: Minghao Xu, Hang Wang, Bingbing Ni, Hongyu Guo, Jian Tang
HIGHLIGHT: In this paper, we propose a unified framework called Local-instance and Global-semantic Learning (GraphLoG) for self-supervised whole-graph representation learning.

1052, TITLE: Conformal prediction interval for dynamic time-series
http://proceedings.mlr.press/v139/xu21h.html
AUTHORS: Chen Xu, Yao Xie
HIGHLIGHT: We develop a method to construct distribution-free prediction intervals for dynamic time-series, called \Verb|EnbPI| that wraps around any bootstrap ensemble estimator to construct sequential prediction intervals.

1053, TITLE: Learner-Private Convex Optimization
http://proceedings.mlr.press/v139/xu21i.html
AUTHORS: Jiaming Xu, Kuang Xu, Dana Yang
HIGHLIGHT: In this paper, we study how to optimally obfuscate the learner's queries in convex optimization with first-order feedback, so that their learned optimal value is provably difficult to estimate for the eavesdropping adversary.

1054, TITLE: Doubly Robust Off-Policy Actor-Critic: Convergence and Optimality

http://proceedings.mlr.press/v139/xu21j.html
AUTHORS: Tengyu Xu, Zhuoran Yang, Zhaoran Wang, Yingbin Liang
HIGHLIGHT: In this paper, we develop a doubly robust off-policy AC (DR-Off-PAC) for discounted MDP, which can take advantage of learned nuisance functions to reduce estimation errors.

1055, TITLE: Optimization of Graph Neural Networks: Implicit Acceleration by Skip Connections and More Depth
http://proceedings.mlr.press/v139/xu21k.html
AUTHORS: Keyulu Xu, Mozhi Zhang, Stefanie Jegelka, Kenji Kawaguchi
HIGHLIGHT: We take the first step towards analyzing GNN training by studying the gradient dynamics of GNNs.

1056, TITLE: Group-Sparse Matrix Factorization for Transfer Learning of Word Embeddings
http://proceedings.mlr.press/v139/xu21l.html
AUTHORS: Kan Xu, Xuanyi Zhao, Hamsa Bastani, Osbert Bastani
HIGHLIGHT: We propose a novel group-sparse penalty that exploits this sparsity to perform transfer learning when there is very little text data available in the target domain—e.g., a single article of text.

1057, TITLE: KNAS: Green Neural Architecture Search
http://proceedings.mlr.press/v139/xu21m.html
AUTHORS: Jingjing Xu, Liang Zhao, Junyang Lin, Rundong Gao, Xu Sun, Hongxia Yang
HIGHLIGHT: According to this hypothesis, we propose a new kernel based architecture search approach KNAS.

1058, TITLE: Structured Convolutional Kernel Networks for Airline Crew Scheduling
http://proceedings.mlr.press/v139/yaakoubi21a.html
AUTHORS: Yassine Yaakoubi, Francois Soumis, Simon Lacoste-Julien
HIGHLIGHT: Motivated by the needs from an airline crew scheduling application, we introduce structured convolutional kernel networks (Struct-CKN), which combine CKNs from Mairal et al. (2014) in a structured prediction framework that supports constraints on the outputs.

1059, TITLE: Mediated Uncoupled Learning: Learning Functions without Direct Input-output Correspondences
http://proceedings.mlr.press/v139/yamane21a.html
AUTHORS: Ikko Yamane, Junya Honda, Florian Yger, Masashi Sugiyama
HIGHLIGHT: In this paper, we consider the task of predicting Y from X when we have no paired data of them, but we have two separate, independent datasets of X and Y each observed with some mediating variable U, that is, we have two datasets $S_X = \{(X_i, U_i)\}$ and $S_Y = \{(U'_j, Y'_j)\}$.

1060, TITLE: EL-Attention: Memory Efficient Lossless Attention for Generation
http://proceedings.mlr.press/v139/yan21a.html
AUTHORS: Yu Yan, Jiusheng Chen, Weizhen Qi, Nikhil Bhendawade, Yeyun Gong, Nan Duan, Ruofei Zhang
HIGHLIGHT: We propose memory-efficient lossless attention (called EL-attention) to address this issue.

1061, TITLE: Link Prediction with Persistent Homology: An Interactive View
http://proceedings.mlr.press/v139/yan21b.html
AUTHORS: Zuoyu Yan, Tengfei Ma, Liangcai Gao, Zhi Tang, Chao Chen
HIGHLIGHT: In this paper, we propose a novel topological approach to characterize interactions between two nodes.

1062, TITLE: CATE: Computation-aware Neural Architecture Encoding with Transformers
http://proceedings.mlr.press/v139/yan21c.html
AUTHORS: Shen Yan, Kaiqiang Song, Fei Liu, Mi Zhang
HIGHLIGHT: In this work, we introduce a Computation-Aware Transformer-based Encoding method called CATE.

1063, TITLE: On Perceptual Lossy Compression: The Cost of Perceptual Reconstruction and An Optimal Training Framework
http://proceedings.mlr.press/v139/yan21d.html
AUTHORS: Zeyu Yan, Fei Wen, Rendong Ying, Chao Ma, Peilin Liu
HIGHLIGHT: This paper provides nontrivial results theoretically revealing that, 1) the cost of achieving perfect perception quality is exactly a doubling of the lowest achievable MSE distortion, 2) an optimal encoder for the "classic" rate-distortion problem is also optimal for the perceptual compression problem, 3) distortion loss is unnecessary for training a perceptual decoder.

1064, TITLE: CIFS: Improving Adversarial Robustness of CNNs via Channel-wise Importance-based Feature Selection
http://proceedings.mlr.press/v139/yan21e.html

AUTHORS: Hanshu Yan, Jingfeng Zhang, Gang Niu, Jiashi Feng, Vincent Tan, Masashi Sugiyama
HIGHLIGHT: To examine this hypothesis, we introduce a novel mechanism, \textit{i.e.}, \underline{C}hannel-wise \underline{I}mportance-based \underline{F}eature \underline{S}election (CIFS).

1065, TITLE: Exact Gap between Generalization Error and Uniform Convergence in Random Feature Models
http://proceedings.mlr.press/v139/yang21a.html
AUTHORS: Zitong Yang, Yu Bai, Song Mei
HIGHLIGHT: To better understand this gap, we study the uniform convergence in the nonlinear random feature model and perform a precise theoretical analysis on how uniform convergence depends on the sample size and the number of parameters.

1066, TITLE: Learning Optimal Auctions with Correlated Valuations from Samples
http://proceedings.mlr.press/v139/yang21b.html
AUTHORS: Chunxue Yang, Xiaohui Bei
HIGHLIGHT: In this work, we investigate the robustness of the optimal auction with correlated valuations via sample complexity analysis.

1067, TITLE: Tensor Programs IV: Feature Learning in Infinite-Width Neural Networks
http://proceedings.mlr.press/v139/yang21c.html
AUTHORS: Greg Yang, Edward J. Hu
HIGHLIGHT: We propose simple modifications to the standard parametrization to allow for feature learning in the limit.

1068, TITLE: LARNet: Lie Algebra Residual Network for Face Recognition
http://proceedings.mlr.press/v139/yang21d.html
AUTHORS: Xiaolong Yang, Xiaohong Jia, Dihong Gong, Dong-Ming Yan, Zhifeng Li, Wei Liu
HIGHLIGHT: In this paper, we propose a novel method with Lie algebra theory to explore how face rotation in the 3D space affects the deep feature generation process of convolutional neural networks (CNNs).

1069, TITLE: BASGD: Buffered Asynchronous SGD for Byzantine Learning
http://proceedings.mlr.press/v139/yang21e.html
AUTHORS: Yi-Rui Yang, Wu-Jun Li
HIGHLIGHT: In this paper, we propose a novel method, called buffered asynchronous stochastic gradient descent (BASGD), for ABL.

1070, TITLE: Tensor Programs IIb: Architectural Universality Of Neural Tangent Kernel Training Dynamics
http://proceedings.mlr.press/v139/yang21f.html
AUTHORS: Greg Yang, Etai Littwin
HIGHLIGHT: To achieve this result, we apply the Tensor Programs technique: Write the entire SGD dynamics inside a Tensor Program and analyze it via the Master Theorem.

1071, TITLE: Graph Neural Networks Inspired by Classical Iterative Algorithms
http://proceedings.mlr.press/v139/yang21g.html
AUTHORS: Yongyi Yang, Tang Liu, Yangkun Wang, Jinjing Zhou, Quan Gan, Zhewei Wei, Zheng Zhang, Zengfeng Huang, David Wipf
HIGHLIGHT: To at least partially address these issues within a simple transparent framework, we consider a new family of GNN layers designed to mimic and integrate the update rules of two classical iterative algorithms, namely, proximal gradient descent and iterative reweighted least squares (IRLS).

1072, TITLE: Representation Matters: Offline Pretraining for Sequential Decision Making
http://proceedings.mlr.press/v139/yang21h.html
AUTHORS: Mengjiao Yang, Ofir Nachum
HIGHLIGHT: In this paper, we consider a slightly different approach to incorporating offline data into sequential decision-making.

1073, TITLE: Accelerating Safe Reinforcement Learning with Constraint-mismatched Baseline Policies
http://proceedings.mlr.press/v139/yang21i.html
AUTHORS: Tsung-Yen Yang, Justinian Rosca, Karthik Narasimhan, Peter J Ramadge
HIGHLIGHT: In order to safely learn from baseline policies, we propose an iterative policy optimization algorithm that alternates between maximizing expected return on the task, minimizing distance to the baseline policy, and projecting the policy onto the constraint-satisfying set.

1074, TITLE: Voice2Series: Reprogramming Acoustic Models for Time Series Classification
http://proceedings.mlr.press/v139/yang21j.html
AUTHORS: Chao-Han Huck Yang, Yun-Yun Tsai, Pin-Yu Chen
HIGHLIGHT: Motivated by the advances in deep speech processing models and the fact that voice data are univariate temporal signals, in this paper we propose Voice2Serie (V2S), a novel end-to-end approach that reprograms acoustic models for time series classification, through input transformation learning and output label mapping.

1075, TITLE: When All We Need is a Piece of the Pie: A Generic Framework for Optimizing Two-way Partial AUC
http://proceedings.mlr.press/v139/yang21k.html
AUTHORS: Zhiyong Yang, Qianqian Xu, Shilong Bao, Yuan He, Xiaochun Cao, Qingming Huang
HIGHLIGHT: To address this issue, we propose a generic framework to construct surrogate optimization problems, which supports efficient end-to-end training with deep-learning.

1076, TITLE: Rethinking Rotated Object Detection with Gaussian Wasserstein Distance Loss
http://proceedings.mlr.press/v139/yang21l.html
AUTHORS: Xue Yang, Junchi Yan, Qi Ming, Wentao Wang, Xiaopeng Zhang, Qi Tian
HIGHLIGHT: In this paper, we propose a novel regression loss based on Gaussian Wasserstein distance as a fundamental approach to solve the problem.

1077, TITLE: Delving into Deep Imbalanced Regression
http://proceedings.mlr.press/v139/yang21m.html
AUTHORS: Yuzhe Yang, Kaiwen Zha, Yingcong Chen, Hao Wang, Dina Katabi
HIGHLIGHT: Motivated by the intrinsic difference between categorical and continuous label space, we propose distribution smoothing for both labels and features, which explicitly acknowledges the effects of nearby targets, and calibrates both label and learned feature distributions.

1078, TITLE: Backpropagated Neighborhood Aggregation for Accurate Training of Spiking Neural Networks
http://proceedings.mlr.press/v139/yang21n.html
AUTHORS: Yukun Yang, Wenrui Zhang, Peng Li
HIGHLIGHT: We propose a novel BP-like method, called neighborhood aggregation (NA), which computes accurate error gradients guiding weight updates that may lead to discontinuous modifications of firing activities.

1079, TITLE: SimAM: A Simple, Parameter-Free Attention Module for Convolutional Neural Networks
http://proceedings.mlr.press/v139/yang21o.html
AUTHORS: Lingxiao Yang, Ru-Yuan Zhang, Lida Li, Xiaohua Xie
HIGHLIGHT: In this paper, we propose a conceptually simple but very effective attention module for Convolutional Neural Networks (ConvNets).

1080, TITLE: HAWQ-V3: Dyadic Neural Network Quantization
http://proceedings.mlr.press/v139/yao21a.html
AUTHORS: Zhewei Yao, Zhen Dong, Zhangcheng Zheng, Amir Gholami, Jiali Yu, Eric Tan, Leyuan Wang, Qijing Huang, Yida Wang, Michael Mahoney, Kurt Keutzer
HIGHLIGHT: To address this, we present HAWQ-V3, a novel mixed-precision integer-only quantization framework.

1081, TITLE: Improving Generalization in Meta-learning via Task Augmentation
http://proceedings.mlr.press/v139/yao21b.html
AUTHORS: Huaxiu Yao, Long-Kai Huang, Linjun Zhang, Ying Wei, Li Tian, James Zou, Junzhou Huang, Zhenhui () Li
HIGHLIGHT: Concretely, we propose two task augmentation methods, including MetaMix and Channel Shuffle.

1082, TITLE: Deep Learning for Functional Data Analysis with Adaptive Basis Layers
http://proceedings.mlr.press/v139/yao21c.html
AUTHORS: Junwen Yao, Jonas Mueller, Jane-Ling Wang
HIGHLIGHT: We introduce neural networks that employ a new Basis Layer whose hidden units are each basis functions themselves implemented as a micro neural network.

1083, TITLE: Addressing Catastrophic Forgetting in Few-Shot Problems
http://proceedings.mlr.press/v139/yap21a.html
AUTHORS: Pauching Yap, Hippolyt Ritter, David Barber
HIGHLIGHT: We demonstrate that the popular gradient-based model-agnostic meta-learning algorithm (MAML) indeed suffers from catastrophic forgetting and introduce a Bayesian online meta-learning framework that tackles this problem.

1084, TITLE: Reinforcement Learning with Prototypical Representations
http://proceedings.mlr.press/v139/yarats21a.html
AUTHORS: Denis Yarats, Rob Fergus, Alessandro Lazaric, Lerrel Pinto
HIGHLIGHT: To address these challenges we propose Proto-RL, a self-supervised framework that ties representation learning with exploration through prototypical representations.

1085, TITLE: Elementary superexpressive activations
http://proceedings.mlr.press/v139/yarotsky21a.html
AUTHORS: Dmitry Yarotsky
HIGHLIGHT: We call a finite family of activation functions \emph{superexpressive} if any multivariate continuous function can be approximated by a neural network that uses these activations and has a fixed architecture only depending on the number of input variables (i.e., to achieve any accuracy we only need to adjust the weights, without increasing the number of neurons).

1086, TITLE: Break-It-Fix-It: Unsupervised Learning for Program Repair
http://proceedings.mlr.press/v139/yasunaga21a.html
AUTHORS: Michihiro Yasunaga, Percy Liang
HIGHLIGHT: To bridge this gap, we propose a new training approach, Break-It-Fix-It (BIFI), which has two key ideas: (i) we use the critic to check a fixer's output on real bad inputs and add good (fixed) outputs to the training data, and (ii) we train a breaker to generate realistic bad code from good code.
Existing works create training data consisting of (bad, good) pairs by corrupting good examples using heuristics (e.g., dropping tokens).

1087, TITLE: Improving Gradient Regularization using Complex-Valued Neural Networks
http://proceedings.mlr.press/v139/yeats21a.html
AUTHORS: Eric C Yeats, Yiran Chen, Hai Li
HIGHLIGHT: A form of complex-valued neural network (CVNN) is proposed to improve the performance of gradient regularization on classification tasks of real-valued input in adversarial settings.

1088, TITLE: Neighborhood Contrastive Learning Applied to Online Patient Monitoring
http://proceedings.mlr.press/v139/yeche21a.html
AUTHORS: Hugo Y?che, Gideon Dresdner, Francesco Locatello, Matthias H?ser, Gunnar R?tsch
HIGHLIGHT: In this work, we overcome this limitation by supplementing time-series data augmentation techniques with a novel contrastive learning objective which we call neighborhood contrastive learning (NCL).

1089, TITLE: From Local Structures to Size Generalization in Graph Neural Networks
http://proceedings.mlr.press/v139/yehudai21a.html
AUTHORS: Gilad Yehudai, Ethan Fetaya, Eli Meirom, Gal Chechik, Haggai Maron
HIGHLIGHT: In this paper, we identify an important type of data where generalization from small to large graphs is challenging: graph distributions for which the local structure depends on the graph size.

1090, TITLE: Improved OOD Generalization via Adversarial Training and Pretraing
http://proceedings.mlr.press/v139/yi21a.html
AUTHORS: Mingyang Yi, Lu Hou, Jiacheng Sun, Lifeng Shang, Xin Jiang, Qun Liu, Zhiming Ma
HIGHLIGHT: In this paper, after defining OOD generalization by Wasserstein distance, we theoretically justify that a model robust to input perturbation also generalizes well on OOD data.

1091, TITLE: Regret and Cumulative Constraint Violation Analysis for Online Convex Optimization with Long Term Constraints
http://proceedings.mlr.press/v139/yi21b.html
AUTHORS: Xinlei Yi, Xiuxian Li, Tao Yang, Lihua Xie, Tianyou Chai, Karl Johansson
HIGHLIGHT: This paper considers online convex optimization with long term constraints, where constraints can be violated in intermediate rounds, but need to be satisfied in the long run.

1092, TITLE: Continuous-time Model-based Reinforcement Learning
http://proceedings.mlr.press/v139/yildiz21a.html
AUTHORS: Cagatay Yildiz, Markus Heinonen, Harri L?hdesm?ki
HIGHLIGHT: To avoid time-discretization approximation of the underlying process, we propose a continuous-time MBRL framework based on a novel actor-critic method.

1093, TITLE: Distributed Nystr?m Kernel Learning with Communications

http://proceedings.mlr.press/v139/yin21a.html
AUTHORS: Rong Yin, Weiping Wang, Dan Meng
HIGHLIGHT: We study the statistical performance for distributed kernel ridge regression with Nyström (DKRR-NY) and with Nyström and iterative solvers (DKRR-NY-PCG) and successfully derive the optimal learning rates, which can improve the ranges of the number of local processors p to the optimal in existing state-of-art bounds.

1094, TITLE: Path Planning using Neural A* Search
http://proceedings.mlr.press/v139/yonetani21a.html
AUTHORS: Ryo Yonetani, Tatsunori Taniai, Mohammadamin Barekatain, Mai Nishimura, Asako Kanezaki
HIGHLIGHT: In this work, we reformulate a canonical A* search algorithm to be differentiable and couple it with a convolutional encoder to form an end-to-end trainable neural network planner.

1095, TITLE: SinIR: Efficient General Image Manipulation with Single Image Reconstruction
http://proceedings.mlr.press/v139/yoo21a.html
AUTHORS: Jihyeong Yoo, Qifeng Chen
HIGHLIGHT: We propose SinIR, an efficient reconstruction-based framework trained on a single natural image for general image manipulation, including super-resolution, editing, harmonization, paint-to-image, photo-realistic style transfer, and artistic style transfer.

1096, TITLE: Conditional Temporal Neural Processes with Covariance Loss
http://proceedings.mlr.press/v139/yoo21b.html
AUTHORS: Boseon Yoo, Jiwoo Lee, Janghoon Ju, Seijun Chung, Soyeon Kim, Jaesik Choi
HIGHLIGHT: We introduce a novel loss function, Covariance Loss, which is conceptually equivalent to conditional neural processes and has a form of regularization so that is applicable to many kinds of neural networks.

1097, TITLE: Adversarial Purification with Score-based Generative Models
http://proceedings.mlr.press/v139/yoon21a.html
AUTHORS: Jongmin Yoon, Sung Ju Hwang, Juho Lee
HIGHLIGHT: In this paper, we propose a novel adversarial purification method based on an EBM trained with DSM.

1098, TITLE: Federated Continual Learning with Weighted Inter-client Transfer
http://proceedings.mlr.press/v139/yoon21b.html
AUTHORS: Jaehong Yoon, Wonyong Jeong, Giwoong Lee, Eunho Yang, Sung Ju Hwang
HIGHLIGHT: To resolve these issues, we propose a novel federated continual learning framework, Federated Weighted Inter-client Transfer (FedWeIT), which decomposes the network weights into global federated parameters and sparse task-specific parameters, and each client receives selective knowledge from other clients by taking a weighted combination of their task-specific parameters.

1099, TITLE: Autoencoding Under Normalization Constraints
http://proceedings.mlr.press/v139/yoon21c.html
AUTHORS: Sangwoong Yoon, Yung-Kyun Noh, Frank Park
HIGHLIGHT: We propose the Normalized Autoencoder (NAE), a normalized probabilistic model constructed from an autoencoder.

1100, TITLE: Accelerated Algorithms for Smooth Convex-Concave Minimax Problems with O(1/k^2) Rate on Squared Gradient Norm
http://proceedings.mlr.press/v139/yoon21d.html
AUTHORS: Taeho Yoon, Ernest K Ryu
HIGHLIGHT: In this work, we study the computational complexity of reducing the squared gradient magnitude for smooth minimax optimization problems.

1101, TITLE: Lower-Bounded Proper Losses for Weakly Supervised Classification
http://proceedings.mlr.press/v139/yoshida21a.html
AUTHORS: Shuhei M Yoshida, Takashi Takenouchi, Masashi Sugiyama
HIGHLIGHT: This paper discusses the problem of weakly supervised classification, in which instances are given weak labels that are produced by some label-corruption process.

1102, TITLE: Graph Contrastive Learning Automated
http://proceedings.mlr.press/v139/you21a.html
AUTHORS: Yuning You, Tianlong Chen, Yang Shen, Zhangyang Wang

HIGHLIGHT: Aiming to fill in this crucial gap, this paper proposes a unified bi-level optimization framework to automatically, adaptively and dynamically select data augmentations when performing GraphCL on specific graph data.

1103, TITLE: LogME: Practical Assessment of Pre-trained Models for Transfer Learning
http://proceedings.mlr.press/v139/you21b.html
AUTHORS: Kaichao You, Yong Liu, Jianmin Wang, Mingsheng Long
HIGHLIGHT: In pursuit of a practical assessment method, we propose to estimate the maximum value of label evidence given features extracted by pre-trained models.

1104, TITLE: Exponentially Many Local Minima in Quantum Neural Networks
http://proceedings.mlr.press/v139/you21c.html
AUTHORS: Xuchen You, Xiaodi Wu
HIGHLIGHT: We conduct a quantitative investigation on the landscape of loss functions of QNNs and identify a class of simple yet extremely hard QNN instances for training.

1105, TITLE: DAGs with No Curl: An Efficient DAG Structure Learning Approach
http://proceedings.mlr.press/v139/yu21a.html
AUTHORS: Yue Yu, Tian Gao, Naiyu Yin, Qiang Ji
HIGHLIGHT: To further improve efficiency, we propose a novel learning framework to model and learn the weighted adjacency matrices in the DAG space directly.

1106, TITLE: Provably Efficient Algorithms for Multi-Objective Competitive RL
http://proceedings.mlr.press/v139/yu21b.html
AUTHORS: Tiancheng Yu, Yi Tian, Jingzhao Zhang, Suvrit Sra
HIGHLIGHT: Our results extend Blackwell's approachability theorem \citep{blackwell1956analog} to tabular RL, where strategic exploration becomes essential.

1107, TITLE: Whittle Networks: A Deep Likelihood Model for Time Series
http://proceedings.mlr.press/v139/yu21c.html
AUTHORS: Zhongjie Yu, Fabrizio G Ventola, Kristian Kersting
HIGHLIGHT: To this end, we propose the first probabilistic circuits (PCs) approach for modeling the joint distribution of multivariate time series, called Whittle sum-product networks (WSPNs).

1108, TITLE: Deep Latent Graph Matching
http://proceedings.mlr.press/v139/yu21d.html
AUTHORS: Tianshu Yu, Runzhong Wang, Junchi Yan, Baoxin Li
HIGHLIGHT: To address this, we propose to learn the (distribution of) latent topology, which can better support the downstream GM task.

1109, TITLE: Learning Generalized Intersection Over Union for Dense Pixelwise Prediction
http://proceedings.mlr.press/v139/yu21e.html
AUTHORS: Jiaqian Yu, Jingtao Xu, Yiwei Chen, Weiming Li, Qiang Wang, Byungin Yoo, Jae-Joon Han
HIGHLIGHT: In this paper, we propose PixIoU, a generalized IoU for pixelwise prediction that is sensitive to the distance for non-overlapping cases and the locations in prediction.

1110, TITLE: Large Scale Private Learning via Low-rank Reparametrization
http://proceedings.mlr.press/v139/yu21f.html
AUTHORS: Da Yu, Huishuai Zhang, Wei Chen, Jian Yin, Tie-Yan Liu
HIGHLIGHT: We propose a reparametrization scheme to address the challenges of applying differentially private SGD on large neural networks, which are 1) the huge memory cost of storing individual gradients, 2) the added noise suffering notorious dimensional dependence.

1111, TITLE: Federated Deep AUC Maximization for Hetergeneous Data with a Constant Communication Complexity
http://proceedings.mlr.press/v139/yuan21a.html
AUTHORS: Zhuoning Yuan, Zhishuai Guo, Yi Xu, Yiming Ying, Tianbao Yang
HIGHLIGHT: In this paper, we propose improved FDAM algorithms for heterogeneous data by solving the popular non-convex strongly-concave min-max formulation of DAM in a distributed fashion, which can also be applied to a class of non-convex strongly-concave min-max problems.

1112, TITLE: Neural Tangent Generalization Attacks

http://proceedings.mlr.press/v139/yuan21b.html
AUTHORS: Chia-Hung Yuan, Shan-Hung Wu
HIGHLIGHT: In this paper, we study the generalization attacks against DNNs, where an attacker aims to slightly modify training data in order to spoil the training process such that a trained network lacks generalizability.

1113, TITLE: On Explainability of Graph Neural Networks via Subgraph Explorations
http://proceedings.mlr.press/v139/yuan21c.html
AUTHORS: Hao Yuan, Haiyang Yu, Jie Wang, Kang Li, Shuiwang Ji
HIGHLIGHT: In this work, we propose a novel method, known as SubgraphX, to explain GNNs by identifying important subgraphs.

1114, TITLE: Federated Composite Optimization
http://proceedings.mlr.press/v139/yuan21d.html
AUTHORS: Honglin Yuan, Manzil Zaheer, Sashank Reddi
HIGHLIGHT: In this paper, we study the Federated Composite Optimization (FCO) problem, in which the loss function contains a non-smooth regularizer.

1115, TITLE: Three Operator Splitting with a Nonconvex Loss Function
http://proceedings.mlr.press/v139/yurtsever21a.html
AUTHORS: Alp Yurtsever, Varun Mangalick, Suvrit Sra
HIGHLIGHT: We consider the problem of minimizing the sum of three functions, one of which is nonconvex but differentiable, and the other two are convex but possibly nondifferentiable.

1116, TITLE: Grey-box Extraction of Natural Language Models
http://proceedings.mlr.press/v139/zanella-beguelin21a.html
AUTHORS: Santiago Zanella-Beguelin, Shruti Tople, Andrew Paverd, Boris K?pf
HIGHLIGHT: In this paper we present algebraic and hybrid algebraic/learning-based attacks on large-scale natural language models.

1117, TITLE: Exponential Lower Bounds for Batch Reinforcement Learning: Batch RL can be Exponentially Harder than Online RL
http://proceedings.mlr.press/v139/zanette21a.html
AUTHORS: Andrea Zanette
HIGHLIGHT: For both tasks we derive exponential information-theoretic lower bounds in discounted infinite horizon MDPs with a linear function representation for the action value function even if 1) realizability holds, 2) the batch algorithm observes the exact reward and transition functions, and 3) the batch algorithm is given the best a priori data distribution for the problem class.

1118, TITLE: Learning Binary Decision Trees by Argmin Differentiation
http://proceedings.mlr.press/v139/zantedeschi21a.html
AUTHORS: Valentina Zantedeschi, Matt Kusner, Vlad Niculae
HIGHLIGHT: We propose to learn discrete parameters (i.e., for tree traversals and node pruning) and continuous parameters (i.e., for tree split functions and prediction functions) simultaneously using argmin differentiation.

1119, TITLE: Barlow Twins: Self-Supervised Learning via Redundancy Reduction
http://proceedings.mlr.press/v139/zbontar21a.html
AUTHORS: Jure Zbontar, Li Jing, Ishan Misra, Yann Lecun, Stephane Deny
HIGHLIGHT: We propose an objective function that naturally avoids collapse by measuring the cross-correlation matrix between the outputs of two identical networks fed with distorted versions of a sample, and making it as close to the identity matrix as possible.

1120, TITLE: You Only Sample (Almost) Once: Linear Cost Self-Attention Via Bernoulli Sampling
http://proceedings.mlr.press/v139/zeng21a.html
AUTHORS: Zhanpeng Zeng, Yunyang Xiong, Sathya Ravi, Shailesh Acharya, Glenn M Fung, Vikas Singh
HIGHLIGHT: In this paper, we show that a Bernoulli sampling attention mechanism based on Locality Sensitive Hashing (LSH), decreases the quadratic complexity of such models to linear.

1121, TITLE: DouZero: Mastering DouDizhu with Self-Play Deep Reinforcement Learning
http://proceedings.mlr.press/v139/zha21a.html
AUTHORS: Daochen Zha, Jingru Xie, Wenye Ma, Sheng Zhang, Xiangru Lian, Xia Hu, Ji Liu
HIGHLIGHT: In this work, we propose a conceptually simple yet effective DouDizhu AI system, namely DouZero, which enhances traditional Monte-Carlo methods with deep neural networks, action encoding, and parallel actors.

1122, TITLE: DORO: Distributional and Outlier Robust Optimization
http://proceedings.mlr.press/v139/zhai21a.html
AUTHORS: Runtian Zhai, Chen Dan, Zico Kolter, Pradeep Ravikumar
HIGHLIGHT: To resolve this issue, we propose the framework of DORO, for Distributional and Outlier Robust Optimization.

1123, TITLE: Can Subnetwork Structure Be the Key to Out-of-Distribution Generalization?
http://proceedings.mlr.press/v139/zhang21a.html
AUTHORS: Dinghuai Zhang, Kartik Ahuja, Yilun Xu, Yisen Wang, Aaron Courville
HIGHLIGHT: In this paper, we use a functional modular probing method to analyze deep model structures under OOD setting.

1124, TITLE: Towards Certifying L-infinity Robustness using Neural Networks with L-inf-dist Neurons
http://proceedings.mlr.press/v139/zhang21b.html
AUTHORS: Bohang Zhang, Tianle Cai, Zhou Lu, Di He, Liwei Wang
HIGHLIGHT: In this paper, we seek for a new approach to develop a theoretically principled neural network that inherently resists ℓ_∞ perturbations.

1125, TITLE: Efficient Lottery Ticket Finding: Less Data is More
http://proceedings.mlr.press/v139/zhang21c.html
AUTHORS: Zhenyu Zhang, Xuxi Chen, Tianlong Chen, Zhangyang Wang
HIGHLIGHT: This paper explores a new perspective on finding lottery tickets more efficiently, by doing so only with a specially selected subset of data, called Pruning-Aware Critical set (PrAC set), rather than using the full training set.

1126, TITLE: Robust Policy Gradient against Strong Data Corruption
http://proceedings.mlr.press/v139/zhang21d.html
AUTHORS: Xuezhou Zhang, Yiding Chen, Xiaojin Zhu, Wen Sun
HIGHLIGHT: We study the problem of robust reinforcement learning under adversarial corruption on both rewards and transitions.

1127, TITLE: Near Optimal Reward-Free Reinforcement Learning
http://proceedings.mlr.press/v139/zhang21e.html
AUTHORS: Zihan Zhang, Simon Du, Xiangyang Ji
HIGHLIGHT: We study the reward-free reinforcement learning framework, which is particularly suitable for batch reinforcement learning and scenarios where one needs policies for multiple reward functions.

1128, TITLE: Bayesian Attention Belief Networks
http://proceedings.mlr.press/v139/zhang21f.html
AUTHORS: Shujian Zhang, Xinjie Fan, Bo Chen, Mingyuan Zhou
HIGHLIGHT: This paper introduces Bayesian attention belief networks, which construct a decoder network by modeling unnormalized attention weights with a hierarchy of gamma distributions, and an encoder network by stacking Weibull distributions with a deterministic-upward-stochastic-downward structure to approximate the posterior.

1129, TITLE: Understanding Failures in Out-of-Distribution Detection with Deep Generative Models
http://proceedings.mlr.press/v139/zhang21g.html
AUTHORS: Lily Zhang, Mark Goldstein, Rajesh Ranganath
HIGHLIGHT: Deep generative models (DGMs) seem a natural fit for detecting out-of-distribution (OOD) inputs, but such models have been shown to assign higher probabilities or densities to OOD images than images from the training distribution. In this work, we explain why this behavior should be attributed to model misestimation.

1130, TITLE: Poolingformer: Long Document Modeling with Pooling Attention
http://proceedings.mlr.press/v139/zhang21h.html
AUTHORS: Hang Zhang, Yeyun Gong, Yelong Shen, Weisheng Li, Jiancheng Lv, Nan Duan, Weizhu Chen
HIGHLIGHT: In this paper, we introduce a two-level attention schema, Poolingformer, for long document modeling.

1131, TITLE: Probabilistic Generating Circuits
http://proceedings.mlr.press/v139/zhang21i.html
AUTHORS: Honghua Zhang, Brendan Juba, Guy Van Den Broeck
HIGHLIGHT: In this paper, we explore their use as a tractable probabilistic model, and propose probabilistic generating circuits (PGCs) for their efficient representation.

1132, TITLE: PAPRIKA: Private Online False Discovery Rate Control
http://proceedings.mlr.press/v139/zhang21j.html
AUTHORS: Wanrong Zhang, Gautam Kamath, Rachel Cummings
HIGHLIGHT: In this work, we study False Discovery Rate (FDR) control in multiple hypothesis testing under the constraint of differential privacy for the sample.

1133, TITLE: Learning from Noisy Labels with No Change to the Training Process
http://proceedings.mlr.press/v139/zhang21k.html
AUTHORS: Mingyuan Zhang, Jane Lee, Shivani Agarwal
HIGHLIGHT: In this paper, we show that this is really unnecessary: one can simply perform class probability estimation (CPE) on the noisy examples, e.g. using a standard (multiclass) logistic regression algorithm, and then apply noise-correction only in the final prediction step.

1134, TITLE: Progressive-Scale Boundary Blackbox Attack via Projective Gradient Estimation
http://proceedings.mlr.press/v139/zhang21l.html
AUTHORS: Jiawei Zhang, Linyi Li, Huichen Li, Xiaolu Zhang, Shuang Yang, Bo Li
HIGHLIGHT: In this paper, we show that such efficiency highly depends on the scale at which the attack is applied, and attacking at the optimal scale significantly improves the efficiency.

1135, TITLE: FOP: Factorizing Optimal Joint Policy of Maximum-Entropy Multi-Agent Reinforcement Learning
http://proceedings.mlr.press/v139/zhang21m.html
AUTHORS: Tianhao Zhang, Yueheng Li, Chen Wang, Guangming Xie, Zongqing Lu
HIGHLIGHT: In this paper, we present a novel multi-agent actor-critic method, FOP, which can factorize the optimal joint policy induced by maximum-entropy multi-agent reinforcement learning (MARL) into individual policies.

1136, TITLE: Learning Noise Transition Matrix from Only Noisy Labels via Total Variation Regularization
http://proceedings.mlr.press/v139/zhang21n.html
AUTHORS: Yivan Zhang, Gang Niu, Masashi Sugiyama
HIGHLIGHT: In this work, we propose a theoretically grounded method that can estimate the noise transition matrix and learn a classifier simultaneously, without relying on the error-prone noisy class-posterior estimation.

1137, TITLE: Quantile Bandits for Best Arms Identification
http://proceedings.mlr.press/v139/zhang21o.html
AUTHORS: Mengyan Zhang, Cheng Soon Ong
HIGHLIGHT: Motivated by risk-averse decision-making problems, our goal is to identify a set of m arms with the highest τ-quantile values within a fixed budget.

1138, TITLE: Towards Better Robust Generalization with Shift Consistency Regularization
http://proceedings.mlr.press/v139/zhang21p.html
AUTHORS: Shufei Zhang, Zhuang Qian, Kaizhu Huang, Qiufeng Wang, Rui Zhang, Xinping Yi
HIGHLIGHT: Towards better robust generalization, we propose a new regularization method {–} shift consistency regularization (SCR) {–} to steer the same-class latent features of both natural and adversarial data into a common direction during adversarial training.

1139, TITLE: On-Policy Deep Reinforcement Learning for the Average-Reward Criterion
http://proceedings.mlr.press/v139/zhang21q.html
AUTHORS: Yiming Zhang, Keith W Ross
HIGHLIGHT: We develop theory and algorithms for average-reward on-policy Reinforcement Learning (RL).

1140, TITLE: Differentiable Dynamic Quantization with Mixed Precision and Adaptive Resolution
http://proceedings.mlr.press/v139/zhang21r.html
AUTHORS: Zhaoyang Zhang, Wenqi Shao, Jinwei Gu, Xiaogang Wang, Ping Luo
HIGHLIGHT: Unlike prior arts that carefully tune these values, we present a fully differentiable approach to learn all of them, named Differentiable Dynamic Quantization (DDQ), which has several benefits.

1141, TITLE: iDARTS: Differentiable Architecture Search with Stochastic Implicit Gradients
http://proceedings.mlr.press/v139/zhang21s.html
AUTHORS: Miao Zhang, Steven W. Su, Shirui Pan, Xiaojun Chang, Ehsan M Abbasnejad, Reza Haffari
HIGHLIGHT: In this paper, we tackle the hypergradient computation in DARTS based on the implicit function theorem, making it only depends on the obtained solution to the inner-loop optimization and agnostic to the optimization path.

1142, TITLE: Deep Coherent Exploration for Continuous Control
http://proceedings.mlr.press/v139/zhang21t.html
AUTHORS: Yijie Zhang, Herke Van Hoof
HIGHLIGHT: In this paper, we introduce deep coherent exploration, a general and scalable exploration framework for deep RL algorithms for continuous control, that generalizes step-based and trajectory-based exploration.

1143, TITLE: Average-Reward Off-Policy Policy Evaluation with Function Approximation
http://proceedings.mlr.press/v139/zhang21u.html
AUTHORS: Shangtong Zhang, Yi Wan, Richard S Sutton, Shimon Whiteson
HIGHLIGHT: To address the deadly triad, we propose two novel algorithms, reproducing the celebrated success of Gradient TD algorithms in the average-reward setting.

1144, TITLE: Matrix Sketching for Secure Collaborative Machine Learning
http://proceedings.mlr.press/v139/zhang21v.html
AUTHORS: Mengjiao Zhang, Shusen Wang
HIGHLIGHT: We propose a practical defense which we call Double-Blind Collaborative Learning (DBCL).

1145, TITLE: MetaCURE: Meta Reinforcement Learning with Empowerment-Driven Exploration
http://proceedings.mlr.press/v139/zhang21w.html
AUTHORS: Jin Zhang, Jianhao Wang, Hao Hu, Tong Chen, Yingfeng Chen, Changjie Fan, Chongjie Zhang
HIGHLIGHT: To address this challenge, we explicitly model an exploration policy learning problem for meta-RL, which is separated from exploitation policy learning, and introduce a novel empowerment-driven exploration objective, which aims to maximize information gain for task identification.

1146, TITLE: World Model as a Graph: Learning Latent Landmarks for Planning
http://proceedings.mlr.press/v139/zhang21x.html
AUTHORS: Lunjun Zhang, Ge Yang, Bradly C Stadie
HIGHLIGHT: In this work, we propose to learn graph-structured world models composed of sparse, multi-step transitions.

1147, TITLE: Breaking the Deadly Triad with a Target Network
http://proceedings.mlr.press/v139/zhang21y.html
AUTHORS: Shangtong Zhang, Hengshuai Yao, Shimon Whiteson
HIGHLIGHT: In this paper, we investigate the target network as a tool for breaking the deadly triad, providing theoretical support for the conventional wisdom that a target network stabilizes training.

1148, TITLE: Multiscale Invertible Generative Networks for High-Dimensional Bayesian Inference
http://proceedings.mlr.press/v139/zhang21z.html
AUTHORS: Shumao Zhang, Pengchuan Zhang, Thomas Y Hou
HIGHLIGHT: We propose a Multiscale Invertible Generative Network (MsIGN) and associated training algorithm that leverages multiscale structure to solve high-dimensional Bayesian inference.

1149, TITLE: Meta Learning for Support Recovery in High-dimensional Precision Matrix Estimation
http://proceedings.mlr.press/v139/zhang21aa.html
AUTHORS: Qian Zhang, Yilin Zheng, Jean Honorio
HIGHLIGHT: In this paper, we study meta learning for support (i.e., the set of non-zero entries) recovery in high-dimensional precision matrix estimation where we reduce the sufficient sample complexity in a novel task with the information learned from other auxiliary tasks.

1150, TITLE: Model-Free Reinforcement Learning: from Clipped Pseudo-Regret to Sample Complexity
http://proceedings.mlr.press/v139/zhang21ab.html
AUTHORS: Zihan Zhang, Yuan Zhou, Xiangyang Ji
HIGHLIGHT: In this paper we consider the problem of learning an ϵ-optimal policy for a discounted Markov Decision Process (MDP).

1151, TITLE: Learning to Rehearse in Long Sequence Memorization
http://proceedings.mlr.press/v139/zhang21ac.html
AUTHORS: Zhu Zhang, Chang Zhou, Jianxin Ma, Zhijie Lin, Jingren Zhou, Hongxia Yang, Zhou Zhao
HIGHLIGHT: In this paper, we propose the Rehearsal Memory (RM) to enhance long-sequence memorization by self-supervised rehearsal with a history sampler.

1152, TITLE: Dataset Condensation with Differentiable Siamese Augmentation
http://proceedings.mlr.press/v139/zhao21a.html
AUTHORS: Bo Zhao, Hakan Bilen
HIGHLIGHT: In this paper, we focus on condensing large training sets into significantly smaller synthetic sets which can be used to train deep neural networks from scratch with minimum drop in performance.

1153, TITLE: Joining datasets via data augmentation in the label space for neural networks
http://proceedings.mlr.press/v139/zhao21b.html
AUTHORS: Junbo Zhao, Mingfeng Ou, Linji Xue, Yunkai Cui, Sai Wu, Gang Chen
HIGHLIGHT: In this article, we are interested in systematic ways to join datasets that are made of similar purposes.

1154, TITLE: Calibrate Before Use: Improving Few-shot Performance of Language Models
http://proceedings.mlr.press/v139/zhao21c.html
AUTHORS: Zihao Zhao, Eric Wallace, Shi Feng, Dan Klein, Sameer Singh
HIGHLIGHT: GPT-3 can perform numerous tasks when provided a natural language prompt that contains a few training examples. We show that this type of few-shot learning can be unstable: the choice of prompt format, training examples, and even the order of the examples can cause accuracy to vary from near chance to near state-of-the-art.

1155, TITLE: Few-Shot Neural Architecture Search
http://proceedings.mlr.press/v139/zhao21d.html
AUTHORS: Yiyang Zhao, Linnan Wang, Yuandong Tian, Rodrigo Fonseca, Tian Guo
HIGHLIGHT: In this paper, we propose few-shot NAS that uses multiple supernetworks, called sub-supernet, each covering different regions of the search space to alleviate the undesired co-adaption.

1156, TITLE: Expressive 1-Lipschitz Neural Networks for Robust Multiple Graph Learning against Adversarial Attacks
http://proceedings.mlr.press/v139/zhao21e.html
AUTHORS: Xin Zhao, Zeru Zhang, Zijie Zhang, Lingfei Wu, Jiayin Jin, Yang Zhou, Ruoming Jin, Dejing Dou, Da Yan
HIGHLIGHT: This paper proposes an attack-agnostic graph-adaptive 1-Lipschitz neural network, ERNN, for improving the robustness of deep multiple graph learning while achieving remarkable expressive power.

1157, TITLE: Fused Acoustic and Text Encoding for Multimodal Bilingual Pretraining and Speech Translation
http://proceedings.mlr.press/v139/zheng21a.html
AUTHORS: Renjie Zheng, Junkun Chen, Mingbo Ma, Liang Huang
HIGHLIGHT: To address these problems, we propose a Fused Acoustic and Text Masked Language Model (FAT-MLM) which jointly learns a unified representation for both acoustic and text input from various types of corpora including parallel data for speech recognition and machine translation, and even pure speech and text data.

1158, TITLE: Two Heads are Better Than One: Hypergraph-Enhanced Graph Reasoning for Visual Event Ratiocination
http://proceedings.mlr.press/v139/zheng21b.html
AUTHORS: Wenbo Zheng, Lan Yan, Chao Gou, Fei-Yue Wang
HIGHLIGHT: To this end, we propose a novel multi-modal model, Hypergraph-Enhanced Graph Reasoning.

1159, TITLE: How Framelets Enhance Graph Neural Networks
http://proceedings.mlr.press/v139/zheng21c.html
AUTHORS: Xuebin Zheng, Bingxin Zhou, Junbin Gao, Yuguang Wang, Pietro Li?, Ming Li, Guido Montufar
HIGHLIGHT: This paper presents a new approach for assembling graph neural networks based on framelet transforms.

1160, TITLE: Probabilistic Sequential Shrinking: A Best Arm Identification Algorithm for Stochastic Bandits with Corruptions
http://proceedings.mlr.press/v139/zhong21a.html
AUTHORS: Zixin Zhong, Wang Chi Cheung, Vincent Tan
HIGHLIGHT: We consider a best arm identification (BAI) problem for stochastic bandits with adversarial corruptions in the fixed-budget setting of T steps.

1161, TITLE: Towards Distraction-Robust Active Visual Tracking
http://proceedings.mlr.press/v139/zhong21b.html
AUTHORS: Fangwei Zhong, Peng Sun, Wenhan Luo, Tingyun Yan, Yizhou Wang
HIGHLIGHT: To address this issue, we propose a mixed cooperative-competitive multi-agent game, where a target and multiple distractors form a collaborative team to play against a tracker and make it fail to follow.

1162, TITLE: Provably Efficient Reinforcement Learning for Discounted MDPs with Feature Mapping
http://proceedings.mlr.press/v139/zhou21a.html
AUTHORS: Dongruo Zhou, Jiafan He, Quanquan Gu
HIGHLIGHT: In this paper, we study reinforcement learning for discounted Markov Decision Processes (MDPs), where the transition kernel can be parameterized as a linear function of certain feature mapping.

1163, TITLE: Amortized Conditional Normalized Maximum Likelihood: Reliable Out of Distribution Uncertainty Estimation
http://proceedings.mlr.press/v139/zhou21b.html
AUTHORS: Aurick Zhou, Sergey Levine
HIGHLIGHT: In this paper, we propose the amortized conditional normalized maximum likelihood (ACNML) method as a scalable general-purpose approach for uncertainty estimation, calibration, and out-of-distribution robustness with deep networks.

1164, TITLE: Optimal Estimation of High Dimensional Smooth Additive Function Based on Noisy Observations
http://proceedings.mlr.press/v139/zhou21c.html
AUTHORS: Fan Zhou, Ping Li
HIGHLIGHT: We inherit the idea from a recent work which introduced an effective bias reduction technique through iterative bootstrap and derive a bias-reducing estimator.

1165, TITLE: Incentivized Bandit Learning with Self-Reinforcing User Preferences
http://proceedings.mlr.press/v139/zhou21d.html
AUTHORS: Tianchen Zhou, Jia Liu, Chaosheng Dong, Jingyuan Deng
HIGHLIGHT: In this paper, we investigate a new multi-armed bandit (MAB) online learning model that considers real-world phenomena in many recommender systems: (i) the learning agent cannot pull the arms by itself and thus has to offer rewards to users to incentivize arm-pulling indirectly; and (ii) if users with specific arm preferences are well rewarded, they induce a "self-reinforcing" effect in the sense that they will attract more users of similar arm preferences.

1166, TITLE: Towards Defending against Adversarial Examples via Attack-Invariant Features
http://proceedings.mlr.press/v139/zhou21e.html
AUTHORS: Dawei Zhou, Tongliang Liu, Bo Han, Nannan Wang, Chunlei Peng, Xinbo Gao
HIGHLIGHT: To solve this problem, in this paper, we propose to remove adversarial noise by learning generalizable invariant features across attacks which maintain semantic classification information.

1167, TITLE: Asymmetric Loss Functions for Learning with Noisy Labels
http://proceedings.mlr.press/v139/zhou21f.html
AUTHORS: Xiong Zhou, Xianming Liu, Junjun Jiang, Xin Gao, Xiangyang Ji
HIGHLIGHT: In this work, we propose a new class of loss functions, namely asymmetric loss functions, which are robust to learning from noisy labels for arbitrary noise type.

1168, TITLE: Examining and Combating Spurious Features under Distribution Shift
http://proceedings.mlr.press/v139/zhou21g.html
AUTHORS: Chunting Zhou, Xuezhe Ma, Paul Michel, Graham Neubig
HIGHLIGHT: In this paper, we define and analyze robust and spurious representations using the information-theoretic concept of minimal sufficient statistics.

1169, TITLE: Sparse and Imperceptible Adversarial Attack via a Homotopy Algorithm
http://proceedings.mlr.press/v139/zhu21a.html
AUTHORS: Mingkang Zhu, Tianlong Chen, Zhangyang Wang
HIGHLIGHT: In this paper, we address this challenge by proposing a homotopy algorithm, to jointly tackle the sparsity and the perturbation bound in one unified framework.

1170, TITLE: Data-Free Knowledge Distillation for Heterogeneous Federated Learning
http://proceedings.mlr.press/v139/zhu21b.html
AUTHORS: Zhuangdi Zhu, Junyuan Hong, Jiayu Zhou
HIGHLIGHT: Inspired by the prior art, we propose a data-free knowledge distillation approach to address heterogeneous FL, where the server learns a lightweight generator to ensemble user information in a data-free manner, which is then broadcasted to users, regulating local training using the learned knowledge as an inductive bias.

1171, TITLE: Spectral vertex sparsifiers and pair-wise spanners over distributed graphs
http://proceedings.mlr.press/v139/zhu21c.html

AUTHORS: Chunjiang Zhu, Qinqing Liu, Jinbo Bi
HIGHLIGHT: In this work, we design communication-efficient distributed algorithms for constructing spectral vertex sparsifiers, which closely preserve effective resistance distances on a subset of vertices of interest in the original graphs, under the well-established message passing communication model.

1172, TITLE: Few-shot Language Coordination by Modeling Theory of Mind
http://proceedings.mlr.press/v139/zhu21d.html
AUTHORS: Hao Zhu, Graham Neubig, Yonatan Bisk
HIGHLIGHT: Drawing inspiration from the study of theory-of-mind (ToM; Premack & Woodruff (1978)), we study the effect of the speaker explicitly modeling the listener's mental state.

1173, TITLE: Clusterability as an Alternative to Anchor Points When Learning with Noisy Labels
http://proceedings.mlr.press/v139/zhu21e.html
AUTHORS: Zhaowei Zhu, Yiwen Song, Yang Liu
HIGHLIGHT: Our main contribution is the discovery of an efficient estimation procedure based on a clusterability condition.

1174, TITLE: Commutative Lie Group VAE for Disentanglement Learning
http://proceedings.mlr.press/v139/zhu21f.html
AUTHORS: Xinqi Zhu, Chang Xu, Dacheng Tao
HIGHLIGHT: A simple model named Commutative Lie Group VAE is introduced to realize the group-based disentanglement learning.

1175, TITLE: Accumulated Decoupled Learning with Gradient Staleness Mitigation for Convolutional Neural Networks
http://proceedings.mlr.press/v139/zhuang21a.html
AUTHORS: Huiping Zhuang, Zhenyu Weng, Fulin Luo, Toh Kar-Ann, Haizhou Li, Zhiping Lin
HIGHLIGHT: In this paper, we propose an accumulated decoupled learning (ADL), which includes a module-wise gradient accumulation in order to mitigate the gradient staleness.

1176, TITLE: Demystifying Inductive Biases for (Beta-)VAE Based Architectures
http://proceedings.mlr.press/v139/zietlow21a.html
AUTHORS: Dominik Zietlow, Michal Rolinek, Georg Martius
HIGHLIGHT: In this work, we shed light on the inductive bias responsible for the success of VAE-based architectures.

1177, TITLE: Recovering AES Keys with a Deep Cold Boot Attack
http://proceedings.mlr.press/v139/zimerman21a.html
AUTHORS: Itamar Zimerman, Eliya Nachmani, Lior Wolf
HIGHLIGHT: In this work we combine a deep error correcting code technique together with a modified SAT solver scheme in order to apply the attack to AES keys.

1178, TITLE: Learning Fair Policies in Decentralized Cooperative Multi-Agent Reinforcement Learning
http://proceedings.mlr.press/v139/zimmer21a.html
AUTHORS: Matthieu Zimmer, Claire Glanois, Umer Siddique, Paul Weng
HIGHLIGHT: As a solution method, we propose a novel neural network architecture, which is composed of two sub-networks specifically designed for taking into account these two aspects of fairness.

1179, TITLE: Contrastive Learning Inverts the Data Generating Process
http://proceedings.mlr.press/v139/zimmermann21a.html
AUTHORS: Roland S. Zimmermann, Yash Sharma, Steffen Schneider, Matthias Bethge, Wieland Brendel
HIGHLIGHT: We here prove that feedforward models trained with objectives belonging to the commonly used InfoNCE family learn to implicitly invert the underlying generative model of the observed data.

1180, TITLE: Exploration in Approximate Hyper-State Space for Meta Reinforcement Learning
http://proceedings.mlr.press/v139/zintgraf21a.html
AUTHORS: Luisa M Zintgraf, Leo Feng, Cong Lu, Maximilian Igl, Kristian Hartikainen, Katja Hofmann, Shimon Whiteson
HIGHLIGHT: To address this, we propose HyperX, which uses novel reward bonuses for meta-training to explore in approximate hyper-state space (where hyper-states represent the environment state and the agent's task belief).

1181, TITLE: Provable Robustness of Adversarial Training for Learning Halfspaces with Noise
http://proceedings.mlr.press/v139/zou21a.html
AUTHORS: Difan Zou, Spencer Frei, Quanquan Gu

HIGHLIGHT: To the best of our knowledge, this is the first work showing that adversarial training provably yields robust classifiers in the presence of noise.

1182, TITLE: On the Convergence of Hamiltonian Monte Carlo with Stochastic Gradients
http://proceedings.mlr.press/v139/zou21b.html
AUTHORS: Difan Zou, Quanquan Gu
HIGHLIGHT: In this paper, we propose a general framework for proving the convergence rate of HMC with stochastic gradient estimators, for sampling from strongly log-concave and log-smooth target distributions.

1183, TITLE: A Functional Perspective on Learning Symmetric Functions with Neural Networks
http://proceedings.mlr.press/v139/zweig21a.html
AUTHORS: Aaron Zweig, Joan Bruna
HIGHLIGHT: In this work we treat symmetric functions (of any size) as functions over probability measures, and study the learning and representation of neural networks defined on measures.