

1, TITLE: Improved Regret Bounds for Thompson Sampling in Linear Quadratic Control Problems  
<http://proceedings.mlr.press/v80/abeille18a.html>  
AUTHORS: Marc Abeille, Alessandro Lazaric  
HIGHLIGHT: In this paper, we study an instance of TS in the challenging setting of the infinite-horizon linear quadratic (LQ) control, which models problems with continuous state-action variables, linear dynamics, and quadratic cost.

2, TITLE: State Abstractions for Lifelong Reinforcement Learning  
<http://proceedings.mlr.press/v80/abel18a.html>  
AUTHORS: David Abel, Dilip Arumugam, Lucas Lehnert, Michael Littman  
HIGHLIGHT: We introduce two new classes of abstractions: (1) transitive state abstractions, whose optimal form can be computed efficiently, and (2) PAC state abstractions, which are guaranteed to hold with respect to a distribution of tasks.

3, TITLE: Policy and Value Transfer in Lifelong Reinforcement Learning  
<http://proceedings.mlr.press/v80/abel18b.html>  
AUTHORS: David Abel, Yuu Jinnai, Sophie Yue Guo, George Konidaris, Michael Littman  
HIGHLIGHT: We consider the problem of how best to use prior experience to bootstrap lifelong learning, where an agent faces a series of task instances drawn from some task distribution.

4, TITLE: INSPECTRE: Privately Estimating the Unseen  
<http://proceedings.mlr.press/v80/acharya18a.html>  
AUTHORS: Jayadev Acharya, Gautam Kamath, Ziteng Sun, Huanyu Zhang  
HIGHLIGHT: We develop differentially private methods for estimating various distributional properties.

5, TITLE: Learning Representations and Generative Models for 3D Point Clouds  
<http://proceedings.mlr.press/v80/achlioptas18a.html>  
AUTHORS: Panos Achlioptas, Olga Diamanti, Ioannis Mitliagkas, Leonidas Guibas  
HIGHLIGHT: In this paper, we look at geometric data represented as point clouds.

6, TITLE: Discovering Interpretable Representations for Both Deep Generative and Discriminative Models  
<http://proceedings.mlr.press/v80/adel18a.html>  
AUTHORS: Tameem Adel, Zoubin Ghahramani, Adrian Weller  
HIGHLIGHT: We propose two interpretability frameworks.

7, TITLE: A Reductions Approach to Fair Classification  
<http://proceedings.mlr.press/v80/agarwal18a.html>  
AUTHORS: Alekh Agarwal, Alina Beygelzimer, Miroslav Dudik, John Langford, Hanna Wallach  
HIGHLIGHT: We present a systematic approach for achieving fairness in a binary classification setting.

8, TITLE: Accelerated Spectral Ranking  
<http://proceedings.mlr.press/v80/agarwal18b.html>  
AUTHORS: Arpit Agarwal, Prathamesh Patil, Shivani Agarwal  
HIGHLIGHT: In this paper, we design a provably faster spectral ranking algorithm, which we call accelerated spectral ranking (ASR), that is also consistent under the MNL/BTL models.

9, TITLE: MISSION: Ultra Large-Scale Feature Selection using Count-Sketches  
<http://proceedings.mlr.press/v80/aghazadeh18a.html>  
AUTHORS: Amirali Aghazadeh, Ryan Spring, Daniel Lejeune, Gautam Dasarathy, Anshumali Shrivastava, baraniuk  
HIGHLIGHT: In this paper, we present MISSION, a novel framework for ultra large-scale feature selection that performs stochastic gradient descent while maintaining an efficient representation of the features in memory using a Count-Sketch data structure.

- 10, TITLE: Minimal I-MAP MCMC for Scalable Structure Discovery in Causal DAG Models  
<http://proceedings.mlr.press/v80/agrawal18a.html>  
AUTHORS: Raj Agrawal, Caroline Uhler, Tamara Broderick  
HIGHLIGHT: We here propose an alternative posterior approximation based on the observation that, if we incorporate empirical conditional independence tests, we can focus on a high-probability DAG associated with each order of the vertices.
- 11, TITLE: Proportional Allocation: Simple, Distributed, and Diverse Matching with High Entropy  
<http://proceedings.mlr.press/v80/agrawal18b.html>  
AUTHORS: Shipra Agrawal, Morteza Zadimoghaddam, Vahab Mirrokni  
HIGHLIGHT: Inspired by many applications of bipartite matching in online advertising and machine learning, we study a simple and natural iterative proportional allocation algorithm: Maintain a priority score  $\text{\$priority\_a}$  for each node  $s \in \mathcal{A}$  on one side of the bipartition, initialized as  $\text{\$priority\_a}=1$ .
- 12, TITLE: Bucket Renormalization for Approximate Inference  
<http://proceedings.mlr.press/v80/ahn18a.html>  
AUTHORS: Sungsoo Ahn, Michael Chertkov, Adrian Weller, Jinwoo Shin  
HIGHLIGHT: In this paper, we instead consider computing the partition function via sequential summation over variables.
- 13, TITLE: oi-VAE: Output Interpretable VAEs for Nonlinear Group Factor Analysis  
<http://proceedings.mlr.press/v80/ainsworth18a.html>  
AUTHORS: Samuel K. Ainsworth, Nicholas J. Foti, Adrian K. C. Lee, Emily B. Fox  
HIGHLIGHT: We present an output interpretable VAE (oi-VAE) for grouped data that models complex, nonlinear latent-to-observed relationships.
- 14, TITLE: Limits of Estimating Heterogeneous Treatment Effects: Guidelines for Practical Algorithm Design  
<http://proceedings.mlr.press/v80/alaa18a.html>  
AUTHORS: Ahmed Alaa, Mihaela Schaar  
HIGHLIGHT: In this paper, we provide such a guideline by characterizing the fundamental limits of estimating heterogeneous treatment effects, and establishing conditions under which these limits can be achieved.
- 15, TITLE: AutoPrognosis: Automated Clinical Prognostic Modeling via Bayesian Optimization with Structured Kernel Learning  
<http://proceedings.mlr.press/v80/alaa18b.html>  
AUTHORS: Ahmed Alaa, Mihaela Schaar  
HIGHLIGHT: To enable off-the-shelf usage of machine learning (ML) in prognostic research, we developed AUTOPROGNOSIS: a system for automating the design of predictive modeling pipelines tailored for clinical prognosis.
- 16, TITLE: Information Theoretic Guarantees for Empirical Risk Minimization with Applications to Model Selection and Large-Scale Optimization  
<http://proceedings.mlr.press/v80/alabdulmohsin18a.html>  
AUTHORS: Ibrahim Alabdulmohsin  
HIGHLIGHT: In this paper, we derive bounds on the mutual information of the empirical risk minimization (ERM) procedure for both 0-1 and strongly-convex loss classes.
- 17, TITLE: Fixing a Broken ELBO  
<http://proceedings.mlr.press/v80/alemi18a.html>  
AUTHORS: Alexander Alemi, Ben Poole, Ian Fischer, Joshua Dillon, Rif A. Saurous, Kevin Murphy  
HIGHLIGHT: Recent work in unsupervised representation learning has focused on learning deep directed latentvariable models.

- 18, TITLE: Differentially Private Identity and Equivalence Testing of Discrete Distributions  
<http://proceedings.mlr.press/v80/aliakbarpour18a.html>  
AUTHORS: Maryam Aliakbarpour, Ilias Diakonikolas, Ronitt Rubinfeld  
HIGHLIGHT: Our goal is to develop efficient testers while guaranteeing differential privacy to the individuals of the population.
- 19, TITLE: Katyusha X: Simple Momentum Method for Stochastic Sum-of-Nonconvex Optimization  
<http://proceedings.mlr.press/v80/allen-zhu18a.html>  
AUTHORS: Zeyuan Allen-Zhu  
HIGHLIGHT: We show how to provably obtain an accelerated stochastic algorithm for minimizing sum-of-nonconvex functions, by adding one additional line to the well-known SVRG method.
- 20, TITLE: Make the Minority Great Again: First-Order Regret Bound for Contextual Bandits  
<http://proceedings.mlr.press/v80/allen-zhu18b.html>  
AUTHORS: Zeyuan Allen-Zhu, Sebastien Bubeck, Yuanzhi Li  
HIGHLIGHT: In the present paper, we resolve this open problem by presenting a new strategy based on augmenting the policy space.
- 21, TITLE: Augmented CycleGAN: Learning Many-to-Many Mappings from Unpaired Data  
<http://proceedings.mlr.press/v80/almahairi18a.html>  
AUTHORS: Amjad Almahairi, Sai Rajeshwar, Alessandro Sordani, Philip Bachman, Aaron Courville  
HIGHLIGHT: We propose a new model, called Augmented CycleGAN, which learns many-to-many mappings between domains.
- 22, TITLE: Meta-Learning by Adjusting Priors Based on Extended PAC-Bayes Theory  
<http://proceedings.mlr.press/v80/amit18a.html>  
AUTHORS: Ron Amit, Ron Meir  
HIGHLIGHT: We present a framework for meta-learning that is based on generalization error bounds, allowing us to extend various PAC-Bayes bounds to meta-learning.
- 23, TITLE: MAGAN: Aligning Biological Manifolds  
<http://proceedings.mlr.press/v80/amodio18a.html>  
AUTHORS: Matthew Amodio, Smita Krishnaswamy  
HIGHLIGHT: We present the new Manifold Aligning GAN (MAGAN) that aligns two manifolds such that related points in each measurement space are aligned.
- 24, TITLE: Subspace Embedding and Linear Regression with Orlicz Norm  
<http://proceedings.mlr.press/v80/andoni18a.html>  
AUTHORS: Alexandr Andoni, Chengyu Lin, Ying Sheng, Peilin Zhong, Ruiqi Zhong  
HIGHLIGHT: Specifically, we show how to efficiently find an  $m \times n$  embedding matrix  $S$  ( $m \ll n$ ), such that for every  $d$ -dimensional vector  $x$ , we have  $\Omega(1/(d \log n)) \|Ax\|_G \leq \|SAx\|_2 \leq O(d^2 \log n) \|Ax\|_G$ . By applying this subspace embedding technique, we show an approximation algorithm for the regression problem  $\min_x \|Ax - b\|_G$ , up to a  $O(d \log^2 n)$  factor.
- 25, TITLE: Efficient Gradient-Free Variational Inference using Policy Search  
<http://proceedings.mlr.press/v80/arenz18a.html>  
AUTHORS: Oleg Arenz, Gerhard Neumann, Mingjun Zhong  
HIGHLIGHT: We propose an efficient, gradient-free method for learning general GMM approximations of multimodal distributions based on recent insights from stochastic search methods.
- 26, TITLE: On the Optimization of Deep Networks: Implicit Acceleration by Overparameterization  
<http://proceedings.mlr.press/v80/arora18a.html>  
AUTHORS: Sanjeev Arora, Nadav Cohen, Elad Hazan  
HIGHLIGHT: This paper suggests that, sometimes, increasing depth can speed up optimization.

27, TITLE: Stronger Generalization Bounds for Deep Nets via a Compression Approach  
<http://proceedings.mlr.press/v80/arora18b.html>  
AUTHORS: Sanjeev Arora, Rong Ge, Behnam Neyshabur, Yi Zhang  
HIGHLIGHT: Recent works try to give an explanation using PAC-Bayes and Margin-based analyses, but do not as yet result in sample complexity bounds better than naive parameter counting.

28, TITLE: Lipschitz Continuity in Model-based Reinforcement Learning  
<http://proceedings.mlr.press/v80/asadi18a.html>  
AUTHORS: Kavosh Asadi, Dipendra Misra, Michael Littman  
HIGHLIGHT: We examine the impact of learning Lipschitz continuous models in the context of model-based reinforcement learning.

29, TITLE: Obfuscated Gradients Give a False Sense of Security: Circumventing Defenses to Adversarial Examples  
<http://proceedings.mlr.press/v80/athalye18a.html>  
AUTHORS: Anish Athalye, Nicholas Carlini, David Wagner  
HIGHLIGHT: We describe characteristic behaviors of defenses exhibiting the effect, and for each of the three types of obfuscated gradients we discover, we develop attack techniques to overcome it.

30, TITLE: Synthesizing Robust Adversarial Examples  
<http://proceedings.mlr.press/v80/athalye18b.html>  
AUTHORS: Anish Athalye, Logan Engstrom, Andrew Ilyas, Kevin Kwok  
HIGHLIGHT: We demonstrate the existence of robust 3D adversarial objects, and we present the first algorithm for synthesizing examples that are adversarial over a chosen distribution of transformations.

31, TITLE: Contextual Graph Markov Model: A Deep and Generative Approach to Graph Processing  
<http://proceedings.mlr.press/v80/bacciu18a.html>  
AUTHORS: Davide Bacciu, Federico Errica, Alessio Micheli  
HIGHLIGHT: We introduce the Contextual Graph Markov Model, an approach combining ideas from generative models and neural networks for the processing of graph data.

32, TITLE: Greed is Still Good: Maximizing Monotone Submodular+Supermodular (BP) Functions  
<http://proceedings.mlr.press/v80/bai18a.html>  
AUTHORS: Wenruo Bai, Jeff Bilmes  
HIGHLIGHT: Using the curvature  $\text{curv}_f$  of the submodular term, and introducing  $\text{curv}^g$  for the supermodular term (a natural dual curvature for supermodular functions), however, both of which are computable in linear time, we show that BP maximization can be efficiently approximated by both the greedy and the semi-gradient based algorithm.

33, TITLE: Comparing Dynamics: Deep Neural Networks versus Glassy Systems  
<http://proceedings.mlr.press/v80/baity-jesi18a.html>  
AUTHORS: Marco Baity-Jesi, Levent Sagun, Mario Geiger, Stefano Spigler, Gerard Ben Arous, Chiara Cammarota, Yann LeCun, Matthieu Wyart, Giulio Biroli  
HIGHLIGHT: The two main issues we address are the complexity of the loss-landscape and of the dynamics within it, and to what extent DNNs share similarities with glassy systems.

34, TITLE: SMAC: Simultaneous Mapping and Clustering Using Spectral Decompositions  
<http://proceedings.mlr.press/v80/bajaj18a.html>  
AUTHORS: Chandrajit Bajaj, Tingran Gao, Zihang He, Qixing Huang, Zhenxiao Liang  
HIGHLIGHT: We introduce a principled approach for simultaneous mapping and clustering (SMAC) for establishing consistent maps across heterogeneous object collections (e.g., 2D images or 3D shapes).

- 35, TITLE: A Boo(n) for Evaluating Architecture Performance  
<http://proceedings.mlr.press/v80/bajgar18a.html>  
AUTHORS: Ondrej Bajgar, Rudolf Kadlec, Jan Kleindienst  
HIGHLIGHT: We point out important problems with the common practice of using the best single model performance for comparing deep learning architectures, and we propose a method that corrects these flaws.
- 36, TITLE: Learning to Branch  
<http://proceedings.mlr.press/v80/balcan18a.html>  
AUTHORS: Maria-Florina Balcan, Travis Dick, Tuomas Sandholm, Ellen Vitercik  
HIGHLIGHT: Many partitioning techniques have been proposed, but no theory describes which is optimal.
- 37, TITLE: The Mechanics of n-Player Differentiable Games  
<http://proceedings.mlr.press/v80/balduzzi18a.html>  
AUTHORS: David Balduzzi, Sebastien Racaniere, James Martens, Jakob Foerster, Karl Tuyls, Thore Graepel  
HIGHLIGHT: In this paper, we develop new techniques to understand and control the dynamics in general games.
- 38, TITLE: Spline Filters For End-to-End Deep Learning  
<http://proceedings.mlr.press/v80/balestriero18a.html>  
AUTHORS: Randall Balestriero, Romain Cosentino, Herve Glotin, Richard Baraniuk  
HIGHLIGHT: We propose to tackle the problem of end-to-end learning for raw waveform signals by introducing learnable continuous time-frequency atoms.
- 39, TITLE: A Spline Theory of Deep Learning  
<http://proceedings.mlr.press/v80/balestriero18b.html>  
AUTHORS: Randall Balestriero, baraniuk  
HIGHLIGHT: Going further, we propose a simple penalty term that can be added to the cost function of any DN learning algorithm to force the templates to be orthogonal with each other; this leads to significantly improved classification performance and reduced overfitting with no change to the DN architecture.
- 40, TITLE: Approximation Guarantees for Adaptive Sampling  
<http://proceedings.mlr.press/v80/balkanski18a.html>  
AUTHORS: Eric Balkanski, Yaron Singer  
HIGHLIGHT: In this paper we analyze an adaptive sampling approach for submodular maximization.
- 41, TITLE: Improving the Gaussian Mechanism for Differential Privacy: Analytical Calibration and Optimal Denoising  
<http://proceedings.mlr.press/v80/balle18a.html>  
AUTHORS: Borja Balle, Yu-Xiang Wang  
HIGHLIGHT: In this paper we revisit the Gaussian mechanism and show that the original analysis has several important limitations.
- 42, TITLE: Dissecting Adam: The Sign, Magnitude and Variance of Stochastic Gradients  
<http://proceedings.mlr.press/v80/balles18a.html>  
AUTHORS: Lukas Balles, Philipp Hennig  
HIGHLIGHT: We interpret ADAM as a combination of two aspects: for each weight, the update direction is determined by the sign of stochastic gradients, whereas the update magnitude is determined by an estimate of their relative variance.
- 43, TITLE: Differentially Private Database Release via Kernel Mean Embeddings  
<http://proceedings.mlr.press/v80/balog18a.html>  
AUTHORS: Matej Balog, Ilya Tolstikhin, Bernhard Schölkopf  
HIGHLIGHT: We describe two instantiations of the proposed framework, suitable under different scenarios, and prove theoretical results guaranteeing differential privacy of the resulting algorithms and the consistency of estimators constructed from their outputs.

- 44, TITLE: Improving Optimization for Models With Continuous Symmetry Breaking  
<http://proceedings.mlr.press/v80/bamler18a.html>  
AUTHORS: Robert Bamler, Stephan Mandt  
HIGHLIGHT: We propose a new optimization algorithm that speeds up convergence using ideas from gauge theory in physics.
- 45, TITLE: Improved Training of Generative Adversarial Networks Using Representative Features  
<http://proceedings.mlr.press/v80/bang18a.html>  
AUTHORS: Duhyeon Bang, Hyunjung Shim  
HIGHLIGHT: This paper achieves both aims simultaneously by improving the stability of training GANs.
- 46, TITLE: Using Inherent Structures to design Lean 2-layer RBMs  
<http://proceedings.mlr.press/v80/bansal18a.html>  
AUTHORS: Abhishek Bansal, Abhinav Anand, Chiranjib Bhattacharyya  
HIGHLIGHT: Motivated from the approach of Inherent Structure formalism (Stillinger & Weber, 1982), extensively used in analysing Spin Glasses, we propose a novel measure called Inherent Structure Capacity (ISC), which characterizes the representation capacity of a fixed architecture RBM by the expected number of modes of distributions emanating from the RBM with parameters drawn from a prior distribution.
- 47, TITLE: Classification from Pairwise Similarity and Unlabeled Data  
<http://proceedings.mlr.press/v80/bao18a.html>  
AUTHORS: Han Bao, Gang Niu, Masashi Sugiyama  
HIGHLIGHT: To overcome this problem, we propose a new weakly-supervised learning setting where only similar (S) data pairs (two examples belong to the same class) and unlabeled (U) data points are needed instead of fully labeled data, which is called SU classification.
- 48, TITLE: Bayesian Optimization of Combinatorial Structures  
<http://proceedings.mlr.press/v80/baptista18a.html>  
AUTHORS: Ricardo Baptista, Matthias Poloczek  
HIGHLIGHT: This article proposes, to the best of our knowledge, the first algorithm to overcome these challenges, based on an adaptive, scalable model that identifies useful combinatorial structure even when data is scarce.
- 49, TITLE: Geodesic Convolutional Shape Optimization  
<http://proceedings.mlr.press/v80/baque18a.html>  
AUTHORS: Pierre Baque, Edoardo Remelli, Francois Fleuret, Pascal Fua  
HIGHLIGHT: In this work, we introduce a new way to optimize complex shapes fast and accurately.
- 50, TITLE: Learning to Coordinate with Coordination Graphs in Repeated Single-Stage Multi-Agent Decision Problems  
<http://proceedings.mlr.press/v80/bargiacchi18a.html>  
AUTHORS: Eugenio Bargiacchi, Timothy Verstraeten, Diederik Roijers, Ann Now?, Hado Hasselt  
HIGHLIGHT: In this paper we study learning in repeated fully cooperative games, multi-agent multi-armed bandits (MAMABs), in which the expected rewards can be expressed as a coordination graph.
- 51, TITLE: Testing Sparsity over Known and Unknown Bases  
<http://proceedings.mlr.press/v80/barman18a.html>  
AUTHORS: Siddharth Barman, Arnab Bhattacharyya, Suprovat Ghoshal  
HIGHLIGHT: In this work, we describe property testing algorithms for sparsity that observe a low-dimensional projection of the input.
- 52, TITLE: Transfer in Deep Reinforcement Learning Using Successor Features and Generalised Policy Improvement

<http://proceedings.mlr.press/v80/barreto18a.html>

AUTHORS: Andre Barreto, Diana Borsa, John Quan, Tom Schaul, David Silver, Matteo Hessel, Daniel Mankowitz, Augustin Zidek, Remi Munos  
HIGHLIGHT: In this paper we extend the SF&GPI framework in two ways.

53, TITLE: Measuring abstract reasoning in neural networks

<http://proceedings.mlr.press/v80/barrett18a.html>

AUTHORS: David Barrett, Felix Hill, Adam Santoro, Ari Morcos, Timothy Lillicrap  
HIGHLIGHT: Here, we propose a dataset and challenge designed to probe abstract reasoning, inspired by a well-known human IQ test.

54, TITLE: Gradient descent with identity initialization efficiently learns positive definite linear transformations by deep residual networks

<http://proceedings.mlr.press/v80/bartlett18a.html>

AUTHORS: Peter Bartlett, Dave Helmbold, Philip Long  
HIGHLIGHT: We analyze algorithms for approximating a function  $f(x) = \Phi^T x$  mapping  $\mathbb{R}^d$  to  $\mathbb{R}^d$  using deep linear neural networks, i.e. that learn a function  $h$  parameterized by matrices  $\{\Theta_1, \dots, \Theta_L\}$  and defined by  $h(x) = \Theta_L \Theta_{L-1} \dots \Theta_1 x$ .

55, TITLE: Mutual Information Neural Estimation

<http://proceedings.mlr.press/v80/belghazi18a.html>

AUTHORS: Mohamed Ishmael Belghazi, Aristide Baratin, Sai Rajeshwar, Sherjil Ozair, Yoshua Bengio, Aaron Courville, Devon Hjelm  
HIGHLIGHT: We present a Mutual Information Neural Estimator (MINE) that is linearly scalable in dimensionality as well as in sample size, trainable through back-prop, and strongly consistent.

56, TITLE: To Understand Deep Learning We Need to Understand Kernel Learning

<http://proceedings.mlr.press/v80/belkin18a.html>

AUTHORS: Mikhail Belkin, Siyuan Ma, Soumik Mandal  
HIGHLIGHT: Generalization performance of classifiers in deep learning has recently become a subject of intense study.

57, TITLE: Understanding and Simplifying One-Shot Architecture Search

<http://proceedings.mlr.press/v80/bender18a.html>

AUTHORS: Gabriel Bender, Pieter-Jan Kindermans, Barret Zoph, Vijay Vasudevan, Quoc Le  
HIGHLIGHT: We aim to understand weight sharing for one-shot architecture search.

58, TITLE: signSGD: Compressed Optimisation for Non-Convex Problems

<http://proceedings.mlr.press/v80/bernstein18a.html>

AUTHORS: Jeremy Bernstein, Yu-Xiang Wang, Kamyar Azizzadenesheli, Animashree Anandkumar  
HIGHLIGHT: We extend our theory to the distributed setting, where the parameter server uses majority vote to aggregate gradient signs from each worker enabling 1-bit compression of worker-server communication in both directions.

59, TITLE: Distributed Clustering via LSH Based Data Partitioning

<http://proceedings.mlr.press/v80/bhaskara18a.html>

AUTHORS: Aditya Bhaskara, Maheshakya Wijewardena  
HIGHLIGHT: We propose a novel data partitioning idea to overcome this bottleneck, and in effect, have different machines focus on “finding different clusters”.

60, TITLE: Autoregressive Convolutional Neural Networks for Asynchronous Time Series

<http://proceedings.mlr.press/v80/binkowski18a.html>

AUTHORS: Mikolaj Binkowski, Gautier Marti, Philippe Donnat

**HIGHLIGHT:** We propose Significance-Offset Convolutional Neural Network, a deep convolutional network architecture for regression of multivariate asynchronous time series.

61, **TITLE:** Adaptive Sampled Softmax with Kernel Based Sampling  
<http://proceedings.mlr.press/v80/blanc18a.html>  
**AUTHORS:** Guy Blanc, Steffen Rendle  
**HIGHLIGHT:** In this work, we propose a new class of kernel based sampling methods and develop an efficient sampling algorithm.

62, **TITLE:** Optimizing the Latent Space of Generative Networks  
<http://proceedings.mlr.press/v80/bojanowski18a.html>  
**AUTHORS:** Piotr Bojanowski, Armand Joulin, David Lopez-Pas, Arthur Szlam  
**HIGHLIGHT:** The goal of this paper is to disentangle the contribution of these two factors to the success of GANs.

63, **TITLE:** NetGAN: Generating Graphs via Random Walks  
<http://proceedings.mlr.press/v80/bojchevski18a.html>  
**AUTHORS:** Aleksandar Bojchevski, Oleksandr Shchur, Daniel Z?gner, Stephan G?nnemann  
**HIGHLIGHT:** We propose NetGAN - the first implicit generative model for graphs able to mimic real-world networks.  
We pose the problem of graph generation as learning the distribution of biased random walks over the input graph.

64, **TITLE:** A Progressive Batching L-BFGS Method for Machine Learning  
<http://proceedings.mlr.press/v80/bollapragada18a.html>  
**AUTHORS:** Raghu Bollapragada, Jorge Nocedal, Dheevatsa Mudigere, Hao-Jun Shi, Ping Tak Peter Tang  
**HIGHLIGHT:** In this paper, we present a new version of the L-BFGS algorithm that combines three basic components - progressive batching, a stochastic line search, and stable quasi-Newton updating - and that performs well on training logistic regression and deep neural networks.

65, **TITLE:** Prediction Rule Reshaping  
<http://proceedings.mlr.press/v80/bonakdarpour18a.html>  
**AUTHORS:** Matt Bonakdarpour, Sabyasachi Chatterjee, Rina Foygel Barber, John Lafferty  
**HIGHLIGHT:** Two methods are proposed for high-dimensional shape-constrained regression and classification.

66, **TITLE:** QuantTree: Histograms for Change Detection in Multivariate Data Streams  
<http://proceedings.mlr.press/v80/boracchi18a.html>  
**AUTHORS:** Giacomo Boracchi, Diego Carrera, Cristiano Cervellera, Danilo Macci?  
**HIGHLIGHT:** We present QuantTree, a recursive binary splitting scheme that adaptively defines the histogram bins to ease the detection of any distribution change.

67, **TITLE:** Matrix Norms in Data Streams: Faster, Multi-Pass and Row-Order  
<http://proceedings.mlr.press/v80/braverman18a.html>  
**AUTHORS:** Vladimir Braverman, Stephen Chestnut, Robert Krauthgamer, Yi Li, David Woodruff, Lin Yang  
**HIGHLIGHT:** A central problem in mining massive data streams is characterizing which functions of an underlying frequency vector can be approximated efficiently.

68, **TITLE:** Predict and Constrain: Modeling Cardinality in Deep Structured Prediction  
<http://proceedings.mlr.press/v80/brukhim18a.html>  
**AUTHORS:** Nataly Brukhim, Amir Globerson  
**HIGHLIGHT:** Recently, several deep learning approaches to structured prediction have been proposed.

69, **TITLE:** Quasi-Monte Carlo Variational Inference  
<http://proceedings.mlr.press/v80/buchholz18a.html>



AUTHORS: Alexander Buchholz, Florian Wenzel, Stephan Mandt  
HIGHLIGHT: As a prominent example, we focus on Monte Carlo variational inference (MCVI) in this paper.

70, TITLE: Path-Level Network Transformation for Efficient Architecture Search  
<http://proceedings.mlr.press/v80/cai18a.html>  
AUTHORS: Han Cai, Jiacheng Yang, Weinan Zhang, Song Han, Yong Yu  
HIGHLIGHT: We introduce a new function-preserving transformation for efficient neural architecture search.

71, TITLE: Improved large-scale graph learning through ridge spectral sparsification  
<http://proceedings.mlr.press/v80/calandriello18a.html>  
AUTHORS: Daniele Calandriello, Alessandro Lazaric, Ioannis Koutis, Michal Valko  
HIGHLIGHT: In this paper, we combine a spectral sparsification routine with Laplacian learning.

72, TITLE: Bayesian Coreset Construction via Greedy Iterative Geodesic Ascent  
<http://proceedings.mlr.press/v80/campbell18a.html>  
AUTHORS: Trevor Campbell, Tamara Broderick  
HIGHLIGHT: To address this shortcoming, we develop greedy iterative geodesic ascent (GIGA), a novel algorithm for Bayesian coreset construction that scales the coreset log-likelihood optimally.

73, TITLE: Adversarial Learning with Local Coordinate Coding  
<http://proceedings.mlr.press/v80/cao18a.html>  
AUTHORS: Jiezhong Cao, Yong Guo, Qingyao Wu, Chunhua Shen, Junzhou Huang, Minghui Tan  
HIGHLIGHT: In this paper, rather than sampling from the pre-defined prior distribution, we propose a Local Coordinate Coding (LCC) based sampling method to improve GANs.

74, TITLE: Fair and Diverse DPP-Based Data Summarization  
<http://proceedings.mlr.press/v80/celis18a.html>  
AUTHORS: Elisa Celis, Vijay Keswani, Damian Straszak, Amit Deshpande, Tarun Kathuria, Nisheeth Vishnoi  
HIGHLIGHT: In this paper we initiate a study of the problem of outputting a diverse and fair summary of a given dataset.

75, TITLE: Conditional Noise-Contrastive Estimation of Unnormalised Models  
<http://proceedings.mlr.press/v80/ceylan18a.html>  
AUTHORS: Ciwan Ceylan, Michael U. Gutmann  
HIGHLIGHT: We here propose a new method that addresses this issue.

76, TITLE: Adversarial Time-to-Event Modeling  
<http://proceedings.mlr.press/v80/chapfuwa18a.html>  
AUTHORS: Paidamoyo Chapfuwa, Chenyang Tao, Chunyuan Li, Courtney Page, Benjamin Goldstein, Lawrence Carin Duke, Ricardo Henao  
HIGHLIGHT: We present a deep-network-based approach that leverages adversarial learning to address a key challenge in modern time-to-event modeling: nonparametric estimation of event-time distributions.

77, TITLE: Stability and Generalization of Learning Algorithms that Converge to Global Optima  
<http://proceedings.mlr.press/v80/charles18a.html>  
AUTHORS: Zachary Charles, Dimitris Papailiopoulos  
HIGHLIGHT: We use our results to establish the stability of first-order methods such as stochastic gradient descent (SGD), gradient descent (GD), randomized coordinate descent (RCD), and the stochastic variance reduced gradient method (SVRG), in both the PL and the strongly convex setting.

78, TITLE: Learning and Memorization  
<http://proceedings.mlr.press/v80/chatterjee18a.html>  
AUTHORS: Satrajit Chatterjee

**HIGHLIGHT:** In this work we examine to what extent this tension exists by exploring if it is possible to generalize by memorizing alone.

79, **TITLE:** On the Theory of Variance Reduction for Stochastic Gradient Monte Carlo  
<http://proceedings.mlr.press/v80/chatterji18a.html>

**AUTHORS:** Niladri Chatterji, Nicolas Flammarion, Yian Ma, Peter Bartlett, Michael Jordan

**HIGHLIGHT:** We analyze these methods under a uniform set of assumptions on the log-posterior distribution, assuming it to be smooth, strongly convex and Hessian Lipschitz.

80, **TITLE:** Hierarchical Clustering with Structural Constraints  
<http://proceedings.mlr.press/v80/chatziafratis18a.html>

**AUTHORS:** Vaggos Chatziafratis, Rad Niazadeh, Moses Charikar

**HIGHLIGHT:** In this paper, we provide provable approximation guarantees for two simple top-down algorithms, using a recently introduced optimization viewpoint of hierarchical clustering with pairwise similarity information (Dasgupta, 2016).

81, **TITLE:** Hierarchical Deep Generative Models for Multi-Rate Multivariate Time Series  
<http://proceedings.mlr.press/v80/che18a.html>

**AUTHORS:** Zhengping Che, Sanjay Purushotham, Guangyu Li, Bo Jiang, Yan Liu

**HIGHLIGHT:** To address this challenge, we propose the Multi-Rate Hierarchical Deep Markov Model (MR-HDMM), a novel deep generative model which uses the latent hierarchical structure with a learnable switch mechanism to capture the temporal dependencies of MR-MTS.

82, **TITLE:** GradNorm: Gradient Normalization for Adaptive Loss Balancing in Deep Multitask Networks  
<http://proceedings.mlr.press/v80/chen18a.html>

**AUTHORS:** Zhao Chen, Vijay Badrinarayanan, Chen-Yu Lee, Andrew Rabinovich

**HIGHLIGHT:** We present a gradient normalization (GradNorm) algorithm that automatically balances training in deep multitask models by dynamically tuning gradient magnitudes.

83, **TITLE:** Weakly Submodular Maximization Beyond Cardinality Constraints: Does Randomization Help Greedy?  
<http://proceedings.mlr.press/v80/chen18b.html>

**AUTHORS:** Lin Chen, Moran Feldman, Amin Karbasi

**HIGHLIGHT:** In this paper, we prove that a randomized version of the greedy algorithm achieves an approximation ratio of  $(1 + 1/\gamma)^{-2}$  for weakly submodular maximization subject to a general matroid constraint, where  $\gamma$  is a parameter measuring the distance from submodularity.

84, **TITLE:** Projection-Free Online Optimization with Stochastic Gradient: From Convexity to Submodularity  
<http://proceedings.mlr.press/v80/chen18c.html>

**AUTHORS:** Lin Chen, Christopher Harshaw, Hamed Hassani, Amin Karbasi

**HIGHLIGHT:** In this work, we propose Meta-Frank-Wolfe, the first online projection-free algorithm that uses stochastic gradient estimates.

85, **TITLE:** Continuous-Time Flows for Efficient Inference and Density Estimation  
<http://proceedings.mlr.press/v80/chen18d.html>

**AUTHORS:** Changyou Chen, Chunyuan Li, Liqun Chen, Wenlin Wang, Yunchen Pu, Lawrence Carin Duke

**HIGHLIGHT:** In this paper, we propose the concept of continuous-time flows (CTFs), a family of diffusion-based methods that are able to asymptotically approach a target distribution.

86, **TITLE:** Scalable Bilinear  $\pi$  Learning Using State and Action Features  
<http://proceedings.mlr.press/v80/chen18e.html>

**AUTHORS:** Yichen Chen, Lihong Li, Mengdi Wang

**HIGHLIGHT:** In this work, we study a primal-dual formulation of the ALP, and develop a scalable, model-free algorithm called bilinear  $\pi$  learning for reinforcement learning when a sampling oracle is provided.

- 87, TITLE: Stein Points  
http://proceedings.mlr.press/v80/chen18f.html  
AUTHORS: Wilson Ye Chen, Lester Mackey, Jackson Gorham, Francois-Xavier Briol, Chris Oates  
HIGHLIGHT: To this end, we present Stein Points.
- 88, TITLE: Learning K-way D-dimensional Discrete Codes for Compact Embedding Representations  
http://proceedings.mlr.press/v80/chen18g.html  
AUTHORS: Ting Chen, Martin Renqiang Min, Yizhou Sun  
HIGHLIGHT: In this work, we propose a much more compact K-way D-dimensional discrete encoding scheme to replace the "one-hot" encoding.
- 89, TITLE: PixelSNAIL: An Improved Autoregressive Generative Model  
http://proceedings.mlr.press/v80/chen18h.html  
AUTHORS: XI Chen, Nikhil Mishra, Mostafa Rohaninejad, Pieter Abbeel  
HIGHLIGHT: In this paper, we describe the resulting model and present state-of-the-art log-likelihood results on heavily benchmarked datasets: CIFAR-10, \$32 \times 32\$ ImageNet and \$64 \times 64\$ ImageNet.
- 90, TITLE: Dynamical Isometry and a Mean Field Theory of RNNs: Gating Enables Signal Propagation in Recurrent Neural Networks  
http://proceedings.mlr.press/v80/chen18i.html  
AUTHORS: Minmin Chen, Jeffrey Pennington, Samuel Schoenholz  
HIGHLIGHT: To simplify our discussion, we introduce a new RNN cell with a simple gating mechanism that we call the minimalRNN and compare it with vanilla RNNs.
- 91, TITLE: Learning to Explain: An Information-Theoretic Perspective on Model Interpretation  
http://proceedings.mlr.press/v80/chen18j.html  
AUTHORS: Jianbo Chen, Le Song, Martin Wainwright, Michael Jordan  
HIGHLIGHT: We introduce instancewise feature selection as a methodology for model interpretation.
- 92, TITLE: Variational Inference and Model Selection with Generalized Evidence Bounds  
http://proceedings.mlr.press/v80/chen18k.html  
AUTHORS: Liqun Chen, Chenyang Tao, Ruiyi Zhang, Ricardo Henao, Lawrence Carin Duke  
HIGHLIGHT: In this work we propose a new variational bound formulation, yielding an estimator that extends beyond the conventional variational bound.
- 93, TITLE: DRACO: Byzantine-resilient Distributed Training via Redundant Gradients  
http://proceedings.mlr.press/v80/chen18l.html  
AUTHORS: Lingjiao Chen, Hongyi Wang, Zachary Charles, Dimitris Papailiopoulos  
HIGHLIGHT: In this work, we present DRACO, a scalable framework for robust distributed training that uses ideas from coding theory.
- 94, TITLE: SADAGRAD: Strongly Adaptive Stochastic Gradient Methods  
http://proceedings.mlr.press/v80/chen18m.html  
AUTHORS: Zaiyi Chen, Yi Xu, Enhong Chen, Tianbao Yang  
HIGHLIGHT: To alleviate this bad dependence, we propose a simple yet novel variant of ADAGRAD for stochastic (weakly) strongly convex optimization.
- 95, TITLE: Covariate Adjusted Precision Matrix Estimation via Nonconvex Optimization  
http://proceedings.mlr.press/v80/chen18n.html  
AUTHORS: Jinghui Chen, Pan Xu, Lingxiao Wang, Jian Ma, Quanquan Gu  
HIGHLIGHT: To solve this estimator, we propose an alternating gradient descent algorithm with hard thresholding.

- 96, TITLE: End-to-End Learning for the Deep Multivariate Probit Model  
http://proceedings.mlr.press/v80/chen18o.html  
AUTHORS: Di Chen, Yexiang Xue, Carla Gomes  
HIGHLIGHT: We propose a flexible deep generalization of the classic MVP, the Deep Multivariate Probit Model (DMVP), which is an end-to-end learning scheme that uses an efficient parallel sampling process of the multivariate probit model to exploit GPU-boosted deep neural networks.
- 97, TITLE: Stochastic Training of Graph Convolutional Networks with Variance Reduction  
http://proceedings.mlr.press/v80/chen18p.html  
AUTHORS: Jianfei Chen, Jun Zhu, Le Song  
HIGHLIGHT: In this paper, we develop control variate based algorithms with new theoretical guarantee to converge to a local optimum of GCN regardless of the neighbor sampling size.
- 98, TITLE: Extreme Learning to Rank via Low Rank Assumption  
http://proceedings.mlr.press/v80/cheng18a.html  
AUTHORS: Minhao Cheng, Ian Davidson, Cho-Jui Hsieh  
HIGHLIGHT: To address this situation, we propose a Factorization RankSVM algorithm which learns a series of  $k$  basic ranking functions and then constructs for each user a local ranking function that is a combination of them.
- 99, TITLE: Learning a Mixture of Two Multinomial Logits  
http://proceedings.mlr.press/v80/chierichetti18a.html  
AUTHORS: Flavio Chierichetti, Ravi Kumar, Andrew Tomkins  
HIGHLIGHT: In this paper we give the first polynomial-time algorithms for exact learning of uniform mixtures of two MNLs.
- 100, TITLE: Structured Evolution with Compact Architectures for Scalable Policy Optimization  
http://proceedings.mlr.press/v80/choromanski18a.html  
AUTHORS: Krzysztof Choromanski, Mark Rowland, Vikas Sindhwani, Richard Turner, Adrian Weller  
HIGHLIGHT: We present a new method of blackbox optimization via gradient approximation with the use of structured random orthogonal matrices, providing more accurate estimators than baselines and with provable theoretical guarantees.
- 101, TITLE: Path Consistency Learning in Tsallis Entropy Regularized MDPs  
http://proceedings.mlr.press/v80/chow18a.html  
AUTHORS: Yinlam Chow, Ofir Nachum, Mohammad Ghavamzadeh  
HIGHLIGHT: In this paper, we follow the work of Nachum et al. (2017) in the soft ERL setting, and propose a class of novel path consistency learning (PCL) algorithms, called sparse PCL, for the sparse ERL problem that can work with both on-policy and off-policy data.
- 102, TITLE: An Iterative, Sketching-based Framework for Ridge Regression  
http://proceedings.mlr.press/v80/chowdhury18a.html  
AUTHORS: Agniva Chowdhury, Jiasen Yang, Petros Drineas  
HIGHLIGHT: We present a simple, iterative, sketching-based algorithm for ridge regression that guarantees high-quality approximations to the optimal solution vector.
- 103, TITLE: Stochastic Wasserstein Barycenters  
http://proceedings.mlr.press/v80/claici18a.html  
AUTHORS: Sebastian Claiici, Edward Chien, Justin Solomon  
HIGHLIGHT: We present a stochastic algorithm to compute the barycenter of a set of probability distributions under the Wasserstein metric from optimal transport.

104, TITLE: Self-Consistent Trajectory Autoencoder: Hierarchical Reinforcement Learning with Trajectory Embeddings  
<http://proceedings.mlr.press/v80/co-reyes18a.html>  
AUTHORS: John Co-Reyes, YuXuan Liu, Abhishek Gupta, Benjamin Eysenbach, Pieter Abbeel, Sergey Levine  
HIGHLIGHT: In this work, we take a representation learning perspective on hierarchical reinforcement learning, where the problem of learning lower layers in a hierarchy is transformed into the problem of learning trajectory-level generative models.

105, TITLE: On Acceleration with Noise-Corrupted Gradients  
<http://proceedings.mlr.press/v80/cohen18a.html>  
AUTHORS: Michael Cohen, Jelena Diakonikolas, Lorenzo Orecchia  
HIGHLIGHT: This study leverages the simplicity of AGDP and its analysis to clarify the interaction between noise and acceleration and to suggest modifications to the algorithm that reduce the mean and variance of the error incurred due to the gradient noise.

106, TITLE: Online Linear Quadratic Control  
<http://proceedings.mlr.press/v80/cohen18b.html>  
AUTHORS: Alon Cohen, Avinatan Hasidim, Tomer Koren, Nevena Lazic, Yishay Mansour, Kunal Talwar  
HIGHLIGHT: We present the first efficient online learning algorithms in this setting that guarantee  $\mathcal{O}(\sqrt{T})$  regret under mild assumptions, where  $T$  is the time horizon.

107, TITLE: GEP-PG: Decoupling Exploration and Exploitation in Deep Reinforcement Learning Algorithms  
<http://proceedings.mlr.press/v80/colas18a.html>  
AUTHORS: Cédric Colas, Olivier Sigaud, Pierre-Yves Oudeyer  
HIGHLIGHT: In this paper, we present the GEP-PG approach, taking the best of both worlds by sequentially combining a Goal Exploration Process and two variants of DDPG .

108, TITLE: Efficient Model-Based Deep Reinforcement Learning with Variational State Tabulation  
<http://proceedings.mlr.press/v80/corneil18a.html>  
AUTHORS: Dane Corneil, Wulfram Gerstner, Johanni Brea  
HIGHLIGHT: In this article we introduce Variational State Tabulation (VaST), which maps an environment with a high-dimensional state space (e.g. the space of visual inputs) to an abstract tabular model.

109, TITLE: Online Learning with Abstention  
<http://proceedings.mlr.press/v80/cortes18a.html>  
AUTHORS: Corinna Cortes, Giulia DeSalvo, Claudio Gentile, Mehryar Mohri, Scott Yang  
HIGHLIGHT: We present an extensive study of a key problem in online learning where the learner can opt to abstain from making a prediction, at a certain cost.

110, TITLE: Constrained Interacting Submodular Groupings  
<http://proceedings.mlr.press/v80/cotter18a.html>  
AUTHORS: Andrew Cotter, Mahdi Milani Fard, Seungil You, Maya Gupta, Jeff Bilmes  
HIGHLIGHT: We introduce the problem of grouping a finite ground set into blocks where each block is a subset of the ground set and where: (i) the blocks are individually highly valued by a submodular function (both robustly and in the average case) while satisfying block-specific matroid constraints; and (ii) block scores interact where blocks are jointly scored highly, thus making the blocks mutually non-redundant.

111, TITLE: Inference Suboptimality in Variational Autoencoders  
<http://proceedings.mlr.press/v80/cremer18a.html>  
AUTHORS: Chris Cremer, Xuechen Li, David Duvenaud  
HIGHLIGHT: We examine approximate inference in variational autoencoders in terms of these factors.

112, TITLE: Mix & Match Agent Curricula for Reinforcement Learning  
<http://proceedings.mlr.press/v80/czarnecki18a.html>

AUTHORS: Wojciech Czarnecki, Siddhant Jayakumar, Max Jaderberg, Leonard Hasenclever, Yee Whye Teh, Nicolas Heess, Simon Osindero, Razvan Pascanu  
HIGHLIGHT: We introduce Mix and match (M&M) – a training framework designed to facilitate rapid and effective learning in RL agents that would be too slow or too challenging to train otherwise. The key innovation is a procedure that allows us to automatically form a curriculum over agents.

113, TITLE: Implicit Quantile Networks for Distributional Reinforcement Learning  
<http://proceedings.mlr.press/v80/dabney18a.html>

AUTHORS: Will Dabney, Georg Ostrovski, David Silver, Remi Munos  
HIGHLIGHT: In this work, we build on recent advances in distributional reinforcement learning to give a generally applicable, flexible, and state-of-the-art distributional variant of DQN.

114, TITLE: Learning Steady-States of Iterative Algorithms over Graphs

<http://proceedings.mlr.press/v80/dai18a.html>

AUTHORS: Hanjun Dai, Zornitsa Kozareva, Bo Dai, Alex Smola, Le Song  
HIGHLIGHT: In this paper, we propose an embedding representation for iterative algorithms over graphs, and design a learning method which alternates between updating the embeddings and projecting them onto the steady-state constraints.

115, TITLE: Adversarial Attack on Graph Structured Data

<http://proceedings.mlr.press/v80/dai18b.html>

AUTHORS: Hanjun Dai, Hui Li, Tian Tian, Xin Huang, Lin Wang, Jun Zhu, Le Song  
HIGHLIGHT: In this paper, we focus on the adversarial attacks that fool deep learning models by modifying the combinatorial structure of data.

116, TITLE: SBED: Convergent Reinforcement Learning with Nonlinear Function Approximation

<http://proceedings.mlr.press/v80/dai18c.html>

AUTHORS: Bo Dai, Albert Shaw, Lihong Li, Lin Xiao, Niao He, Zhen Liu, Jianshu Chen, Le Song  
HIGHLIGHT: In this paper, we revisit the Bellman equation, and reformulate it into a novel primal-dual optimization problem using Nesterov’s smoothing technique and the Legendre-Fenchel transformation.

117, TITLE: Compressing Neural Networks using the Variational Information Bottleneck

<http://proceedings.mlr.press/v80/dai18d.html>

AUTHORS: Bin Dai, Chen Zhu, Baining Guo, David Wipf  
HIGHLIGHT: In this paper we focus on pruning individual neurons, which can simultaneously trim model size, FLOPs, and run-time memory.

118, TITLE: Asynchronous Byzantine Machine Learning (the case of SGD)

<http://proceedings.mlr.press/v80/damaskinos18a.html>

AUTHORS: Georgios Damaskinos, El Mahdi El Mhamdi, Rachid Guerraoui, Rhicheck Patra, Mahsa Taziki  
HIGHLIGHT: We introduce Kardam, the first distributed asynchronous stochastic gradient descent (SGD) algorithm that copes with Byzantine workers.

119, TITLE: Escaping Saddles with Stochastic Gradients

<http://proceedings.mlr.press/v80/daneshmand18a.html>

AUTHORS: Hadi Daneshmand, Jonas Kohler, Aurelien Lucchi, Thomas Hofmann  
HIGHLIGHT: Based upon this observation we propose a new assumption under which we show that the injection of explicit, isotropic noise usually applied to make gradient descent escape saddle points can successfully be replaced by a simple SGD step.

120, TITLE: Minibatch Gibbs Sampling on Large Graphical Models

<http://proceedings.mlr.press/v80/desa18a.html>

AUTHORS: Chris De Sa, Vincent Chen, Wing Wong

**HIGHLIGHT:** In this paper, we show how this cost can be reduced by using minibatching: subsampling the factors to form an estimate of their sum.

121, **TITLE:** Stochastic Video Generation with a Learned Prior

<http://proceedings.mlr.press/v80/denton18a.html>

**AUTHORS:** Emily Denton, Rob Fergus

**HIGHLIGHT:** In this paper we introduce a video generation model with a learned prior over stochastic latent variables at each time step.

122, **TITLE:** Decomposition of Uncertainty in Bayesian Deep Learning for Efficient and Risk-sensitive Learning

<http://proceedings.mlr.press/v80/depeweg18a.html>

**AUTHORS:** Stefan Depeweg, Jose-Miguel Hernandez-Lobato, Finale Doshi-Velez, Steffen Udluft

**HIGHLIGHT:** Using these models we show how to perform and utilize a decomposition of uncertainty in aleatoric and epistemic components for decision making purposes.

123, **TITLE:** Accurate Inference for Adaptive Linear Models

<http://proceedings.mlr.press/v80/deshpande18a.html>

**AUTHORS:** Yash Deshpande, Lester Mackey, Vasilis Syrgkanis, Matt Taddy

**HIGHLIGHT:** We develop a general method –  $\mathbf{W}$ -decorrelation – for transforming the bias of adaptive linear regression estimators into variance.

124, **TITLE:** Variational Network Inference: Strong and Stable with Concrete Support

<http://proceedings.mlr.press/v80/dezfouli18a.html>

**AUTHORS:** Amir Dezfouli, Edwin Bonilla, Richard Nock

**HIGHLIGHT:** We address these limitations by proposing a model that incorporates a Gaussian process prior on a network-independent component and formally proving that we get algorithmic stability for free while providing a novel perspective on model stability as well as robustness results and precise intervals for key inference parameters.

125, **TITLE:** Modeling Sparse Deviations for Compressed Sensing using Generative Models

<http://proceedings.mlr.press/v80/dhar18a.html>

**AUTHORS:** Manik Dhar, Aditya Grover, Stefano Ermon

**HIGHLIGHT:** We propose Sparse-Gen, a framework that allows for sparse deviations from the support set, thereby achieving the best of both worlds by using a domain specific prior and allowing reconstruction over the full space of signals.

126, **TITLE:** Alternating Randomized Block Coordinate Descent

<http://proceedings.mlr.press/v80/diakonikolas18a.html>

**AUTHORS:** Jelena Diakonikolas, Lorenzo Orecchia

**HIGHLIGHT:** We show that the answer is “no” as long as the least smooth block can be optimized exactly – an assumption that is also needed in the setting of alternating minimization.

127, **TITLE:** Learning to Act in Decentralized Partially Observable MDPs

<http://proceedings.mlr.press/v80/dibangoye18a.html>

**AUTHORS:** Jilles Dibangoye, Olivier Buffet

**HIGHLIGHT:** In this paper, we restrict attention to plans, which are simpler to store and update than policies.

128, **TITLE:** Leveraging Well-Conditioned Bases: Streaming and Distributed Summaries in Minkowski  $\ell_p$ -Norms

<http://proceedings.mlr.press/v80/dickens18a.html>

**AUTHORS:** Charlie Dickens, Graham Cormode, David Woodruff

**HIGHLIGHT:** Unlike previous algorithms for such norms, we give algorithms that are (1) deterministic, (2) work simultaneously for every  $\ell_p$ , including  $\ell_p = \infty$ , and (3) can be implemented in both distributed and streaming environments.

- 129, TITLE: Noisin: Unbiased Regularization for Recurrent Neural Networks  
<http://proceedings.mlr.press/v80/dieng18a.html>  
AUTHORS: Adji Bousso Dieng, Rajesh Ranganath, Jaan Altosaar, David Blei  
HIGHLIGHT: In this paper we develop Noisin, a new method for regularizing RNNs.
- 130, TITLE: Discovering and Removing Exogenous State Variables and Rewards for Reinforcement Learning  
<http://proceedings.mlr.press/v80/dietterich18a.html>  
AUTHORS: Thomas Dietterich, George Trimponias, Zhitang Chen  
HIGHLIGHT: We develop two algorithms for discovering the exogenous variables and test them on several MDPs.
- 131, TITLE: Coordinated Exploration in Concurrent Reinforcement Learning  
<http://proceedings.mlr.press/v80/dimakopoulou18a.html>  
AUTHORS: Maria Dimakopoulou, Benjamin Van Roy  
HIGHLIGHT: As an alternative, we propose seed sampling, which extends posterior sampling in a manner that meets these requirements.
- 132, TITLE: Probabilistic Recurrent State-Space Models  
<http://proceedings.mlr.press/v80/doerr18a.html>  
AUTHORS: Andreas Doerr, Christian Daniel, Martin Schiegg, Nguyen-Tuong Duy, Stefan Schaal, Marc Toussaint, Trimpe Sebastian  
HIGHLIGHT: We propose a novel model formulation and a scalable training algorithm based on doubly stochastic variational inference and Gaussian processes.
- 133, TITLE: Randomized Block Cubic Newton Method  
<http://proceedings.mlr.press/v80/doikov18a.html>  
AUTHORS: Nikita Doikov, Peter Richtarik, University Edinburgh  
HIGHLIGHT: To this effect we propose and analyze a randomized block cubic Newton (RBCN) method, which in each iteration builds a model of the objective function formed as the sum of the natural models of its three components: a linear model with a quadratic regularizer for the differentiable term, a quadratic model with a cubic regularizer for the twice differentiable term, and perfect (proximal) model for the nonsmooth term.
- 134, TITLE: Low-Rank Riemannian Optimization on Positive Semidefinite Stochastic Matrices with Applications to Graph Clustering  
<http://proceedings.mlr.press/v80/douik18a.html>  
AUTHORS: Ahmed Douik, Babak Hassibi  
HIGHLIGHT: This paper develops a Riemannian optimization framework for solving optimization problems on the set of symmetric positive semidefinite stochastic matrices.
- 135, TITLE: Essentially No Barriers in Neural Network Energy Landscape  
<http://proceedings.mlr.press/v80/draxler18a.html>  
AUTHORS: Felix Draxler, Kambis Veschgini, Manfred Salmhofer, Fred Hamprecht  
HIGHLIGHT: Relaxing from linear interpolations, we construct continuous paths between minima of recent neural network architectures on CIFAR10 and CIFAR100.
- 136, TITLE: Weakly Consistent Optimal Pricing Algorithms in Repeated Posted-Price Auctions with Strategic Buyer  
<http://proceedings.mlr.press/v80/drutsa18a.html>  
AUTHORS: Alexey Drutsa  
HIGHLIGHT: We propose a novel algorithm that never decreases offered prices and has a tight strategic regret bound of  $\Theta(\log \log T)$ .
- 137, TITLE: On the Power of Over-parametrization in Neural Networks with Quadratic Activation



<http://proceedings.mlr.press/v80/du18a.html>

AUTHORS: Simon Du, Jason Lee

HIGHLIGHT: We provide new theoretical insights on why over-parametrization is effective in learning neural networks.

138, TITLE: Gradient Descent Learns One-hidden-layer CNN: Don't be Afraid of Spurious Local Minima

<http://proceedings.mlr.press/v80/du18b.html>

AUTHORS: Simon Du, Jason Lee, Yuandong Tian, Aarti Singh, Barnabas Poczos

HIGHLIGHT: We consider the problem of learning an one-hidden-layer neural network with non-overlapping convolutional layer and ReLU activation function, i.e.,  $f(Z; w, a) = \sum_j a_j \sigma(w^{\top} Z_j)$ , in which both the convolutional weights  $w$  and the output weights  $a$  are parameters to be learned.

139, TITLE: Investigating Human Priors for Playing Video Games

<http://proceedings.mlr.press/v80/dubey18a.html>

AUTHORS: Rachit Dubey, Pulkit Agrawal, Deepak Pathak, Tom Griffiths, Alexei Efros

HIGHLIGHT: This paper investigates the role of human priors for solving video games.

140, TITLE: A Distributed Second-Order Algorithm You Can Trust

<http://proceedings.mlr.press/v80/duenner18a.html>

AUTHORS: Celestine Duenner, Aurelien Lucchi, Matilde Gargiani, An Bian, Thomas Hofmann, Martin Jaggi

HIGHLIGHT: In this paper we present a new algorithm for distributed training of generalized linear models that only requires the computation of diagonal blocks of the Hessian matrix on the individual workers.

141, TITLE: Computational Optimal Transport: Complexity by Accelerated Gradient Descent Is Better Than by Sinkhorn's Algorithm

<http://proceedings.mlr.press/v80/dvurechensky18a.html>

AUTHORS: Pavel Dvurechensky, Alexander Gasnikov, Alexey Kroshnin

HIGHLIGHT: We analyze two algorithms for approximating the general optimal transport (OT) distance between two discrete distributions of size  $n$ , up to accuracy  $\epsilon$ .

142, TITLE: Entropy-SGD optimizes the prior of a PAC-Bayes bound: Generalization properties of Entropy-SGD and data-dependent priors

<http://proceedings.mlr.press/v80/dziugaite18a.html>

AUTHORS: Gintare Karolina Dziugaite, Daniel Roy

HIGHLIGHT: We show that Entropy-SGD (Chaudhari et al., 2017), when viewed as a learning algorithm, optimizes a PAC-Bayes bound on the risk of a Gibbs (posterior) classifier, i.e., a randomized classifier obtained by a risk-sensitive perturbation of the weights of a learned classifier.

143, TITLE: Beyond the One-Step Greedy Approach in Reinforcement Learning

<http://proceedings.mlr.press/v80/efroni18a.html>

AUTHORS: Yonathan Efroni, Gal Dalal, Bruno Scherrer, Shie Mannor

HIGHLIGHT: In this work, we introduce the first such analysis.

144, TITLE: Parallel and Streaming Algorithms for K-Core Decomposition

<http://proceedings.mlr.press/v80/esfandiari18a.html>

AUTHORS: Hossein Esfandiari, Silvio Lattanzi, Vahab Mirrokni

HIGHLIGHT: We present the first distributed and the first streaming algorithms to compute and maintain an approximate k-core decomposition with provable guarantees.

145, TITLE: IMPALA: Scalable Distributed Deep-RL with Importance Weighted Actor-Learner Architectures

<http://proceedings.mlr.press/v80/espeholt18a.html>

AUTHORS: Lasse Espeholt, Hubert Soyer, Remi Munos, Karen Simonyan, Vlad Mnih, Tom Ward, Yotam Doron, Vlad Firoiu, Tim Harley, Iain Dunning, Shane Legg, Koray Kavukcuoglu

**HIGHLIGHT:** In this work we aim to solve a large collection of tasks using a single reinforcement learning agent with a single set of parameters.

146, **TITLE:** Scalable Gaussian Processes with Grid-Structured Eigenfunctions (GP-GRIEF)

<http://proceedings.mlr.press/v80/evans18a.html>

**AUTHORS:** Trefor Evans, Prasanth Nair

**HIGHLIGHT:** We introduce a kernel approximation strategy that enables computation of the Gaussian process log marginal likelihood and all hyperparameter derivatives in  $O(p)$  time.

147, **TITLE:** The Limits of Maxing, Ranking, and Preference Learning

<http://proceedings.mlr.press/v80/falahatgar18a.html>

**AUTHORS:** Moein Falahatgar, Ayush Jain, Alon Orlitsky, Venkatadheeraj Pichapati, Vaishakh Ravindrakumar

**HIGHLIGHT:** We present a comprehensive understanding of three important problems in PAC preference learning: maximum selection (maxing), ranking, and estimating all pairwise preference probabilities, in the adaptive setting.

148, **TITLE:** BOHB: Robust and Efficient Hyperparameter Optimization at Scale

<http://proceedings.mlr.press/v80/falkner18a.html>

**AUTHORS:** Stefan Falkner, Aaron Klein, Frank Hutter

**HIGHLIGHT:** We propose a new practical state-of-the-art hyperparameter optimization method, which consistently outperforms both Bayesian optimization and Hyperband on a wide range of problem types, including high-dimensional toy functions, support vector machines, feed-forward neural networks, Bayesian neural networks, deep reinforcement learning, and convolutional neural networks.

149, **TITLE:** More Robust Doubly Robust Off-policy Evaluation

<http://proceedings.mlr.press/v80/farajtabar18a.html>

**AUTHORS:** Mehrdad Farajtabar, Yinlam Chow, Mohammad Ghavamzadeh

**HIGHLIGHT:** In this paper, we propose alternative DR estimators, called more robust doubly robust (MRDR), that learn the model parameter by minimizing the variance of the DR estimator.

150, **TITLE:** Efficient and Consistent Adversarial Bipartite Matching

<http://proceedings.mlr.press/v80/fathony18a.html>

**AUTHORS:** Rizal Fathony, Sima Behpour, Xinhua Zhang, Brian Ziebart

**HIGHLIGHT:** We propose adversarial bipartite matching to avoid both of these limitations.

151, **TITLE:** Global Convergence of Policy Gradient Methods for the Linear Quadratic Regulator

<http://proceedings.mlr.press/v80/fazel18a.html>

**AUTHORS:** Maryam Fazel, Rong Ge, Sham Kakade, Mehran Mesbahi

**HIGHLIGHT:** Direct policy gradient methods for reinforcement learning and continuous control problems are a popular approach for a variety of reasons: 1) they are easy to implement without explicit knowledge of the underlying model, 2) they are an “end-to-end” approach, directly optimizing the performance metric of interest, 3) they inherently allow for richly parameterized policies.

152, **TITLE:** CRVI: Convex Relaxation for Variational Inference

<http://proceedings.mlr.press/v80/fazelnia18a.html>

**AUTHORS:** Ghazal Fazelnia, John Paisley

**HIGHLIGHT:** We present a new technique for solving non-convex variational inference optimization problems.

153, **TITLE:** Fourier Policy Gradients

<http://proceedings.mlr.press/v80/fellows18a.html>

**AUTHORS:** Matthew Fellows, Kamil Ciosek, Shimon Whiteson

**HIGHLIGHT:** We propose a new way of deriving policy gradient updates for reinforcement learning.

154, TITLE: Nonparametric variable importance using an augmented neural network with multi-task learning  
<http://proceedings.mlr.press/v80/feng18a.html>  
AUTHORS: Jean Feng, Brian Williamson, Noah Simon, Marco Carone  
HIGHLIGHT: In this paper, we connect these ideas in nonparametric variable importance to machine learning, and provide a method for efficient estimation of variable importance when building a predictive model using a neural network.

155, TITLE: Closed-form Marginal Likelihood in Gamma-Poisson Matrix Factorization  
<http://proceedings.mlr.press/v80/filstroff18a.html>  
AUTHORS: Louis Filstroff, Alberto Lumbreras, Cédric Fève  
HIGHLIGHT: We present novel understandings of the Gamma-Poisson (GaP) model, a probabilistic matrix factorization model for count data.

156, TITLE: Automatic Goal Generation for Reinforcement Learning Agents  
<http://proceedings.mlr.press/v80/florensa18a.html>  
AUTHORS: Carlos Florensa, David Held, Xinyang Geng, Pieter Abbeel  
HIGHLIGHT: Instead, we propose a method that allows an agent to automatically discover the range of tasks that it is capable of performing in its environment.

157, TITLE: DiCE: The Infinitely Differentiable Monte Carlo Estimator  
<http://proceedings.mlr.press/v80/foerster18a.html>  
AUTHORS: Jakob Foerster, Gregory Farquhar, Maruan Al-Shedivat, Tim Rocktäschel, Eric Xing, Shimon Whiteson  
HIGHLIGHT: To address all these shortcomings in a unified way, we introduce DiCE, which provides a single objective that can be differentiated repeatedly, generating correct estimators of derivatives of any order in SCGs.

158, TITLE: Practical Contextual Bandits with Regression Oracles  
<http://proceedings.mlr.press/v80/foster18a.html>  
AUTHORS: Dylan Foster, Alekh Agarwal, Miroslav Dudik, Haipeng Luo, Robert Schapire  
HIGHLIGHT: We present a new technique that has the empirical and computational advantages of realizability-based approaches combined with the flexibility of agnostic methods.

159, TITLE: Generative Temporal Models with Spatial Memory for Partially Observed Environments  
<http://proceedings.mlr.press/v80/fraccaro18a.html>  
AUTHORS: Marco Fraccaro, Danilo Rezende, Yori Zwols, Alexander Pritzel, S. M. Ali Eslami, Fabio Viola  
HIGHLIGHT: In this work we introduce a novel action-conditioned generative model of such challenging environments.

160, TITLE: ADMM and Accelerated ADMM as Continuous Dynamical Systems  
<http://proceedings.mlr.press/v80/franca18a.html>  
AUTHORS: Guilherme Franca, Daniel Robinson, Rene Vidal  
HIGHLIGHT: We employ the direct method of Lyapunov to analyze the stability of critical points of the dynamical systems and to obtain associated convergence rates.

161, TITLE: Bilevel Programming for Hyperparameter Optimization and Meta-Learning  
<http://proceedings.mlr.press/v80/franceschi18a.html>  
AUTHORS: Luca Franceschi, Paolo Frasconi, Saverio Salzo, Riccardo Grazi, Massimiliano Pontil  
HIGHLIGHT: We introduce a framework based on bilevel programming that unifies gradient-based hyperparameter optimization and meta-learning.

162, TITLE: Efficient Bias-Span-Constrained Exploration-Exploitation in Reinforcement Learning  
<http://proceedings.mlr.press/v80/fruit18a.html>  
AUTHORS: Ronan Fruit, Matteo Pirota, Alessandro Lazaric, Ronald Ortner

**HIGHLIGHT:** We introduce SCAL, an algorithm designed to perform efficient exploration-exploration in any unknown weakly-communicating Markov Decision Process (MDP) for which an upper bound  $c$  on the span of the optimal bias function is known.

163, **TITLE:** Addressing Function Approximation Error in Actor-Critic Methods  
<http://proceedings.mlr.press/v80/fujimoto18a.html>

**AUTHORS:** Scott Fujimoto, Herke Hoof, David Meger

**HIGHLIGHT:** We evaluate our method on the suite of OpenAI gym tasks, outperforming the state of the art in every environment tested.

164, **TITLE:** Clipped Action Policy Gradient  
<http://proceedings.mlr.press/v80/fujita18a.html>

**AUTHORS:** Yasuhiro Fujita, Shin-ichi Maeda

**HIGHLIGHT:** We propose a policy gradient estimator that exploits the knowledge of actions being clipped to reduce the variance in estimation.

165, **TITLE:** Born Again Neural Networks  
<http://proceedings.mlr.press/v80/furlanello18a.html>

**AUTHORS:** Tommaso Furlanello, Zachary Lipton, Michael Tschannen, Laurent Itti, Anima Anandkumar

**HIGHLIGHT:** We study KD from a new perspective: rather than compressing models, we train students parameterized identically to their teachers.

166, **TITLE:** The Generalization Error of Dictionary Learning with Moreau Envelopes  
<http://proceedings.mlr.press/v80/georgogiannis18a.html>

**AUTHORS:** Alexandros Georgogiannis

**HIGHLIGHT:** The Generalization Error of Dictionary Learning with Moreau Envelopes

167, **TITLE:** Local Private Hypothesis Testing: Chi-Square Tests  
<http://proceedings.mlr.press/v80/gaboardi18a.html>

**AUTHORS:** Marco Gaboardi, Ryan Rogers

**HIGHLIGHT:** We explore the design of private hypothesis tests in the local model, where each data entry is perturbed to ensure the privacy of each participant.

168, **TITLE:** Inductive Two-Layer Modeling with Parametric Bregman Transfer  
<http://proceedings.mlr.press/v80/ganapathiraman18a.html>

**AUTHORS:** Vignesh Ganapathiraman, Zhan Shi, Xinhua Zhang, Yaoliang Yu

**HIGHLIGHT:** In this paper, we develop a new inductive learning framework for parametric transfer functions using matching losses.

169, **TITLE:** Hyperbolic Entailment Cones for Learning Hierarchical Embeddings  
<http://proceedings.mlr.press/v80/ganea18a.html>

**AUTHORS:** Octavian Ganea, Gary Becigneul, Thomas Hofmann

**HIGHLIGHT:** We here present a novel method to embed directed acyclic graphs.

170, **TITLE:** Parameterized Algorithms for the Matrix Completion Problem  
<http://proceedings.mlr.press/v80/ganian18a.html>

**AUTHORS:** Robert Ganian, Iyad Kanj, Sebastian Ordyniak, Stefan Szeider

**HIGHLIGHT:** We consider two matrix completion problems, in which we are given a matrix with missing entries and the task is to complete the matrix in a way that (1) minimizes the rank, or (2) minimizes the number of distinct rows.

171, **TITLE:** Synthesizing Programs for Images using Reinforced Adversarial Learning  
<http://proceedings.mlr.press/v80/ganin18a.html>

**AUTHORS:** Yaroslav Ganin, Tejas Kulkarni, Igor Babuschkin, S. M. Ali Eslami, Oriol Vinyals  
**HIGHLIGHT:** To mitigate these issues, we present SPIRAL, an adversarially trained agent that generates a program which is executed by a graphics engine to interpret and sample images.

172, **TITLE:** Spotlight: Optimizing Device Placement for Training Deep Neural Networks  
<http://proceedings.mlr.press/v80/gao18a.html>  
**AUTHORS:** Yuanxiang Gao, Li Chen, Baochun Li  
**HIGHLIGHT:** In this paper, we propose Spotlight, a new reinforcement learning algorithm based on proximal policy optimization, designed specifically for finding an optimal device placement for training DNNs.

173, **TITLE:** Parallel Bayesian Network Structure Learning  
<http://proceedings.mlr.press/v80/gao18b.html>  
**AUTHORS:** Tian Gao, Dennis Wei  
**HIGHLIGHT:** As a natural progression, we investigate parallel learning of BN structures via multiple learning agents simultaneously, where each agent learns one local subgraph at a time.

174, **TITLE:** Structured Output Learning with Abstention: Application to Accurate Opinion Prediction  
<http://proceedings.mlr.press/v80/garcia18a.html>  
**AUTHORS:** Alexandre Garcia, Chloé Clavel, Slim ESSID, Florence d'Alché-Buc  
**HIGHLIGHT:** Motivated by Supervised Opinion Analysis, we propose a novel framework devoted to Structured Output Learning with Abstention (SOLA).

175, **TITLE:** Conditional Neural Processes  
<http://proceedings.mlr.press/v80/garnelo18a.html>  
**AUTHORS:** Marta Garnelo, Dan Rosenbaum, Christopher Maddison, Tiago Ramalho, David Saxton, Murray Shanahan, Yee Whye Teh, Danilo Rezende, S. M. Ali Eslami  
**HIGHLIGHT:** In this paper we propose a family of neural models, Conditional Neural Processes (CNPs), that combine the benefits of both.

176, **TITLE:** Temporal Poisson Square Root Graphical Models  
<http://proceedings.mlr.press/v80/geng18a.html>  
**AUTHORS:** Sinong Geng, Zhaobin Kuang, Peggy Peissig, David Page  
**HIGHLIGHT:** We propose temporal Poisson square root graphical models (TPSQRs), a generalization of Poisson square root graphical models (PSQRs) specifically designed for modeling longitudinal event data.

177, **TITLE:** Budgeted Experiment Design for Causal Structure Learning  
<http://proceedings.mlr.press/v80/ghassami18a.html>  
**AUTHORS:** AmirEmad Ghassami, Saber Salehkaleybar, Negar Kiyavash, Elias Bareinboim  
**HIGHLIGHT:** We study the problem of causal structure learning when the experimenter is limited to perform at most  $k$  non-adaptive experiments of size  $1$ .

178, **TITLE:** Linear Spectral Estimators and an Application to Phase Retrieval  
<http://proceedings.mlr.press/v80/ghods18a.html>  
**AUTHORS:** Ramina Ghods, Andrew Lan, Tom Goldstein, Christoph Studer  
**HIGHLIGHT:** We propose a novel class of estimators suitable for general nonlinear measurement systems, called linear spectral estimators (LSPEs), which can be used to compute accurate initialization vectors for phase retrieval problems.

179, **TITLE:** Structured Variational Learning of Bayesian Neural Networks with Horseshoe Priors  
<http://proceedings.mlr.press/v80/ghosh18a.html>  
**AUTHORS:** Soumya Ghosh, Jiayu Yao, Finale Doshi-Velez  
**HIGHLIGHT:** In this work, we propose several modeling and inference advances that consistently improve the compactness of the model learned while maintaining predictive performance, especially in smaller-sample settings including reinforcement learning.

180, TITLE: Learning Maximum-A-Posteriori Perturbation Models for Structured Prediction in Polynomial Time

<http://proceedings.mlr.press/v80/ghoshal18a.html>

AUTHORS: Asish Ghoshal, Jean Honorio

HIGHLIGHT: In this paper, we propose a provably polynomial time randomized algorithm for learning the parameters of perturbed MAP predictors.

181, TITLE: Robust and Scalable Models of Microbiome Dynamics

<http://proceedings.mlr.press/v80/gibson18a.html>

AUTHORS: Travis Gibson, Georg Gerber

HIGHLIGHT: In this work we present a Bayesian nonparametric model and associated efficient inference algorithm that addresses the key conceptual and practical challenges of learning microbial dynamics from time series microbe abundance data.

182, TITLE: Non-linear motor control by local learning in spiking neural networks

<http://proceedings.mlr.press/v80/gilra18a.html>

AUTHORS: Aditya Gilra, Wulfram Gerstner

HIGHLIGHT: We propose a network architecture, termed differential feedforward, and show that it gives a lower test error than other feedforward and recurrent architectures.

183, TITLE: Learning One Convolutional Layer with Overlapping Patches

<http://proceedings.mlr.press/v80/goel18a.html>

AUTHORS: Surbhi Goel, Adam Klivans, Raghu Meka

HIGHLIGHT: We give the first provably efficient algorithm for learning a one hidden layer convolutional network with respect to a general class of (potentially overlapping) patches under mild conditions on the underlying distribution.

184, TITLE: Visualizing and Understanding Atari Agents

<http://proceedings.mlr.press/v80/greydanus18a.html>

AUTHORS: Samuel Greydanus, Anurag Koul, Jonathan Dodge, Alan Fern

HIGHLIGHT: In this paper, we take a step toward explaining deep RL agents through a case study using Atari 2600 environments.

185, TITLE: Learning Policy Representations in Multiagent Systems

<http://proceedings.mlr.press/v80/grover18a.html>

AUTHORS: Aditya Grover, Maruan Al-Shedivat, Jayesh Gupta, Yuri Burda, Harrison Edwards

HIGHLIGHT: We propose a general learning framework for modeling agent behavior in any multiagent system using only a handful of interaction data.

186, TITLE: Faster Derivative-Free Stochastic Algorithm for Shared Memory Machines

<http://proceedings.mlr.press/v80/gu18a.html>

AUTHORS: Bin Gu, Zhouyuan Huo, Cheng Deng, Heng Huang

HIGHLIGHT: Recently, the first asynchronous parallel stochastic zeroth-order algorithm (AsySZO) was proposed.

187, TITLE: Learning to search with MCTSnets

<http://proceedings.mlr.press/v80/guez18a.html>

AUTHORS: Arthur Guez, Theophane Weber, Ioannis Antonoglou, Karen Simonyan, Oriol Vinyals, Daan Wierstra, Remi Munos, David Silver

HIGHLIGHT: In this paper we instead learn where, what and how to search.

188, TITLE: Characterizing Implicit Bias in Terms of Optimization Geometry

<http://proceedings.mlr.press/v80/gunasekar18a.html>

AUTHORS: Suriya Gunasekar, Jason Lee, Daniel Soudry, Nathan Srebro  
HIGHLIGHT: We study the bias of generic optimization methods, including Mirror Descent, Natural Gradient Descent and Steepest Descent with respect to different potentials and norms, when optimizing underdetermined linear models or separable linear classification problems.

189, TITLE: Shampoo: Preconditioned Stochastic Tensor Optimization

<http://proceedings.mlr.press/v80/gupta18a.html>

AUTHORS: Vineet Gupta, Tomer Koren, Yoram Singer  
HIGHLIGHT: We describe and analyze a new structure-aware preconditioning algorithm, called Shampoo, for stochastic optimization over tensor spaces.

190, TITLE: Latent Space Policies for Hierarchical Reinforcement Learning

<http://proceedings.mlr.press/v80/haarnoja18a.html>

AUTHORS: Tuomas Haarnoja, Kristian Hartikainen, Pieter Abbeel, Sergey Levine  
HIGHLIGHT: We address the problem of learning hierarchical deep neural network policies for reinforcement learning.

191, TITLE: Soft Actor-Critic: Off-Policy Maximum Entropy Deep Reinforcement Learning with a Stochastic Actor

<http://proceedings.mlr.press/v80/haarnoja18b.html>

AUTHORS: Tuomas Haarnoja, Aurick Zhou, Pieter Abbeel, Sergey Levine  
HIGHLIGHT: In this paper, we propose soft actor-critic, an off-policy actor-critic deep RL algorithm based on the maximum entropy reinforcement learning framework.

192, TITLE: Comparison-Based Random Forests

<http://proceedings.mlr.press/v80/haghiri18a.html>

AUTHORS: Siavash Haghir, Damien Garreau, Ulrike Luxburg  
HIGHLIGHT: In this paper, we propose a novel random forest algorithm for regression and classification that relies only on such triplet comparisons.

193, TITLE: K-Beam Minimax: Efficient Optimization for Deep Adversarial Learning

<http://proceedings.mlr.press/v80/hamm18a.html>

AUTHORS: Jihun Hamm, Yung-Kyun Noh  
HIGHLIGHT: In this paper, we demonstrate the failure of alternating gradient descent in minimax optimization problems due to the discontinuity of solutions of the inner maximization.

194, TITLE: Candidates vs. Noises Estimation for Large Multi-Class Classification Problem

<http://proceedings.mlr.press/v80/han18a.html>

AUTHORS: Lei Han, Yiheng Huang, Tong Zhang  
HIGHLIGHT: This paper proposes a method for multi-class classification problems, where the number of classes  $K$  is large.

195, TITLE: Stein Variational Gradient Descent Without Gradient

<http://proceedings.mlr.press/v80/han18b.html>

AUTHORS: Jun Han, Qiang Liu  
HIGHLIGHT: In this work, we develop a gradient-free variant of SVGD (GF-SVDG), which replaces the true gradient with a surrogate gradient, and corrects the introduced bias by re-weighting the gradients in a proper form.

196, TITLE: Deep Models of Interactions Across Sets

<http://proceedings.mlr.press/v80/hartford18a.html>

AUTHORS: Jason Hartford, Devon Graham, Kevin Leyton-Brown, Siamak Ravanbakhsh  
HIGHLIGHT: We present a parameter-sharing scheme and prove that it is maximally expressive under the PE constraint.

197, TITLE: Learning Memory Access Patterns  
<http://proceedings.mlr.press/v80/hashemi18a.html>  
AUTHORS: Milad Hashemi, Kevin Swersky, Jamie Smith, Grant Ayers, Heiner Litz, Jichuan Chang, Christos Kozyrakis, Parthasarathy Ranganathan  
HIGHLIGHT: In this paper, we demonstrate the potential of deep learning to address the von Neumann bottleneck of memory performance.

198, TITLE: Fairness Without Demographics in Repeated Loss Minimization  
<http://proceedings.mlr.press/v80/hashimoto18a.html>  
AUTHORS: Tatsunori Hashimoto, Megha Srivastava, Hongseok Namkoong, Percy Liang  
HIGHLIGHT: To mitigate this, we develop an approach based on distributionally robust optimization (DRO), which minimizes the worst case risk over all distributions close to the empirical distribution.

199, TITLE: Multicalibration: Calibration for the (Computationally-Identifiable) Masses  
<http://proceedings.mlr.press/v80/hebert-johnson18a.html>  
AUTHORS: Ursula Hebert-Johnson, Michael Kim, Omer Reingold, Guy Rothblum  
HIGHLIGHT: Along the way, we present algorithms for learning a multicalibrated predictor, study the computational complexity of this task, and illustrate tight connections to the agnostic learning model.

200, TITLE: Recurrent Predictive State Policy Networks  
<http://proceedings.mlr.press/v80/hefny18a.html>  
AUTHORS: Ahmed Hefny, Zita Marinho, Wen Sun, Siddhartha Srinivasa, Geoffrey Gordon  
HIGHLIGHT: We introduce Recurrent Predictive State Policy (RPSP) networks, a recurrent architecture that brings insights from predictive state representations to reinforcement learning in partially observable environments.

201, TITLE: Learning unknown ODE models with Gaussian processes  
<http://proceedings.mlr.press/v80/heinonen18a.html>  
AUTHORS: Markus Heinonen, Cagatay Yildiz, Henrik Mannerström, Jukka Intosalmi, Harri Lohdesmaki  
HIGHLIGHT: We propose to learn non-linear, unknown differential functions from state observations using Gaussian process vector fields within the exact ODE formalism.

202, TITLE: Orthogonal Recurrent Neural Networks with Scaled Cayley Transform  
<http://proceedings.mlr.press/v80/helfrich18a.html>  
AUTHORS: Kyle Helfrich, Devin Willmott, Qiang Ye  
HIGHLIGHT: We propose a simpler and novel update scheme to maintain orthogonal recurrent weight matrices without using complex valued matrices.

203, TITLE: Fast Bellman Updates for Robust MDPs  
<http://proceedings.mlr.press/v80/ho18a.html>  
AUTHORS: Chin Pang Ho, Marek Petrik, Wolfram Wiesemann  
HIGHLIGHT: We describe two efficient, and exact, algorithms for computing Bellman updates in robust Markov decision processes (MDPs).

204, TITLE: CyCADA: Cycle-Consistent Adversarial Domain Adaptation  
<http://proceedings.mlr.press/v80/hoffman18a.html>  
AUTHORS: Judy Hoffman, Eric Tzeng, Taesung Park, Jun-Yan Zhu, Phillip Isola, Kate Saenko, Alexei Efros, Trevor Darrell  
HIGHLIGHT: We propose a model which adapts between domains using both generative image space alignment and latent representation space alignment.

205, TITLE: Sound Abstraction and Decomposition of Probabilistic Programs  
<http://proceedings.mlr.press/v80/holtzen18a.html>



AUTHORS: Steven Holtzen, Guy Broeck, Todd Millstein  
HIGHLIGHT: We propose a form of sound abstraction for probabilistic programs wherein the abstractions are themselves simplified programs.

206, TITLE: Gradient Primal-Dual Algorithm Converges to Second-Order Stationary Solution for Nonconvex Distributed Optimization Over Networks

<http://proceedings.mlr.press/v80/hong18a.html>

AUTHORS: Mingyi Hong, Meisam Razaviyayn, Jason Lee

HIGHLIGHT: In this work, we study two first-order primal-dual based algorithms, the Gradient Primal-Dual Algorithm (GPDA) and the Gradient Alternating Direction Method of Multipliers (GADMM), for solving a class of linearly constrained non-convex optimization problems.

207, TITLE: Variational Bayesian dropout: pitfalls and fixes

<http://proceedings.mlr.press/v80/hron18a.html>

AUTHORS: Jiri Hron, Alex Matthews, Zoubin Ghahramani

HIGHLIGHT: The main contribution of the reinterpretation is in providing a theoretical framework useful for analysing and extending the algorithm.

208, TITLE: Does Distributionally Robust Supervised Learning Give Robust Classifiers?

<http://proceedings.mlr.press/v80/hu18a.html>

AUTHORS: Weihua Hu, Gang Niu, Issei Sato, Masashi Sugiyama

HIGHLIGHT: In this paper, we analyze this DRSL, focusing on the classification scenario.

209, TITLE: Dissipativity Theory for Accelerating Stochastic Variance Reduction: A Unified Analysis of SVRG and Katyusha Using Semidefinite Programs

<http://proceedings.mlr.press/v80/hu18b.html>

AUTHORS: Bin Hu, Stephen Wright, Laurent Lessard

HIGHLIGHT: By leveraging dissipativity theory from control, we provide a new perspective on two important variance-reduction algorithms: SVRG and its direct accelerated variant Katyusha.

210, TITLE: Near Optimal Frequent Directions for Sketching Dense and Sparse Matrices

<http://proceedings.mlr.press/v80/huang18a.html>

AUTHORS: Zengfeng Huang

HIGHLIGHT: Given a large matrix  $A \in \mathbb{R}^{n \times d}$ , we consider the problem of computing a sketch matrix  $B \in \mathbb{R}^{\ell \times d}$  which is significantly smaller than but still well approximates  $A$ .

211, TITLE: Learning Deep ResNet Blocks Sequentially using Boosting Theory

<http://proceedings.mlr.press/v80/huang18b.html>

AUTHORS: Furong Huang, Jordan Ash, John Langford, Robert Schapire

HIGHLIGHT: In other words, we propose a weak learning condition and prove a boosting theory for ResNet under the weak learning condition.

212, TITLE: Learning Hidden Markov Models from Pairwise Co-occurrences with Application to Topic Modeling

<http://proceedings.mlr.press/v80/huang18c.html>

AUTHORS: Kejun Huang, Xiao Fu, Nicholas Sidiropoulos

HIGHLIGHT: We present a new algorithm for identifying the transition and emission probabilities of a hidden Markov model (HMM) from the emitted data.

213, TITLE: Neural Autoregressive Flows

<http://proceedings.mlr.press/v80/huang18d.html>

AUTHORS: Chin-Wei Huang, David Krueger, Alexandre Lacoste, Aaron Courville

**HIGHLIGHT:** We demonstrate that the proposed neural autoregressive flows (NAF) are universal approximators for continuous probability distributions, and their greater expressivity allows them to better capture multimodal target distributions.

214, TITLE: Topological mixture estimation  
<http://proceedings.mlr.press/v80/huntsman18a.html>

AUTHORS: Steve Huntsman

**HIGHLIGHT:** We introduce topological mixture estimation, a completely nonparametric and computationally efficient solution to the problem of estimating a one-dimensional mixture with generic unimodal components.

215, TITLE: Decoupled Parallel Backpropagation with Convergence Guarantee  
<http://proceedings.mlr.press/v80/huo18a.html>

AUTHORS: Zhouyuan Huo, Bin Gu, Yang, Heng Huang

**HIGHLIGHT:** In this paper, we propose decoupled parallel backpropagation algorithm for deep learning optimization with convergence guarantee.

216, TITLE: Using Reward Machines for High-Level Task Specification and Decomposition in Reinforcement Learning

<http://proceedings.mlr.press/v80/icarte18a.html>

AUTHORS: Rodrigo Toro Icarte, Toryn Klassen, Richard Valenzano, Sheila McIlraith

**HIGHLIGHT:** In this paper we propose Reward Machines {—} a type of finite state machine that supports the specification of reward functions while exposing reward function structure to the learner and supporting decomposition.

217, TITLE: Deep Variational Reinforcement Learning for POMDPs

<http://proceedings.mlr.press/v80/igl18a.html>

AUTHORS: Maximilian Igl, Luisa Zintgraf, Tuan Anh Le, Frank Wood, Shimon Whiteson

**HIGHLIGHT:** In this paper, we propose deep variational reinforcement learning (DVRL), which introduces an inductive bias that allows an agent to learn a generative model of the environment and perform inference in that model to effectively aggregate the available information.

218, TITLE: Attention-based Deep Multiple Instance Learning

<http://proceedings.mlr.press/v80/ilse18a.html>

AUTHORS: Maximilian Ilse, Jakub Tomczak, Max Welling

**HIGHLIGHT:** In this paper, we state the MIL problem as learning the Bernoulli distribution of the bag label where the bag label probability is fully parameterized by neural networks.

219, TITLE: Black-box Adversarial Attacks with Limited Queries and Information

<http://proceedings.mlr.press/v80/ilyas18a.html>

AUTHORS: Andrew Ilyas, Logan Engstrom, Anish Athalye, Jessy Lin

**HIGHLIGHT:** We define three realistic threat models that more accurately characterize many real-world classifiers: the query-limited setting, the partial-information setting, and the label-only setting.

220, TITLE: Analysis of Minimax Error Rate for Crowdsourcing and Its Application to Worker Clustering Model

<http://proceedings.mlr.press/v80/imamura18a.html>

AUTHORS: Hideaki Imamura, Issei Sato, Masashi Sugiyama

**HIGHLIGHT:** In this paper, we derive a minimax error rate under more practical setting for a broader class of crowdsourcing models including the DS model as a special case.

221, TITLE: Improving Regression Performance with Distributional Losses

<http://proceedings.mlr.press/v80/imani18a.html>

AUTHORS: Ehsan Imani, Martha White

**HIGHLIGHT:** In this work, we investigate the reasons for this improvement, in a regression setting.

222, TITLE: Deep Density Destructors  
<http://proceedings.mlr.press/v80/inouye18a.html>  
AUTHORS: David Inouye, Pradeep Ravikumar  
HIGHLIGHT: We propose a unified framework for deep density models by formally defining density destructors.

223, TITLE: Unbiased Objective Estimation in Predictive Optimization  
<http://proceedings.mlr.press/v80/ito18a.html>  
AUTHORS: Shinji Ito, Akihiro Yabe, Ryohei Fujimaki  
HIGHLIGHT: This paper investigates such optimistic bias and presents two methods for correcting it.

224, TITLE: Anonymous Walk Embeddings  
<http://proceedings.mlr.press/v80/ivanov18a.html>  
AUTHORS: Sergey Ivanov, Evgeny Burnaev  
HIGHLIGHT: Here, we coherently propose an approach for embedding entire graphs and show that our feature representations with SVM classifier increase classification accuracy of CNN algorithms and traditional graph kernels.

225, TITLE: Learning Binary Latent Variable Models: A Tensor Eigenpair Approach  
<http://proceedings.mlr.press/v80/jaffe18a.html>  
AUTHORS: Ariel Jaffe, Roi Weiss, Boaz Nadler, Shai Carmi, Yuval Kluger  
HIGHLIGHT: In this paper we propose a novel spectral approach to this problem, based on the eigenvectors of both the second order moment matrix and third order moment tensor of the observed data.

226, TITLE: Firing Bandits: Optimizing Crowdfunding  
<http://proceedings.mlr.press/v80/jain18a.html>  
AUTHORS: Lalit Jain, Kevin Jamieson  
HIGHLIGHT: In this paper, we model the problem of optimizing crowdfunding platforms, such as the non-profit Kiva or for-profit Kickstarter, as a variant of the multi-armed bandit problem.

227, TITLE: Differentially Private Matrix Completion Revisited  
<http://proceedings.mlr.press/v80/jain18b.html>  
AUTHORS: Prateek Jain, Om Dipakbhai Thakkar, Abhradeep Thakurta  
HIGHLIGHT: We provide the first provably joint differentially private algorithm with formal utility guarantees for the problem of user-level privacy-preserving collaborative filtering.

228, TITLE: Video Prediction with Appearance and Motion Conditions  
<http://proceedings.mlr.press/v80/jang18a.html>  
AUTHORS: Yunseok Jang, Gunhee Kim, Yale Song  
HIGHLIGHT: We propose an Appearance-Motion Conditional GAN to address this challenge.

229, TITLE: Pathwise Derivatives Beyond the Reparameterization Trick  
<http://proceedings.mlr.press/v80/jankowiak18a.html>  
AUTHORS: Martin Jankowiak, Fritz Obermeyer  
HIGHLIGHT: We observe that gradients computed via the reparameterization trick are in direct correspondence with solutions of the transport equation in the formalism of optimal transport.

230, TITLE: Detecting non-causal artifacts in multivariate linear regression models  
<http://proceedings.mlr.press/v80/janzing18a.html>  
AUTHORS: Dominik Janzing, Bernhard Schölkopf  
HIGHLIGHT: We consider linear models where  $d$  potential causes  $X_1, \dots, X_d$  are correlated with one target quantity  $Y$  and propose a method to infer whether the association is causal or whether it is an artifact caused by overfitting or hidden common causes.

231, TITLE: A Unified Framework for Structured Low-rank Matrix Learning  
<http://proceedings.mlr.press/v80/jawanpuria18a.html>  
AUTHORS: Pratik Jawanpuria, Bamdev Mishra  
HIGHLIGHT: We consider the problem of learning a low-rank matrix, constrained to lie in a linear subspace, and introduce a novel factorization for modeling such matrices.

232, TITLE: Efficient end-to-end learning for quantizable representations  
<http://proceedings.mlr.press/v80/jeong18a.html>  
AUTHORS: Yeonwoo Jeong, Hyun Oh Song  
HIGHLIGHT: To this end, we consider the problem of directly learning a quantizable embedding representation and the sparse binary hash code end-to-end which can be used to construct an efficient hash table not only providing significant search reduction in the number of data but also achieving the state of the art search accuracy outperforming previous state of the art deep metric learning methods.

233, TITLE: Exploring Hidden Dimensions in Accelerating Convolutional Neural Networks  
<http://proceedings.mlr.press/v80/jia18a.html>  
AUTHORS: Zhihao Jia, Sina Lin, Charles R. Qi, Alex Aiken  
HIGHLIGHT: In this paper, we propose layer-wise parallelism that allows each layer in a network to use an individual parallelization strategy.

234, TITLE: Feedback-Based Tree Search for Reinforcement Learning  
<http://proceedings.mlr.press/v80/jiang18a.html>  
AUTHORS: Daniel Jiang, Emmanuel Ekwedike, Han Liu  
HIGHLIGHT: Inspired by recent successes of Monte-Carlo tree search (MCTS) in a number of artificial intelligence (AI) application domains, we propose a reinforcement learning (RL) technique that iteratively applies MCTS on batches of small, finite-horizon versions of the original infinite-horizon Markov decision process.

235, TITLE: Quickshift++: Provably Good Initializations for Sample-Based Mean Shift  
<http://proceedings.mlr.press/v80/jiang18b.html>  
AUTHORS: Heinrich Jiang, Jennifer Jang, Samory Kpotufe  
HIGHLIGHT: We provide initial seedings to the Quick Shift clustering algorithm, which approximate the locally high-density regions of the data.

236, TITLE: MentorNet: Learning Data-Driven Curriculum for Very Deep Neural Networks on Corrupted Labels  
<http://proceedings.mlr.press/v80/jiang18c.html>  
AUTHORS: Lu Jiang, Zhengyuan Zhou, Thomas Leung, Li-Jia Li, Li Fei-Fei  
HIGHLIGHT: To overcome the overfitting on corrupted labels, we propose a novel technique of learning another neural network, called MentorNet, to supervise the training of the base deep networks, namely, StudentNet.

237, TITLE: The Weighted Kendall and High-order Kernels for Permutations  
<http://proceedings.mlr.press/v80/jiao18a.html>  
AUTHORS: Yunlong Jiao, Jean-Philippe Vert  
HIGHLIGHT: We propose new positive definite kernels for permutations.

238, TITLE: Junction Tree Variational Autoencoder for Molecular Graph Generation  
<http://proceedings.mlr.press/v80/jin18a.html>  
AUTHORS: Wengong Jin, Regina Barzilay, Tommi Jaakkola  
HIGHLIGHT: Our primary contribution is the direct realization of molecular graphs, a task previously approached by generating linear SMILES strings instead of graphs.

239, TITLE: Network Global Testing by Counting Graphlets  
<http://proceedings.mlr.press/v80/jin18b.html>

AUTHORS: Jiashun Jin, Zheng Ke, Shengming Luo  
HIGHLIGHT: We construct a class of test statistics using the numbers of short paths and short cycles, and the key to our approach is a general framework for canceling the effects of degree heterogeneity.

240, TITLE: Regret Minimization for Partially Observable Deep Reinforcement Learning  
<http://proceedings.mlr.press/v80/jin18c.html>

AUTHORS: Peter Jin, Kurt Keutzer, Sergey Levine  
HIGHLIGHT: In this work, we propose a new deep reinforcement learning algorithm based on counterfactual regret minimization that iteratively updates an approximation to an advantage-like function and is robust to partially observed state.

241, TITLE: WSNet: Compact and Efficient Networks Through Weight Sampling  
<http://proceedings.mlr.press/v80/jin18d.html>

AUTHORS: Xiaojie Jin, Yingzhen Yang, Ning Xu, Jianchao Yang, Nebojsa Jojic, Jiashi Feng, Shuicheng Yan  
HIGHLIGHT: We present a new approach and a novel architecture, termed WSNet, for learning compact and efficient deep neural networks.

242, TITLE: Large-Scale Cox Process Inference using Variational Fourier Features  
<http://proceedings.mlr.press/v80/john18a.html>

AUTHORS: ST John, James Hensman  
HIGHLIGHT: Here we introduce Cox process inference based on Fourier features.

243, TITLE: Composite Functional Gradient Learning of Generative Adversarial Models  
<http://proceedings.mlr.press/v80/johnson18a.html>

AUTHORS: Rie Johnson, Tong Zhang  
HIGHLIGHT: Based on the theory, we propose a new stable generative adversarial method.

244, TITLE: Kronecker Recurrent Units  
<http://proceedings.mlr.press/v80/jose18a.html>

AUTHORS: Cijo Jose, Moustapha Cisse, Francois Fleuret  
HIGHLIGHT: We present a flexible recurrent neural network model called Kronecker Recurrent Units (KRU).

245, TITLE: Fast Decoding in Sequence Models Using Discrete Latent Variables  
<http://proceedings.mlr.press/v80/kaiser18a.html>

AUTHORS: Lukasz Kaiser, Samy Bengio, Aurko Roy, Ashish Vaswani, Niki Parmar, Jakob Uszkoreit, Noam Shazeer  
HIGHLIGHT: We present a method to extend sequence models using discrete latent variables that makes decoding much more parallel.

246, TITLE: Kernel Recursive ABC: Point Estimation with Intractable Likelihood  
<http://proceedings.mlr.press/v80/kajihara18a.html>

AUTHORS: Takafumi Kajihara, Motonobu Kanagawa, Keisuke Yamazaki, Kenji Fukumizu  
HIGHLIGHT: We propose a novel approach to parameter estimation for simulator-based statistical models with intractable likelihood.

247, TITLE: Efficient Neural Audio Synthesis  
<http://proceedings.mlr.press/v80/kalchbrenner18a.html>

AUTHORS: Nal Kalchbrenner, Erich Elsen, Karen Simonyan, Seb Noury, Norman Casagrande, Edward Lockhart, Florian Stimberg, Aaron Oord, Sander Dieleman, Koray Kavukcuoglu  
HIGHLIGHT: With a focus on text-to-speech synthesis, we describe a set of general techniques for reducing sampling time while maintaining high output quality.

248, TITLE: Learning Diffusion using Hyperparameters

<http://proceedings.mlr.press/v80/kalimeris18a.html>

AUTHORS: Dimitris Kalimeris, Yaron Singer, Karthik Subbian, Udi Weinsberg  
HIGHLIGHT: In this paper we advocate for a hyperparametric approach to learn diffusion in the independent cascade (IC) model.

249, TITLE: Signal and Noise Statistics Oblivious Orthogonal Matching Pursuit

<http://proceedings.mlr.press/v80/kallummil18a.html>

AUTHORS: Sreejith Kallummil, Sheetal Kalyani  
HIGHLIGHT: In this paper, we present a novel technique called residual ratio thresholding (RRT) to operate OMP without any a priori knowledge of sparsity and noise statistics and establish finite sample and large sample support recovery guarantees for the same.

250, TITLE: Residual Unfairness in Fair Machine Learning from Prejudiced Data

<http://proceedings.mlr.press/v80/kallus18a.html>

AUTHORS: Nathan Kallus, Angela Zhou  
HIGHLIGHT: When certain benchmark data is available, we show how sample reweighting can estimate and adjust fairness metrics while accounting for censoring.

251, TITLE: Learn from Your Neighbor: Learning Multi-modal Mappings from Sparse Annotations

<http://proceedings.mlr.press/v80/kalyan18a.html>

AUTHORS: Ashwin Kalyan, Stefan Lee, Anitha Kannan, Dhruv Batra  
HIGHLIGHT: Specifically, we propose an objective that transfers supervision from neighboring examples.

252, TITLE: Semi-Supervised Learning via Compact Latent Space Clustering

<http://proceedings.mlr.press/v80/kamnitsas18a.html>

AUTHORS: Konstantinos Kamnitsas, Daniel Castro, Loic Le Folgoc, Ian Walker, Ryutaro Tanno, Daniel Rueckert, Ben Glocker, Antonio Criminisi, Aditya Nori  
HIGHLIGHT: We present a novel cost function for semi-supervised learning of neural networks that encourages compact clustering of the latent space to facilitate separation.

253, TITLE: Policy Optimization with Demonstrations

<http://proceedings.mlr.press/v80/kang18a.html>

AUTHORS: Bingyi Kang, Zequn Jie, Jiashi Feng  
HIGHLIGHT: We propose to effectively leverage available demonstrations to guide exploration through enforcing occupancy measure matching between the learned policy and current demonstrations, and develop a novel Policy Optimization from Demonstration (POfD) method.

254, TITLE: Improving Sign Random Projections With Additional Information

<http://proceedings.mlr.press/v80/kang18b.html>

AUTHORS: Keegan Kang, Weipin Wong  
HIGHLIGHT: We propose using additional information to improve these estimates which is easy to implement and cost efficient.

255, TITLE: Let's be Honest: An Optimal No-Regret Framework for Zero-Sum Games

<http://proceedings.mlr.press/v80/kangarshahi18a.html>

AUTHORS: Ehsan Asadi Kangarshahi, Ya-Ping Hsieh, Mehmet Fatih Sahin, Volkan Cevher  
HIGHLIGHT: We propose a simple algorithmic framework that simultaneously achieves the best rates for honest regret as well as adversarial regret, and in addition resolves the open problem of removing the logarithmic terms in convergence to the value of the game.

256, TITLE: Continual Reinforcement Learning with Complex Synapses

<http://proceedings.mlr.press/v80/kaplanis18a.html>

AUTHORS: Christos Kaplanis, Murray Shanahan, Claudia Clopath

**HIGHLIGHT:** In this paper, we show that by equipping tabular and deep reinforcement learning agents with a synaptic model that incorporates this biological complexity (Benna & Fusi, 2016), catastrophic forgetting can be mitigated at multiple timescales.

257, **TITLE:** LaVAN: Localized and Visible Adversarial Noise

<http://proceedings.mlr.press/v80/karmon18a.html>

**AUTHORS:** Danny Karmon, Daniel Zoran, Yoav Goldberg

**HIGHLIGHT:** We explore the case where the noise is allowed to be visible but confined to a small, localized patch of the image, without covering any of the main object(s) in the image.

258, **TITLE:** Riemannian Stochastic Recursive Gradient Algorithm

<http://proceedings.mlr.press/v80/kasai18a.html>

**AUTHORS:** Hiroyuki Kasai, Hiroyuki Sato, Bamdev Mishra

**HIGHLIGHT:** The present paper proposes a Riemannian stochastic recursive gradient algorithm (R-SRG), which does not require the inverse of retraction between two distant iterates on the manifold.

259, **TITLE:** Not All Samples Are Created Equal: Deep Learning with Importance Sampling

<http://proceedings.mlr.press/v80/katharopoulos18a.html>

**AUTHORS:** Angelos Katharopoulos, Francois Fleuret

**HIGHLIGHT:** We propose to mitigate this phenomenon with a principled importance sampling scheme that focuses computation on "informative" examples, and reduces the variance of the stochastic gradients during training.

260, **TITLE:** Feasible Arm Identification

<http://proceedings.mlr.press/v80/katz-samuels18a.html>

**AUTHORS:** Julian Katz-Samuels, Clay Scott

**HIGHLIGHT:** We introduce the feasible arm identification problem, a pure exploration multi-armed bandit problem where the agent is given a set of  $D$ -dimensional arms and a polyhedron  $P = \{x : Ax \leq b\} \subset \mathbb{R}^D$ .

261, **TITLE:** Scalable Deletion-Robust Submodular Maximization: Data Summarization with Privacy and Fairness Constraints

<http://proceedings.mlr.press/v80/kazemi18a.html>

**AUTHORS:** Ehsan Kazemi, Morteza Zadimoghaddam, Amin Karbasi

**HIGHLIGHT:** We propose the first memory-efficient centralized, streaming, and distributed methods with constant-factor approximation guarantees against any number of adversarial deletions.

262, **TITLE:** Focused Hierarchical RNNs for Conditional Sequence Processing

<http://proceedings.mlr.press/v80/ke18a.html>

**AUTHORS:** Nan Rosemary Ke, Konrad Zolna, Alessandro Sordoni, Zhouhan Lin, Adam Trischler, Yoshua Bengio, Joelle Pineau, Laurent Charlin, Christopher Pal

**HIGHLIGHT:** We present a mechanism for focusing RNN encoders for sequence modelling tasks which allows them to attend to key parts of the input as needed.

263, **TITLE:** Preventing Fairness Gerrymandering: Auditing and Learning for Subgroup Fairness

<http://proceedings.mlr.press/v80/kearns18a.html>

**AUTHORS:** Michael Kearns, Seth Neel, Aaron Roth, Zhiwei Steven Wu

**HIGHLIGHT:** We thus consider fairness across exponentially or infinitely many subgroups, defined by a structured class of functions over the protected attributes.

264, **TITLE:** Improved nearest neighbor search using auxiliary information and priority functions

<http://proceedings.mlr.press/v80/keivani18a.html>

**AUTHORS:** Omid Keivani, Kaushik Sinha

**HIGHLIGHT:** To address this issue, in this paper we present different search strategies to improve nearest neighbor search performance of a single random projection tree.

- 265, TITLE: ContextNet: Deep learning for Star Galaxy Classification  
http://proceedings.mlr.press/v80/kennamer18a.html  
AUTHORS: Noble Kennamer, David Kirkby, Alexander Ihler, Francisco Javier Sanchez-Lopez  
HIGHLIGHT: We present a framework to compose artificial neural networks in cases where the data cannot be treated as independent events.
- 266, TITLE: Frank-Wolfe with Subsampling Oracle  
http://proceedings.mlr.press/v80/kerdreux18a.html  
AUTHORS: Thomas Kerdreux, Fabian Pedregosa, Alexandre d'Aspremont  
HIGHLIGHT: The first algorithm that we propose is a randomized variant of the original FW algorithm and achieves a  $\mathcal{O}(1/t)$  sublinear convergence rate as in the deterministic counterpart.
- 267, TITLE: Convergence guarantees for a class of non-convex and non-smooth optimization problems  
http://proceedings.mlr.press/v80/khamaru18a.html  
AUTHORS: Koulik Khamaru, Martin Wainwright  
HIGHLIGHT: We consider the problem of finding critical points of a broad class of non-convex problems with non-smooth components.
- 268, TITLE: Fast and Scalable Bayesian Deep Learning by Weight-Perturbation in Adam  
http://proceedings.mlr.press/v80/khan18a.html  
AUTHORS: Mohammad Khan, Didrik Nielsen, Voot Tangkaratt, Wu Lin, Yarin Gal, Akash Srivastava  
HIGHLIGHT: In this paper, we propose new natural-gradient algorithms to reduce such efforts for Gaussian mean-field VI.
- 269, TITLE: Geometry Score: A Method For Comparing Generative Adversarial Networks  
http://proceedings.mlr.press/v80/khrulkov18a.html  
AUTHORS: Valentin Khrulkov, Ivan Oseledets  
HIGHLIGHT: In this work, we construct a novel measure of performance of a GAN by comparing geometrical properties of the underlying data manifold and the generated one, which provides both qualitative and quantitative means for evaluation.
- 270, TITLE: Blind Justice: Fairness with Encrypted Sensitive Attributes  
http://proceedings.mlr.press/v80/kilbertus18a.html  
AUTHORS: Niki Kilbertus, Adria Gascon, Matt Kusner, Michael Veale, Krishna Gummadi, Adrian Weller  
HIGHLIGHT: We introduce methods from secure multi-party computation which allow us to avoid both.
- 271, TITLE: Markov Modulated Gaussian Cox Processes for Semi-Stationary Intensity Modeling of Events Data  
http://proceedings.mlr.press/v80/kim18a.html  
AUTHORS: Minyoung Kim  
HIGHLIGHT: In this paper we bring up two most popular Cox models representing two extremes, and propose a novel semi-stationary Cox process model that can take benefits from both models.
- 272, TITLE: Disentangling by Factorising  
http://proceedings.mlr.press/v80/kim18b.html  
AUTHORS: Hyunjik Kim, Andriy Mnih  
HIGHLIGHT: We propose FactorVAE, a method that disentangles by encouraging the distribution of representations to be factorial and hence independent across the dimensions.
- 273, TITLE: Self-Bounded Prediction Suffix Tree via Approximate String Matching  
http://proceedings.mlr.press/v80/kim18c.html  
AUTHORS: Dongwoo Kim, Christian Walder



**HIGHLIGHT:** We present a provably correct algorithm for learning a PST with approximate suffix matching by relaxing the exact matching condition.

274, TITLE: Interpretability Beyond Feature Attribution: Quantitative Testing with Concept Activation Vectors (TCAV)

<http://proceedings.mlr.press/v80/kim18d.html>

AUTHORS: Been Kim, Martin Wattenberg, Justin Gilmer, Carrie Cai, James Wexler, Fernanda Viegas, Rory sayres

**HIGHLIGHT:** To address these challenges, we introduce Concept Activation Vectors (CAVs), which provide an interpretation of a neural net's internal state in terms of human-friendly concepts.

275, TITLE: Semi-Amortized Variational Autoencoders

<http://proceedings.mlr.press/v80/kim18e.html>

AUTHORS: Yoon Kim, Sam Wiseman, Andrew Miller, David Sontag, Alexander Rush

**HIGHLIGHT:** We propose a hybrid approach, to use AVI to initialize the variational parameters and run stochastic variational inference (SVI) to refine them.

276, TITLE: Neural Relational Inference for Interacting Systems

<http://proceedings.mlr.press/v80/kipf18a.html>

AUTHORS: Thomas Kipf, Ethan Fetaya, Kuan-Chieh Wang, Max Welling, Richard Zemel

**HIGHLIGHT:** In this work, we introduce the neural relational inference (NRI) model: an unsupervised model that learns to infer interactions while simultaneously learning the dynamics purely from observational data.

277, TITLE: An Alternative View: When Does SGD Escape Local Minima?

<http://proceedings.mlr.press/v80/kleinberg18a.html>

AUTHORS: Bobby Kleinberg, Yuanzhi Li, Yang Yuan

**HIGHLIGHT:** In order to understand this phenomenon, we take an alternative view that SGD is working on the convolved (thus smoothed) version of the loss function.

278, TITLE: Crowdsourcing with Arbitrary Adversaries

<http://proceedings.mlr.press/v80/kleindessner18a.html>

AUTHORS: Matthaues Kleindessner, Pranjal Awasthi

**HIGHLIGHT:** In this adversarial scenario, we design an efficient algorithm to consistently estimate the workers' error probabilities.

279, TITLE: Spatio-temporal Bayesian On-line Changepoint Detection with Model Selection

<http://proceedings.mlr.press/v80/knoblach18a.html>

AUTHORS: Jeremias Knoblauch, Theodoros Damoulas

**HIGHLIGHT:** We propose spatially structured Vector Autoregressions (VARs) for modelling the process between changepoints (CPs) and give an upper bound on the approximation error of such models.

280, TITLE: Fast Gradient-Based Methods with Exponential Rate: A Hybrid Control Framework

<http://proceedings.mlr.press/v80/kolarijani18a.html>

AUTHORS: Arman Sharifi Kolarijani, Peyman Mohajerin Esfahani, Tamas Keviczky

**HIGHLIGHT:** In this study, we pursue a similar objective and propose a class of hybrid control systems that adopts a 2nd-order differential equation as its continuous flow.

281, TITLE: Nonconvex Optimization for Regression with Fairness Constraints

<http://proceedings.mlr.press/v80/komiyama18a.html>

AUTHORS: Junpei Komiyama, Akiko Takeda, Junya Honda, Hajime Shimao

**HIGHLIGHT:** Despite such nonconvexity, we show an exact solution is available by using tools of global optimization theory.

- 282, TITLE: On the Generalization of Equivariance and Convolution in Neural Networks to the Action of Compact Groups  
<http://proceedings.mlr.press/v80/kondor18a.html>  
AUTHORS: Risi Kondor, Shubhendu Trivedi  
HIGHLIGHT: In this paper we give a rigorous, theoretical treatment of convolution and equivariance in neural networks with respect to not just translations, but the action of any compact group.
- 283, TITLE: Compiling Combinatorial Prediction Games  
<http://proceedings.mlr.press/v80/koriche18a.html>  
AUTHORS: Frederic Koriche  
HIGHLIGHT: In this paper, we consider the paradigm of compiling the set of combinatorial constraints into a deterministic and Decomposable Negation Normal Form (dDNNF) circuit, for which the tasks of linear optimization and solution sampling take linear time.
- 284, TITLE: Dynamic Evaluation of Neural Sequence Models  
<http://proceedings.mlr.press/v80/krause18a.html>  
AUTHORS: Ben Krause, Emmanuel Kahembwe, Iain Murray, Steve Renals  
HIGHLIGHT: We develop a dynamic evaluation approach that outperforms existing adaptation approaches in our comparisons.
- 285, TITLE: Semiparametric Contextual Bandits  
<http://proceedings.mlr.press/v80/krishnamurthy18a.html>  
AUTHORS: Akshay Krishnamurthy, Zhiwei Steven Wu, Vasilis Syrgkanis  
HIGHLIGHT: We design new algorithms that achieve  $\tilde{O}(\sqrt{T})$  regret over  $T$  rounds, when the linear function is  $d$ -dimensional, which matches the best known bounds for the simpler unconfounded case and improves on a recent result of Greenwald et al. (2017).
- 286, TITLE: Fast Maximization of Non-Submodular, Monotonic Functions on the Integer Lattice  
<http://proceedings.mlr.press/v80/kuhnle18a.html>  
AUTHORS: Alan Kuhnle, J. David Smith, Victoria Crawford, My Thai  
HIGHLIGHT: We propose a general framework for influence maximization on the integer lattice that generalizes prior works on this topic, and we demonstrate the efficiency of our algorithms in this context.
- 287, TITLE: Accurate Uncertainties for Deep Learning Using Calibrated Regression  
<http://proceedings.mlr.press/v80/kuleshov18a.html>  
AUTHORS: Volodymyr Kuleshov, Nathan Fenner, Stefano Ermon  
HIGHLIGHT: Here, we propose a simple procedure for calibrating any regression algorithm; when applied to Bayesian and probabilistic models, it is guaranteed to produce calibrated uncertainty estimates given enough data.
- 288, TITLE: Trainable Calibration Measures for Neural Networks from Kernel Mean Embeddings  
<http://proceedings.mlr.press/v80/kumar18a.html>  
AUTHORS: Aviral Kumar, Sunita Sarawagi, Ujjwal Jain  
HIGHLIGHT: We propose a more principled fix that minimizes an explicit calibration error during training.
- 289, TITLE: Data-Dependent Stability of Stochastic Gradient Descent  
<http://proceedings.mlr.press/v80/kuzborskij18a.html>  
AUTHORS: Ilja Kuzborskij, Christoph Lampert  
HIGHLIGHT: We establish a data-dependent notion of algorithmic stability for Stochastic Gradient Descent (SGD), and employ it to develop novel generalization bounds.
- 290, TITLE: Explicit Inductive Bias for Transfer Learning with Convolutional Networks  
<http://proceedings.mlr.press/v80/li18a.html>  
AUTHORS: Xuhong LI, Yves Grandvalet, Franck Davoine

**HIGHLIGHT:** In this paper, we investigate several regularization schemes that explicitly promote the similarity of the final solution with the initial model.

291, **TITLE:** Understanding the Loss Surface of Neural Networks for Binary Classification  
<http://proceedings.mlr.press/v80/liang18a.html>

**AUTHORS:** SHIYU LIANG, Ruoyu Sun, Yixuan Li, Rayadurgam Srikant

**HIGHLIGHT:** Here we focus on the training performance of neural networks for binary classification, and provide conditions under which the training error is zero at all local minima of appropriately chosen surrogate loss functions.

292, **TITLE:** Mixed batches and symmetric discriminators for GAN training  
<http://proceedings.mlr.press/v80/lucas18a.html>

**AUTHORS:** Thomas LUCAS, Corentin Tallec, Yann Ollivier, Jakob Verbeek

**HIGHLIGHT:** We propose to feed the discriminator with mixed batches of true and fake samples, and train it to predict the ratio of true samples in the batch.

293, **TITLE:** Binary Partitions with Approximate Minimum Impurity  
<http://proceedings.mlr.press/v80/laber18a.html>

**AUTHORS:** Eduardo Laber, Marco Molinaro, Felipe Mello Pereira

**HIGHLIGHT:** To fill this gap, we propose two new splitting procedures that provably achieve near-optimal impurity.

294, **TITLE:** Canonical Tensor Decomposition for Knowledge Base Completion  
<http://proceedings.mlr.press/v80/lacroix18a.html>

**AUTHORS:** Timothee Lacroix, Nicolas Usunier, Guillaume Obozinski

**HIGHLIGHT:** In this work, we attempt to understand the limits of CP for knowledge base completion.

295, **TITLE:** Generalization without Systematicity: On the Compositional Skills of Sequence-to-Sequence Recurrent Networks

<http://proceedings.mlr.press/v80/lake18a.html>

**AUTHORS:** Brenden Lake, Marco Baroni

**HIGHLIGHT:** In this paper, we introduce the SCAN domain, consisting of a set of simple compositional navigation commands paired with the corresponding action sequences.

296, **TITLE:** An Estimation and Analysis Framework for the Rasch Model  
<http://proceedings.mlr.press/v80/lan18a.html>

**AUTHORS:** Andrew Lan, Mung Chiang, Christoph Studer

**HIGHLIGHT:** This paper provides a framework that relies on a novel linear minimum mean-squared error (L-MMSE) estimator which enables an exact, nonasymptotic, and closed-form analysis of the parameter estimation error under the Rasch model.

297, **TITLE:** Partial Optimality and Fast Lower Bounds for Weighted Correlation Clustering  
<http://proceedings.mlr.press/v80/lange18a.html>

**AUTHORS:** Jan-Hendrik Lange, Andreas Karrenbauer, Bjoern Andres

**HIGHLIGHT:** We introduce a re-weighting with the dual solution by which efficient local search algorithms converge to better feasible solutions.

298, **TITLE:** Deep Linear Networks with Arbitrary Loss: All Local Minima Are Global  
<http://proceedings.mlr.press/v80/laurent18a.html>

**AUTHORS:** Thomas Laurent, James Brecht

**HIGHLIGHT:** We consider deep linear networks with arbitrary convex differentiable loss.

299, **TITLE:** The Multilinear Structure of ReLU Networks  
<http://proceedings.mlr.press/v80/laurent18b.html>

- AUTHORS: Thomas Laurent, James Brecht  
HIGHLIGHT: We study the loss surface of neural networks equipped with a hinge loss criterion and ReLU or leaky ReLU nonlinearities.
- 300, TITLE: Hierarchical Imitation and Reinforcement Learning  
<http://proceedings.mlr.press/v80/le18a.html>  
AUTHORS: Hoang Le, Nan Jiang, Alekh Agarwal, Miroslav Dudik, Yisong Yue, Hal Daum?  
HIGHLIGHT: We propose an algorithmic framework, called hierarchical guidance, that leverages the hierarchical structure of the underlying problem to integrate different modes of expert interaction.
- 301, TITLE: Gradient-Based Meta-Learning with Learned Layerwise Metric and Subspace  
<http://proceedings.mlr.press/v80/lee18a.html>  
AUTHORS: Yoonho Lee, Seungjin Choi  
HIGHLIGHT: Our primary contribution is the MT-net, which enables the meta-learner to learn on each layer's activation space a subspace that the task-specific learner performs gradient descent on.
- 302, TITLE: Deep Reinforcement Learning in Continuous Action Spaces: a Case Study in the Game of Simulated Curling  
<http://proceedings.mlr.press/v80/lee18b.html>  
AUTHORS: Kyowoon Lee, Sol-A Kim, Jaesik Choi, Seong-Wan Lee  
HIGHLIGHT: In this paper, we present a new self-play reinforcement learning framework which equips a continuous search algorithm which enables to search in continuous action spaces with a kernel regression method.
- 303, TITLE: Gated Path Planning Networks  
<http://proceedings.mlr.press/v80/lee18c.html>  
AUTHORS: Lisa Lee, Emilio Parisotto, Devendra Singh Chaplot, Eric Xing, Ruslan Salakhutdinov  
HIGHLIGHT: In this work, we reframe VINs as recurrent-convolutional networks which demonstrates that VINs couple recurrent convolutions with an unconventional max-pooling activation.
- 304, TITLE: Deep Asymmetric Multi-task Feature Learning  
<http://proceedings.mlr.press/v80/lee18d.html>  
AUTHORS: Hae Beom Lee, Eunho Yang, Sung Ju Hwang  
HIGHLIGHT: We propose Deep Asymmetric Multitask Feature Learning (Deep-AMTFL) which can learn deep representations shared across multiple tasks while effectively preventing negative transfer that may happen in the feature sharing process.
- 305, TITLE: Noise2Noise: Learning Image Restoration without Clean Data  
<http://proceedings.mlr.press/v80/lehtinen18a.html>  
AUTHORS: Jaakko Lehtinen, Jacob Munkberg, Jon Hasselgren, Samuli Laine, Tero Karras, Miika Aittala, Timo Aila  
HIGHLIGHT: We apply basic statistical reasoning to signal reconstruction by machine learning - learning to map corrupted observations to clean signals - with a simple and powerful conclusion: it is possible to learn to restore images by only looking at corrupted examples, at performance at and sometimes exceeding training using clean data, without explicit image priors or likelihood models of the corruption.
- 306, TITLE: Out-of-sample extension of graph adjacency spectral embedding  
<http://proceedings.mlr.press/v80/levin18a.html>  
AUTHORS: Keith Levin, Fred Roosta, Michael Mahoney, Carey Priebe  
HIGHLIGHT: In this work, we consider the problem of obtaining an out-of-sample extension for the adjacency spectral embedding, a procedure for embedding the vertices of a graph into Euclidean space.
- 307, TITLE: An Optimal Control Approach to Deep Learning and Applications to Discrete-Weight Neural Networks  
<http://proceedings.mlr.press/v80/li18b.html>

AUTHORS: Qianxiao Li, Shuji Hao  
HIGHLIGHT: In particular, we introduce the discrete-time method of successive approximations (MSA), which is based on the Pontryagin's maximum principle, for training neural networks.

308, TITLE: Towards Binary-Valued Gates for Robust LSTM Training  
<http://proceedings.mlr.press/v80/li18c.html>  
AUTHORS: Zhuohan Li, Di He, Fei Tian, Wei Chen, Tao Qin, Liwei Wang, Tieyan Liu  
HIGHLIGHT: In this paper, we propose a new way for LSTM training, which pushes the output values of the gates towards 0 or 1.

309, TITLE: On the Limitations of First-Order Approximation in GAN Dynamics  
<http://proceedings.mlr.press/v80/li18d.html>  
AUTHORS: Jerry Li, Aleksander Madry, John Peebles, Ludwig Schmidt  
HIGHLIGHT: To address this issue, we study GAN dynamics in a simple yet rich parametric model that exhibits several of the common problematic convergence behaviors such as vanishing gradients, mode collapse, and diverging or oscillatory behavior.

310, TITLE: Submodular Hypergraphs: p-Laplacians, Cheeger Inequalities and Spectral Clustering  
<http://proceedings.mlr.press/v80/li18e.html>  
AUTHORS: Pan Li, Olgica Milenkovic  
HIGHLIGHT: We introduce submodular hypergraphs, a family of hypergraphs that have different submodular weights associated with different cuts of hyperedges.

311, TITLE: The Well-Tempered Lasso  
<http://proceedings.mlr.press/v80/li18f.html>  
AUTHORS: Yuanzhi Li, Yoram Singer  
HIGHLIGHT: Alas, the path complexity of the Lasso on artificially designed regression problems is exponential. We use smoothed analysis as a mechanism for bridging the gap between worst case settings and the de facto low complexity.

312, TITLE: Estimation of Markov Chain via Rank-Constrained Likelihood  
<http://proceedings.mlr.press/v80/li18g.html>  
AUTHORS: Xudong Li, Mengdi Wang, Anru Zhang  
HIGHLIGHT: We propose a non-convex estimator based on rank-constrained likelihood maximization.

313, TITLE: Asynchronous Decentralized Parallel Stochastic Gradient Descent  
<http://proceedings.mlr.press/v80/lian18a.html>  
AUTHORS: Xiangru Lian, Wei Zhang, Ce Zhang, Ji Liu  
HIGHLIGHT: In this paper, we propose an asynchronous decentralized stochastic gradient descent algorithm (AD-PSGD) satisfying all above expectations.

314, TITLE: RLlib: Abstractions for Distributed Reinforcement Learning  
<http://proceedings.mlr.press/v80/liang18b.html>  
AUTHORS: Eric Liang, Richard Liaw, Robert Nishihara, Philipp Moritz, Roy Fox, Ken Goldberg, Joseph Gonzalez, Michael Jordan, Ion Stoica  
HIGHLIGHT: We argue for distributing RL components in a composable way by adapting algorithms for top-down hierarchical control, thereby encapsulating parallelism and resource requirements within short-running compute tasks.

315, TITLE: On the Spectrum of Random Features Maps of High Dimensional Data  
<http://proceedings.mlr.press/v80/liao18a.html>  
AUTHORS: Zhenyu Liao, Romain Couillet

**HIGHLIGHT:** In this paper we leverage the "concentration" phenomenon induced by random matrix theory to perform a spectral analysis on the Gram matrix of these random feature maps, here for Gaussian mixture models of simultaneously large dimension and size.

316, **TITLE:** The Dynamics of Learning: A Random Matrix Approach  
<http://proceedings.mlr.press/v80/liao18b.html>

**AUTHORS:** Zhenyu Liao, Romain Couillet

**HIGHLIGHT:** In this paper, we introduce a random matrix-based framework to analyze the learning dynamics of a single-layer linear network on a binary classification problem, for data of simultaneously large dimension and size, trained by gradient descent.

317, **TITLE:** Reviving and Improving Recurrent Back-Propagation  
<http://proceedings.mlr.press/v80/liao18c.html>

**AUTHORS:** Renjie Liao, Yuwen Xiong, Ethan Fetaya, Lisa Zhang, KiJung Yoon, Xaq Pitkow, Raquel Urtasun, Richard Zemel

**HIGHLIGHT:** In this paper, we revisit the recurrent back-propagation (RBP) algorithm, discuss the conditions under which it applies as well as how to satisfy them in deep neural networks.

318, **TITLE:** Optimal Distributed Learning with Multi-pass Stochastic Gradient Methods  
<http://proceedings.mlr.press/v80/lin18a.html>

**AUTHORS:** Junhong Lin, Volkan Cevher

**HIGHLIGHT:** We investigate distributed stochastic gradient methods (SGM), with mini-batches and multi-passes over the data.

319, **TITLE:** Optimal Rates of Sketched-regularized Algorithms for Least-Squares Regression over Hilbert Spaces  
<http://proceedings.mlr.press/v80/lin18b.html>

**AUTHORS:** Junhong Lin, Volkan Cevher

**HIGHLIGHT:** We investigate regularized algorithms combining with projection for least-squares regression problem over a Hilbert space, covering nonparametric regression over a reproducing kernel Hilbert space.

320, **TITLE:** Level-Set Methods for Finite-Sum Constrained Convex Optimization  
<http://proceedings.mlr.press/v80/lin18c.html>

**AUTHORS:** Qihang Lin, Runchao Ma, Tianbao Yang

**HIGHLIGHT:** We consider two level-set methods to solve this class of problems, an existing inexact Newton method and a new feasible level-set method.

321, **TITLE:** Detecting and Correcting for Label Shift with Black Box Predictors  
<http://proceedings.mlr.press/v80/lipton18a.html>

**AUTHORS:** Zachary Lipton, Yu-Xiang Wang, Alexander Smola

**HIGHLIGHT:** We propose Black Box Shift Estimation (BBSE) to estimate the test distribution  $p(y)$ .

322, **TITLE:** Generalized Robust Bayesian Committee Machine for Large-scale Gaussian Process Regression  
<http://proceedings.mlr.press/v80/liu18a.html>

**AUTHORS:** Haitao Liu, Jianfei Cai, Yi Wang, Yew Soon Ong

**HIGHLIGHT:** Generalized Robust Bayesian Committee Machine for Large-scale Gaussian Process Regression

323, **TITLE:** Towards Black-box Iterative Machine Teaching  
<http://proceedings.mlr.press/v80/liu18b.html>

**AUTHORS:** Weiyang Liu, Bo Dai, Xingguo Li, Zhen Liu, James Rehg, Le Song

**HIGHLIGHT:** In this paper, we make an important step towards the black-box machine teaching by considering the cross-space machine teaching, where the teacher and the learner use different feature representations and the teacher can not fully observe the learner's model.

- 324, TITLE: Delayed Impact of Fair Machine Learning  
<http://proceedings.mlr.press/v80/liu18c.html>  
AUTHORS: Lydia T. Liu, Sarah Dean, Esther Rolf, Max Simchowitz, Moritz Hardt  
HIGHLIGHT: We study how static fairness criteria interact with temporal indicators of well-being, such as long-term improvement, stagnation, and decline in a variable of interest.
- 325, TITLE: A Two-Step Computation of the Exact GAN Wasserstein Distance  
<http://proceedings.mlr.press/v80/liu18d.html>  
AUTHORS: Huidong Liu, Xianfeng GU, Dimitris Samaras  
HIGHLIGHT: In this paper, we propose a two-step method to compute the Wasserstein distance in Wasserstein Generative Adversarial Networks (WGANs): 1) The convex part of our objective can be solved by linear programming; 2) The non-convex residual can be approximated by a deep neural network.
- 326, TITLE: Open Category Detection with PAC Guarantees  
<http://proceedings.mlr.press/v80/liu18e.html>  
AUTHORS: Si Liu, Risheek Garrepalli, Thomas Dietterich, Alan Fern, Dan Hendrycks  
HIGHLIGHT: In this paper, we take a step toward addressing this gap by studying a simple, but practically-relevant variant of open category detection.
- 327, TITLE: Fast Variance Reduction Method with Stochastic Batch Size  
<http://proceedings.mlr.press/v80/liu18f.html>  
AUTHORS: Xuanqing Liu, Cho-Jui Hsieh  
HIGHLIGHT: In this paper we study a family of variance reduction methods with randomized batch size—at each step, the algorithm first randomly chooses the batch size and then selects a batch of samples to conduct a variance-reduced stochastic update.
- 328, TITLE: Fast Stochastic AUC Maximization with  $O(1/n)$ -Convergence Rate  
<http://proceedings.mlr.press/v80/liu18g.html>  
AUTHORS: Mingrui Liu, Xiaoxuan Zhang, Zaiyi Chen, Xiaoyu Wang, Tianbao Yang  
HIGHLIGHT: In this paper, we consider statistical learning with AUC (area under ROC curve) maximization in the classical stochastic setting where one random data drawn from an unknown distribution is revealed at each iteration for updating the model.
- 329, TITLE: On Matching Pursuit and Coordinate Descent  
<http://proceedings.mlr.press/v80/locatello18a.html>  
AUTHORS: Francesco Locatello, Anant Raj, Sai Praneeth Karimireddy, Gunnar Raetsch, Bernhard Schölkopf, Sebastian Stich, Martin Jaggi  
HIGHLIGHT: Exploiting the connection between the two algorithms, we present a unified analysis of both, providing affine invariant sublinear  $O(1/t)$  rates on smooth objectives and linear convergence on strongly convex objectives.
- 330, TITLE: PDE-Net: Learning PDEs from Data  
<http://proceedings.mlr.press/v80/long18a.html>  
AUTHORS: Zichao Long, Yiping Lu, Xianzhong Ma, Bin Dong  
HIGHLIGHT: Inspired by the latest development of neural network designs in deep learning, we propose a new feed-forward deep network, called PDE-Net, to fulfill two objectives at the same time: to accurately predict dynamics of complex systems and to uncover the underlying hidden PDE models.
- 331, TITLE: Error Estimation for Randomized Least-Squares Algorithms via the Bootstrap  
<http://proceedings.mlr.press/v80/lopes18a.html>  
AUTHORS: Miles Lopes, Shusen Wang, Michael Mahoney  
HIGHLIGHT: As a more practical alternative, we propose a bootstrap method to compute a posteriori error estimates for randomized LS algorithms.

- 332, TITLE: Constraining the Dynamics of Deep Probabilistic Models  
<http://proceedings.mlr.press/v80/lorenzi18a.html>  
AUTHORS: Marco Lorenzi, Maurizio Filippone  
HIGHLIGHT: We introduce a novel generative formulation of deep probabilistic models implementing "soft" constraints on their function dynamics.
- 333, TITLE: Spectrally Approximating Large Graphs with Smaller Graphs  
<http://proceedings.mlr.press/v80/loukas18a.html>  
AUTHORS: Andreas Loukas, Pierre Vandergheynst  
HIGHLIGHT: We provide conditions such that the principal eigenvalues and eigenspaces of a coarsened and original graph Laplacian matrices are close.
- 334, TITLE: The Edge Density Barrier: Computational-Statistical Tradeoffs in Combinatorial Inference  
<http://proceedings.mlr.press/v80/lu18a.html>  
AUTHORS: Hao Lu, Yuan Cao, Zhuoran Yang, Junwei Lu, Han Liu, Zhaoran Wang  
HIGHLIGHT: In this paper, we quantify the minimum computational complexity required to attain the information-theoretic limits based on an oracle computational model.
- 335, TITLE: Accelerating Greedy Coordinate Descent Methods  
<http://proceedings.mlr.press/v80/lu18b.html>  
AUTHORS: Haihao Lu, Robert Freund, Vahab Mirrokni  
HIGHLIGHT: We introduce and study two algorithms to accelerate greedy coordinate descent in theory and in practice: Accelerated Semi-Greedy Coordinate Descent (ASCD) and Accelerated Greedy Coordinate Descent (AGCD).
- 336, TITLE: Structured Variationally Auto-encoded Optimization  
<http://proceedings.mlr.press/v80/lu18c.html>  
AUTHORS: Xiaoyu Lu, Javier Gonzalez, Zhenwen Dai, Neil Lawrence  
HIGHLIGHT: We use the \as as a case study to describe our approach, that can be easily generalized to other domains.
- 337, TITLE: Beyond Finite Layer Neural Networks: Bridging Deep Architectures and Numerical Differential Equations  
<http://proceedings.mlr.press/v80/lu18d.html>  
AUTHORS: Yiping Lu, Aoxiao Zhong, Quanzheng Li, Bin Dong  
HIGHLIGHT: As an example, we propose a linear multi-step architecture (LM-architecture) which is inspired by the linear multi-step method solving ordinary differential equations.
- 338, TITLE: End-to-end Active Object Tracking via Reinforcement Learning  
<http://proceedings.mlr.press/v80/luo18a.html>  
AUTHORS: Wenhan Luo, Peng Sun, Fangwei Zhong, Wei Liu, Tong Zhang, Yizhou Wang  
HIGHLIGHT: To address these issues, we propose, in this paper, an end-to-end solution via deep reinforcement learning, where a ConvNet-LSTM function approximator is adopted for the direct frame-to-action prediction.
- 339, TITLE: Competitive Caching with Machine Learned Advice  
<http://proceedings.mlr.press/v80/lykouris18a.html>  
AUTHORS: Thodoris Lykouris, Sergei Vassilvtskii  
HIGHLIGHT: We apply this framework to the traditional caching problem  $\{—\}$  creating an eviction strategy for a cache of size  $k$ .
- 340, TITLE: Batch Bayesian Optimization via Multi-objective Acquisition Ensemble for Automated Analog Circuit Design  
<http://proceedings.mlr.press/v80/lyu18a.html>  
AUTHORS: Wenlong Lyu, Fan Yang, Changhao Yan, Dian Zhou, Xuan Zeng



HIGHLIGHT: In this paper, a novel batch Bayesian optimization approach is proposed.

341, TITLE: Celer: a Fast Solver for the Lasso with Dual Extrapolation

<http://proceedings.mlr.press/v80/massias18a.html>

AUTHORS: Mathurin MASSIAS, Alexandre Gramfort, Joseph Salmon

HIGHLIGHT: Here, we propose an extrapolation technique starting from a sequence of iterates in the dual that leads to the construction of improved dual points.

342, TITLE: The Power of Interpolation: Understanding the Effectiveness of SGD in Modern Over-parametrized Learning

<http://proceedings.mlr.press/v80/ma18a.html>

AUTHORS: Siyuan Ma, Raef Bassily, Mikhail Belkin

HIGHLIGHT: In this paper we aim to formally explain the phenomenon of fast convergence of Stochastic Gradient Descent (SGD) observed in modern machine learning.

343, TITLE: Gradient Descent for Sparse Rank-One Matrix Completion for Crowd-Sourced Aggregation of Sparsely Interacting Workers

<http://proceedings.mlr.press/v80/ma18b.html>

AUTHORS: Yao Ma, Alexander Olshevsky, Csaba Szepesvari, Venkatesh Saligrama

HIGHLIGHT: We formulate skill estimation as a rank-one correlation-matrix completion problem, where the observed components correspond to observed label correlation between workers.

344, TITLE: Implicit Regularization in Nonconvex Statistical Estimation: Gradient Descent Converges Linearly for Phase Retrieval and Matrix Completion

<http://proceedings.mlr.press/v80/ma18c.html>

AUTHORS: Cong Ma, Kaizheng Wang, Yuejie Chi, Yuxin Chen

HIGHLIGHT: Focusing on two statistical estimation problems, i.e. solving random quadratic systems of equations and low-rank matrix completion, we establish that gradient descent achieves near-optimal statistical and computational guarantees without explicit regularization.

345, TITLE: Dimensionality-Driven Learning with Noisy Labels

<http://proceedings.mlr.press/v80/ma18d.html>

AUTHORS: Xingjun Ma, Yisen Wang, Michael E. Houle, Shuo Zhou, Sarah Erfani, Shutao Xia, Sudanthi Wijewickrema, James Bailey

HIGHLIGHT: We propose a new perspective for understanding DNN generalization for such datasets, by investigating the dimensionality of the deep representation subspace of training samples.

346, TITLE: Approximate message passing for amplitude based optimization

<http://proceedings.mlr.press/v80/ma18e.html>

AUTHORS: Junjie Ma, Ji Xu, Arian Maleki

HIGHLIGHT: We consider an  $\ell_2$ -regularized non-convex optimization problem for recovering signals from their noisy phaseless observations.

347, TITLE: Orthogonal Machine Learning: Power and Limitations

<http://proceedings.mlr.press/v80/mackey18a.html>

AUTHORS: Lester Mackey, Vasilis Syrgkanis, Ilias Zadik

HIGHLIGHT: We show that the  $n^{-1/4}$  requirement can be improved to  $n^{-1/(2k+2)}$  by employing a  $k$ -th order notion of orthogonality that grants robustness to more complex or higher-dimensional nuisance parameters.

348, TITLE: Learning Adversarially Fair and Transferable Representations

<http://proceedings.mlr.press/v80/madras18a.html>

AUTHORS: David Madras, Elliot Creager, Toniann Pitassi, Richard Zemel

HIGHLIGHT: In this paper, we advocate for representation learning as the key to mitigating unfair prediction outcomes downstream.

- 349, TITLE: An Efficient, Generalized Bellman Update For Cooperative Inverse Reinforcement Learning  
<http://proceedings.mlr.press/v80/malik18a.html>  
AUTHORS: Dhruv Malik, Malayandi Palaniappan, Jaime Fisac, Dylan Hadfield-Menell, Stuart Russell, Anca Dragan  
HIGHLIGHT: In this work, we exploit a specific property of CIRL: the human is a full information agent.
- 350, TITLE: Iterative Amortized Inference  
<http://proceedings.mlr.press/v80/marino18a.html>  
AUTHORS: Joe Marino, Yisong Yue, Stephan Mandt  
HIGHLIGHT: We aim toward closing this gap by proposing iterative inference models, which learn to perform inference optimization through repeatedly encoding gradients.
- 351, TITLE: Streaming Principal Component Analysis in Noisy Setting  
<http://proceedings.mlr.press/v80/marinov18a.html>  
AUTHORS: Teodor Vanislavov Marinov, Poorya Mianjy, Raman Arora  
HIGHLIGHT: We present computationally efficient algorithms with sub-linear regret bounds for PCA in the presence of noise, missing data, and gross outliers.
- 352, TITLE: Fast Approximate Spectral Clustering for Dynamic Networks  
<http://proceedings.mlr.press/v80/martin18a.html>  
AUTHORS: Lionel Martin, Andreas Loukas, Pierre Vandergheynst  
HIGHLIGHT: We show that the proposed algorithm achieves clustering assignments with quality approximating that of spectral clustering and that it can yield significant complexity benefits when the graph dynamics are appropriately bounded.
- 353, TITLE: Bayesian Model Selection for Change Point Detection and Clustering  
<http://proceedings.mlr.press/v80/mazhar18a.html>  
AUTHORS: Othmane Mazhar, Cristian Rojas, Carlo Fischione, Mohammad Reza Hesamzadeh  
HIGHLIGHT: Our approach is to model it as a nonparametric penalized least square model selection on a family of models indexed over the collection of partitions of the design points and propose a computationally efficient algorithm to approximately solve it.
- 354, TITLE: Optimization, fast and slow: optimally switching between local and Bayesian optimization  
<http://proceedings.mlr.press/v80/mcleod18a.html>  
AUTHORS: Mark McLeod, Stephen Roberts, Michael A. Osborne  
HIGHLIGHT: We develop the first Bayesian Optimization algorithm, BLOSSOM, which selects between multiple alternative acquisition functions and traditional local optimization at each step.
- 355, TITLE: Bounds on the Approximation Power of Feedforward Neural Networks  
<http://proceedings.mlr.press/v80/mehrabi18a.html>  
AUTHORS: Mohammad Mehrabi, Aslan Tchamkerten, MANSOOR YOUSEFI  
HIGHLIGHT: Bounds on the Approximation Power of Feedforward Neural Networks
- 356, TITLE: Differentiable Dynamic Programming for Structured Prediction and Attention  
<http://proceedings.mlr.press/v80/mensch18a.html>  
AUTHORS: Arthur Mensch, Mathieu Blondel  
HIGHLIGHT: To address this issue, we propose to smooth the max operator in the dynamic programming recursion, using a strongly convex regularizer.
- 357, TITLE: Ranking Distributions based on Noisy Sorting  
<http://proceedings.mlr.press/v80/mesaoudi-paul18a.html>  
AUTHORS: Adil El Mesaoudi-Paul, Eyke H?llermeier, Robert Busa-Fekete

**HIGHLIGHT:** We propose a new statistical model for ranking data, i.e., a new family of probability distributions on permutations.

358, **TITLE:** Which Training Methods for GANs do actually Converge?

<http://proceedings.mlr.press/v80/mescheder18a.html>

**AUTHORS:** Lars Mescheder, Andreas Geiger, Sebastian Nowozin

**HIGHLIGHT:** In this paper, we show that the requirement of absolute continuity is necessary: we describe a simple yet prototypical counterexample showing that in the more realistic case of distributions that are not absolutely continuous, unregularized GAN training is not always convergent.

359, **TITLE:** Configurable Markov Decision Processes

<http://proceedings.mlr.press/v80/metelli18a.html>

**AUTHORS:** Alberto Maria Metelli, Mirco Mutti, Marcello Restelli

**HIGHLIGHT:** In this paper, we propose a novel framework, Configurable Markov Decision Processes (Conf-MDPs), to model this new type of interaction with the environment.

360, **TITLE:** prDeep: Robust Phase Retrieval with a Flexible Deep Network

<http://proceedings.mlr.press/v80/metzler18a.html>

**AUTHORS:** Christopher Metzler, Phillip Schniter, Ashok Veeraraghavan, baraniuk

**HIGHLIGHT:** In this work we leverage the regularization-by-denoising framework and a convolutional neural network denoiser to create prDeep, a new phase retrieval algorithm that is both robust and broadly applicable.

361, **TITLE:** Pseudo-task Augmentation: From Deep Multitask Learning to Intratask Sharing?and Back

<http://proceedings.mlr.press/v80/meyerson18a.html>

**AUTHORS:** Elliot Meyerson, Risto Miikkulainen

**HIGHLIGHT:** This paper adapts ideas from deep multitask learning to the setting where only a single task is available.

362, **TITLE:** The Hidden Vulnerability of Distributed Learning in Byzantium

<http://proceedings.mlr.press/v80/mhamdi18a.html>

**AUTHORS:** El Mahdi El Mhamdi, Rachid Guerraoui, S?bastien Rouault

**HIGHLIGHT:** We show in this paper that convergence is not enough.

363, **TITLE:** Stochastic PCA with  $\ell_2$  and  $\ell_1$  Regularization

<http://proceedings.mlr.press/v80/mianjy18a.html>

**AUTHORS:** Poorya Mianjy, Raman Arora

**HIGHLIGHT:** Therefore, in this paper, we study variants of stochastic gradient descent for a convex relaxation of PCA with (a)  $\ell_2$ , (b)  $\ell_1$ , and (c) elastic net ( $\ell_1 + \ell_2$ ) regularization in the hope that these variants yield (a) better iteration complexity, (b) better control on the rank of the intermediate iterates, and (c) both, respectively.

364, **TITLE:** On the Implicit Bias of Dropout

<http://proceedings.mlr.press/v80/mianjy18b.html>

**AUTHORS:** Poorya Mianjy, Raman Arora, Rene Vidal

**HIGHLIGHT:** In this paper, we focus on understanding such a bias induced in learning through dropout, a popular technique to avoid overfitting in deep learning.

365, **TITLE:** One-Shot Segmentation in Clutter

<http://proceedings.mlr.press/v80/michaelis18a.html>

**AUTHORS:** Claudio Michaelis, Matthias Bethge, Alexander Ecker

**HIGHLIGHT:** We therefore introduce MaskNet, an improved model that attends to multiple candidate locations, generates segmentation proposals to mask out background clutter and selects among the segmented objects. We propose a novel dataset, which we call cluttered Omniglot.

- 366, TITLE: Differentiable plasticity: training plastic neural networks with backpropagation  
<http://proceedings.mlr.press/v80/miconi18a.html>  
AUTHORS: Thomas Miconi, Kenneth Stanley, Jeff Clune  
HIGHLIGHT: How can we build agents that keep learning from experience, quickly and efficiently, after their initial training?
- 367, TITLE: Training Neural Machines with Trace-Based Supervision  
<http://proceedings.mlr.press/v80/mirman18a.html>  
AUTHORS: Matthew Mirman, Dimitar Dimitrov, Pavle Djordjevic, Timon Gehr, Martin Vechev  
HIGHLIGHT: To define the class of neural machines amenable to trace-based supervision, we introduce the concept of a differential neural computational machine (dNCM) and show that several existing architectures (NTMs, NRAMs) can be described as dNCMs.
- 368, TITLE: Differentiable Abstract Interpretation for Provably Robust Neural Networks  
<http://proceedings.mlr.press/v80/mirman18b.html>  
AUTHORS: Matthew Mirman, Timon Gehr, Martin Vechev  
HIGHLIGHT: We introduce a scalable method for training robust neural networks based on abstract interpretation.
- 369, TITLE: A Delay-tolerant Proximal-Gradient Algorithm for Distributed Learning  
<http://proceedings.mlr.press/v80/mishchenko18a.html>  
AUTHORS: Konstantin Mishchenko, Franck Iutzeler, J?r?me Malick, Massih-Reza Amini  
HIGHLIGHT: In this setting, we propose and analyze a flexible asynchronous optimization algorithm for solving nonsmooth learning problems.
- 370, TITLE: Data Summarization at Scale: A Two-Stage Submodular Approach  
<http://proceedings.mlr.press/v80/mitrovic18a.html>  
AUTHORS: Marko Mitrovic, Ehsan Kazemi, Morteza Zadimoghaddam, Amin Karbasi  
HIGHLIGHT: In this paper, we develop the first streaming and distributed solutions to this problem.
- 371, TITLE: The Hierarchical Adaptive Forgetting Variational Filter  
<http://proceedings.mlr.press/v80/moens18a.html>  
AUTHORS: Vincent Moens  
HIGHLIGHT: We present a hierarchical Bayesian algorithm that aims at learning a time-specific approximate posterior distribution of the parameters describing the distribution of the data observed.
- 372, TITLE: Decentralized Submodular Maximization: Bridging Discrete and Continuous Settings  
<http://proceedings.mlr.press/v80/mokhtari18a.html>  
AUTHORS: Aryan Mokhtari, Hamed Hassani, Amin Karbasi  
HIGHLIGHT: In this paper, we showcase the interplay between discrete and continuous optimization in network-structured settings.
- 373, TITLE: DICOD: Distributed Convolutional Coordinate Descent for Convolutional Sparse Coding  
<http://proceedings.mlr.press/v80/moreau18a.html>  
AUTHORS: Thomas Moreau, Laurent Oudre, Nicolas Vayatis  
HIGHLIGHT: In this paper, we introduce DICOD, a convolutional sparse coding algorithm which builds shift invariant representations for long signals.
- 374, TITLE: WHInter: A Working set algorithm for High-dimensional sparse second order Interaction models  
<http://proceedings.mlr.press/v80/morvan18a.html>  
AUTHORS: Marine Le Morvan, Jean-Philippe Vert  
HIGHLIGHT: Here we present WHInter, a working set algorithm to solve large  $\ell_1$ -regularised problems with two-way interactions for binary design matrices.

- 375, TITLE: Dropout Training, Data-dependent Regularization, and Generalization Bounds  
<http://proceedings.mlr.press/v80/mou18a.html>  
AUTHORS: Wenlong Mou, Yuchen Zhou, Jun Gao, Liwei Wang  
HIGHLIGHT: For neural networks, we estimate upper and lower bounds for the variance induced by truthful dropout, a variant of dropout that we propose to ensure unbiased output and fit into our framework, and the variance bounds exhibits connection to adaptive regularization methods.
- 376, TITLE: Kernelized Synaptic Weight Matrices  
<http://proceedings.mlr.press/v80/muller18a.html>  
AUTHORS: Lorenz Muller, Julien Martel, Giacomo Indiveri  
HIGHLIGHT: In this paper we introduce a novel neural network architecture, in which weight matrices are re-parametrized in terms of low-dimensional vectors, interacting through kernel functions.
- 377, TITLE: Rapid Adaptation with Conditionally Shifted Neurons  
<http://proceedings.mlr.press/v80/munkhdalai18a.html>  
AUTHORS: Tsendsuren Munkhdalai, Xingdi Yuan, Soroush Mehri, Adam Trischler  
HIGHLIGHT: We describe a mechanism by which artificial neural networks can learn rapid adaptation - the ability to adapt on the fly, with little data, to new tasks - that we call conditionally shifted neurons.
- 378, TITLE: On the Relationship between Data Efficiency and Error for Uncertainty Sampling  
<http://proceedings.mlr.press/v80/musmann18a.html>  
AUTHORS: Stephen Musmann, Percy Liang  
HIGHLIGHT: This paper poses a basic question: when is active learning actually helpful?
- 379, TITLE: Fitting New Speakers Based on a Short Untranscribed Sample  
<http://proceedings.mlr.press/v80/nachmani18a.html>  
AUTHORS: Eliya Nachmani, Adam Polyak, Yaniv Taigman, Lior Wolf  
HIGHLIGHT: We present a method that is designed to capture a new speaker from a short untranscribed audio sample.
- 380, TITLE: Smoothed Action Value Functions for Learning Gaussian Policies  
<http://proceedings.mlr.press/v80/nachum18a.html>  
AUTHORS: Ofir Nachum, Mohammad Norouzi, George Tucker, Dale Schuurmans  
HIGHLIGHT: We propose a new notion of action value defined by a Gaussian smoothed version of the expected Q-value.
- 381, TITLE: Nearly Optimal Robust Subspace Tracking  
<http://proceedings.mlr.press/v80/narayanamurthy18a.html>  
AUTHORS: Praneeth Narayanamurthy, Namrata Vaswani  
HIGHLIGHT: This work develops a recursive projected compressive sensing algorithm called “Nearly Optimal RST (NORST)”, and obtains one of the first guarantees for it.
- 382, TITLE: Stochastic Proximal Algorithms for AUC Maximization  
<http://proceedings.mlr.press/v80/natole18a.html>  
AUTHORS: Michael Natole, Yiming Ying, Siwei Lyu  
HIGHLIGHT: In this paper, we develop a novel stochastic proximal algorithm for AUC maximization which is referred to as SPAM.
- 383, TITLE: Mitigating Bias in Adaptive Data Gathering via Differential Privacy  
<http://proceedings.mlr.press/v80/nee18a.html>  
AUTHORS: Seth Neel, Aaron Roth

**HIGHLIGHT:** In this paper, we show that this problem is mitigated if the data collection procedure is differentially private.

384, **TITLE:** Optimization Landscape and Expressivity of Deep CNNs  
<http://proceedings.mlr.press/v80/nguyen18a.html>  
**AUTHORS:** Quynh Nguyen, Matthias Hein  
**HIGHLIGHT:** We analyze the loss landscape and expressiveness of practical deep convolutional neural networks (CNNs) with shared weights and max pooling layers.

385, **TITLE:** Neural Networks Should Be Wide Enough to Learn Disconnected Decision Regions  
<http://proceedings.mlr.press/v80/nguyen18b.html>  
**AUTHORS:** Quynh Nguyen, Mahesh Chandra Mukkamala, Matthias Hein  
**HIGHLIGHT:** In this paper we argue that sufficient width of a feedforward network is equally important by answering the simple question under which conditions the decision regions of a neural network are connected.

386, **TITLE:** SGD and Hogwild! Convergence Without the Bounded Gradients Assumption  
<http://proceedings.mlr.press/v80/nguyen18c.html>  
**AUTHORS:** Lam Nguyen, PHUONG HA NGUYEN, Marten Dijk, Peter Richtarik, Katya Scheinberg, Martin Takac  
**HIGHLIGHT:** We then move on the asynchronous parallel setting, and prove convergence of Hogwild!

387, **TITLE:** Active Testing: An Efficient and Robust Framework for Estimating Accuracy  
<http://proceedings.mlr.press/v80/nguyen18d.html>  
**AUTHORS:** Phuc Nguyen, Deva Ramanan, Charles Fowlkes  
**HIGHLIGHT:** In this work, we re-formulate the problem as one of active testing, and examine strategies for efficiently querying a user so as to obtain an accurate performance estimate with minimal vet-ting.

388, **TITLE:** On Learning Sparsely Used Dictionaries from Incomplete Samples  
<http://proceedings.mlr.press/v80/nguyen18e.html>  
**AUTHORS:** Thanh Nguyen, Akshay Soni, Chinmay Hegde  
**HIGHLIGHT:** In this paper, we provide provable approaches for learning – from incomplete samples – a family of dictionaries whose atoms have sufficiently “spread-out” mass.

389, **TITLE:** Learning Continuous Hierarchies in the Lorentz Model of Hyperbolic Geometry  
<http://proceedings.mlr.press/v80/nickel18a.html>  
**AUTHORS:** Maximillian Nickel, Douwe Kiela  
**HIGHLIGHT:** For this purpose, we study different models of hyperbolic space and find that learning embeddings in the Lorentz model is substantially more efficient than in the Poincaré-ball model.

390, **TITLE:** State Space Gaussian Processes with Non-Gaussian Likelihood  
<http://proceedings.mlr.press/v80/nickisch18a.html>  
**AUTHORS:** Hannes Nickisch, Arno Solin, Alexander Grigorevskiy  
**HIGHLIGHT:** We present means of combining the efficient  $O(n)$  state space methodology with existing inference methods.

391, **TITLE:** SparseMAP: Differentiable Sparse Structured Inference  
<http://proceedings.mlr.press/v80/niculae18a.html>  
**AUTHORS:** Vlad Niculae, Andre Martins, Mathieu Blondel, Claire Cardie  
**HIGHLIGHT:** To tackle it, we introduce SparseMAP, a new method for sparse structured inference, together with corresponding loss functions.

392, **TITLE:** A Theoretical Explanation for Perplexing Behaviors of Backpropagation-based Visualizations  
<http://proceedings.mlr.press/v80/nie18a.html>

- AUTHORS: Weili Nie, Yang Zhang, Ankit Patel  
HIGHLIGHT: Motivated by this, we develop a theoretical explanation revealing that GBP and DeconvNet are essentially doing (partial) image recovery which is unrelated to the network decisions.
- 393, TITLE: Functional Gradient Boosting based on Residual Network Perception  
<http://proceedings.mlr.press/v80/nitanda18a.html>  
AUTHORS: Atsushi Nitanda, Taiji Suzuki  
HIGHLIGHT: In this paper, we adopt this viewpoint to construct a new gradient boosting method, which is known to be very powerful in data analysis.
- 394, TITLE: Beyond 1/2-Approximation for Submodular Maximization on Massive Data Streams  
<http://proceedings.mlr.press/v80/norouzi-fard18a.html>  
AUTHORS: Ashkan Norouzi-Fard, Jakub Tarnawski, Slobodan Mitrovic, Amir Zandieh, Aidasadat Mousavifar, Ola Svensson  
HIGHLIGHT: We consider this question in the streaming setting, where elements arrive over time at a fast pace and thus we need to design an efficient, low-memory algorithm.
- 395, TITLE: The Uncertainty Bellman Equation and Exploration  
<http://proceedings.mlr.press/v80/odonoghue18a.html>  
AUTHORS: Brendan O'Donoghue, Ian Osband, Remi Munos, Vlad Mnih  
HIGHLIGHT: In this paper we consider a similar uncertainty Bellman equation (UBE), which connects the uncertainty at any time-step to the expected uncertainties at subsequent time-steps, thereby extending the potential exploratory benefit of a policy beyond individual time-steps.
- 396, TITLE: Is Generator Conditioning Causally Related to GAN Performance?  
<http://proceedings.mlr.press/v80/odena18a.html>  
AUTHORS: Augustus Odena, Jacob Buckman, Catherine Olsson, Tom Brown, Christopher Olah, Colin Raffel, Ian Goodfellow  
HIGHLIGHT: We then test the hypothesis that this relationship is causal by proposing a “regularization” technique (called Jacobian Clamping) that softly penalizes the condition number of the generator Jacobian.
- 397, TITLE: Learning in Reproducing Kernel Krei?n Spaces  
<http://proceedings.mlr.press/v80/oglic18a.html>  
AUTHORS: Dino Oglic, Thomas Gaertner  
HIGHLIGHT: We present an algorithm that can find a globally optimal solution to this non-convex optimization problem in time cubic in the number of instances.
- 398, TITLE: BOCK : Bayesian Optimization with Cylindrical Kernels  
<http://proceedings.mlr.press/v80/oh18a.html>  
AUTHORS: ChangYong Oh, Efstratios Gavves, Max Welling  
HIGHLIGHT: In this paper, we propose BOCK, Bayesian Optimization with Cylindrical Kernels, whose basic idea is to transform the ball geometry of the search space using a cylindrical transformation.
- 399, TITLE: Self-Imitation Learning  
<http://proceedings.mlr.press/v80/oh18b.html>  
AUTHORS: Junhyuk Oh, Yijie Guo, Satinder Singh, Honglak Lee  
HIGHLIGHT: This paper proposes Self-Imitation Learning (SIL), a simple off-policy actor-critic algorithm that learns to reproduce the agent’s past good decisions.
- 400, TITLE: A probabilistic framework for multi-view feature learning with many-to-many associations via neural networks  
<http://proceedings.mlr.press/v80/okuno18a.html>  
AUTHORS: Akifumi Okuno, Tetsuya Hada, Hidetoshi Shimodaira

**HIGHLIGHT:** A simple framework Probabilistic Multi-view Graph Embedding (PMvGE) is proposed for multi-view feature learning with many-to-many associations so that it generalizes various existing multi-view methods.

401, TITLE: Transformation Autoregressive Networks

<http://proceedings.mlr.press/v80/oliva18a.html>

AUTHORS: Junior Oliva, Avinava Dubey, Manzil Zaheer, Barnabas Poczos, Ruslan Salakhutdinov, Eric Xing, Jeff Schneider

**HIGHLIGHT:** In this work, we attempt to systematically characterize methods for density estimation. Finally we introduce a novel data driven framework for learning a family of distributions.

402, TITLE: Design of Experiments for Model Discrimination Hybridising Analytical and Data-Driven Approaches

<http://proceedings.mlr.press/v80/olofsson18a.html>

AUTHORS: Simon Olofsson, Marc Deisenroth, Ruth Misener

**HIGHLIGHT:** In design of experiments for model discrimination, where the goal is to design maximally informative physical experiments in order to discriminate between rival predictive models, research has focused either on analytical approaches, which cannot manage all functions, or on data-driven approaches, which may have computational difficulties or lack interpretable marginal predictive distributions.

403, TITLE: Parallel WaveNet: Fast High-Fidelity Speech Synthesis

<http://proceedings.mlr.press/v80/oord18a.html>

AUTHORS: Aaron Oord, Yazhe Li, Igor Babuschkin, Karen Simonyan, Oriol Vinyals, Koray Kavukcuoglu, George Driessche, Edward Lockhart, Luis Cobo, Florian Stimberg, Norman Casagrande, Dominik Grewe, Seb Noury, Sander Dieleman, Erich Elsen, Nal Kalchbrenner, Heiga Zen, Alex Graves, Helen King, Tom Walters, Dan Belov, Demis Hassabis

**HIGHLIGHT:** This paper introduces Probability Density Distillation, a new method for training a parallel feed-forward network from a trained WaveNet with no significant difference in quality.

404, TITLE: Learning Localized Spatio-Temporal Models From Streaming Data

<http://proceedings.mlr.press/v80/osama18a.html>

AUTHORS: Muhammad Osama, Dave Zachariah, Thomas Sch?n

**HIGHLIGHT:** We address the problem of predicting spatio-temporal processes with temporal patterns that vary across spatial regions, when data is obtained as a stream.

405, TITLE: Autoregressive Quantile Networks for Generative Modeling

<http://proceedings.mlr.press/v80/ostrovski18a.html>

AUTHORS: Georg Ostrovski, Will Dabney, Remi Munos

**HIGHLIGHT:** We introduce autoregressive implicit quantile networks (AIQN), a fundamentally different approach to generative modeling than those commonly used, that implicitly captures the distribution using quantile regression.

406, TITLE: Efficient First-Order Algorithms for Adaptive Signal Denoising

<http://proceedings.mlr.press/v80/ostrovskii18a.html>

AUTHORS: Dmitrii Ostrovskii, Zaid Harchaoui

**HIGHLIGHT:** We consider the problem of discrete-time signal denoising, focusing on a specific family of non-linear convolution-type estimators.

407, TITLE: Analyzing Uncertainty in Neural Machine Translation

<http://proceedings.mlr.press/v80/ott18a.html>

AUTHORS: Myle Ott, Michael Auli, David Grangier, Marc?Aurelio Ranzato

**HIGHLIGHT:** We propose tools and metrics to assess how uncertainty in the data is captured by the model distribution and how it affects search strategies that generate translations.

408, TITLE: Learning Compact Neural Networks with Regularization



<http://proceedings.mlr.press/v80/oymak18a.html>

AUTHORS: Samet Oymak

HIGHLIGHT: We propose and analyze regularized gradient descent algorithms for learning shallow neural networks.

409, TITLE: Tree Edit Distance Learning via Adaptive Symbol Embeddings

<http://proceedings.mlr.press/v80/paassen18a.html>

AUTHORS: Benjamin Paasen, Claudio Gallicchio, Alessio Micheli, Barbara Hammer

HIGHLIGHT: In this contribution, we propose a novel metric learning approach for trees which we call embedding edit distance learning (BEDL) and which learns an edit distance indirectly by embedding the tree nodes as vectors, such that the Euclidean distance between those vectors supports class discrimination.

410, TITLE: Reinforcement Learning with Function-Valued Action Spaces for Partial Differential Equation Control

<http://proceedings.mlr.press/v80/pan18a.html>

AUTHORS: Yangchen Pan, Amir-massoud Farahmand, Martha White, Saleh Nabi, Piyush Grover, Daniel Nikovski

HIGHLIGHT: In particular, we propose the concept of action descriptors, which encode regularities among spatially-extended action dimensions and enable the agent to control high-dimensional action PDEs.

411, TITLE: Learning to Speed Up Structured Output Prediction

<http://proceedings.mlr.press/v80/pan18b.html>

AUTHORS: Xingyuan Pan, Vivek Srikumar

HIGHLIGHT: In this paper, we focus on reducing the prediction time of a trained black-box structured classifier without losing accuracy.

412, TITLE: Theoretical Analysis of Image-to-Image Translation with Adversarial Learning

<http://proceedings.mlr.press/v80/pan18c.html>

AUTHORS: Xudong Pan, Mi Zhang, Daizong Ding

HIGHLIGHT: In this paper, we reformulate their model from a brand-new geometrical perspective and have eventually reached a full interpretation on some interesting but unclear empirical phenomenons from their experiments.

413, TITLE: Max-Mahalanobis Linear Discriminant Analysis Networks

<http://proceedings.mlr.press/v80/pang18a.html>

AUTHORS: Tianyu Pang, Chao Du, Jun Zhu

HIGHLIGHT: In this paper, we show that a properly designed classifier can improve robustness to adversarial attacks and lead to better prediction results.

414, TITLE: Stochastic Variance-Reduced Policy Gradient

<http://proceedings.mlr.press/v80/papini18a.html>

AUTHORS: Matteo Papini, Damiano Binaghi, Giuseppe Canonaco, Matteo Pirota, Marcello Restelli

HIGHLIGHT: In this paper, we propose a novel reinforcement-learning algorithm consisting in a stochastic variance-reduced version of policy gradient for solving Markov Decision Processes (MDPs).

415, TITLE: Learning Independent Causal Mechanisms

<http://proceedings.mlr.press/v80/parascandolo18a.html>

AUTHORS: Giambattista Parascandolo, Niki Kilbertus, Mateo Rojas-Carulla, Bernhard Schölkopf

HIGHLIGHT: We develop an algorithm to recover a set of independent (inverse) mechanisms from a set of transformed data points.

416, TITLE: Time Limits in Reinforcement Learning

<http://proceedings.mlr.press/v80/pardo18a.html>

AUTHORS: Fabio Pardo, Arash Tavakoli, Vitaly Levnik, Petar Kormushev

**HIGHLIGHT:** In this paper, we provide a formal account for how time limits could effectively be handled in each of the two cases and explain why not doing so can cause state-aliasing and invalidation of experience replay, leading to suboptimal policies and training instability.

417, **TITLE:** Image Transformer  
<http://proceedings.mlr.press/v80/parmar18a.html>  
**AUTHORS:** Niki Parmar, Ashish Vaswani, Jakob Uszkoreit, Lukasz Kaiser, Noam Shazeer, Alexander Ku, Dustin Tran  
**HIGHLIGHT:** In this work, we generalize a recently proposed model architecture based on self-attention, the Transformer, to a sequence modeling formulation of image generation with a tractable likelihood.

418, **TITLE:** PIPPS: Flexible Model-Based Policy Search Robust to the Curse of Chaos  
<http://proceedings.mlr.press/v80/parmas18a.html>  
**AUTHORS:** Paavo Parmas, Carl Edward Rasmussen, Jan Peters, Kenji Doya  
**HIGHLIGHT:** Using our insights, we develop a model-based policy search framework, Probabilistic Inference for Particle-Based Policy Search (PIPPS), which is easily extensible, and allows for almost arbitrary models and policies, while simultaneously matching the performance of previous data-efficient learning algorithms.

419, **TITLE:** High-Quality Prediction Intervals for Deep Learning: A Distribution-Free, Ensembled Approach  
<http://proceedings.mlr.press/v80/pearce18a.html>  
**AUTHORS:** Tim Pearce, Alexandra Brintrup, Mohamed Zaki, Andy Neely  
**HIGHLIGHT:** This paper considers the generation of prediction intervals (PIs) by neural networks for quantifying uncertainty in regression tasks.

420, **TITLE:** Adaptive Three Operator Splitting  
<http://proceedings.mlr.press/v80/pedregosa18a.html>  
**AUTHORS:** Fabian Pedregosa, Gauthier Gidel  
**HIGHLIGHT:** We propose and analyze a novel adaptive step size variant of the Davis-Yin three operator splitting, a method that can solve optimization problems composed of a sum of a smooth term for which we have access to its gradient and an arbitrary number of potentially non-smooth terms for which we have access to their proximal operator.

421, **TITLE:** Efficient Neural Architecture Search via Parameters Sharing  
<http://proceedings.mlr.press/v80/pham18a.html>  
**AUTHORS:** Hieu Pham, Melody Guan, Barret Zoph, Quoc Le, Jeff Dean  
**HIGHLIGHT:** We propose Efficient Neural Architecture Search (ENAS), a fast and inexpensive approach for automatic model design.

422, **TITLE:** Bandits with Delayed, Aggregated Anonymous Feedback  
<http://proceedings.mlr.press/v80/pike-burke18a.html>  
**AUTHORS:** Ciara Pike-Burke, Shipra Agrawal, Csaba Szepesvari, Steffen Grunewalder  
**HIGHLIGHT:** In this paper, we show that this additive regret increase can be maintained in the harder delayed, aggregated anonymous feedback setting when the expected delay (or a bound on it) is known.

423, **TITLE:** Constant-Time Predictive Distributions for Gaussian Processes  
<http://proceedings.mlr.press/v80/pleiss18a.html>  
**AUTHORS:** Geoff Pleiss, Jacob Gardner, Kilian Weinberger, Andrew Gordon Wilson  
**HIGHLIGHT:** In this paper we address these shortcomings by using the Lanczos algorithm to rapidly approximate the predictive covariance matrix.

424, **TITLE:** Local Convergence Properties of SAGA/Prox-SVRG and Acceleration  
<http://proceedings.mlr.press/v80/poon18a.html>  
**AUTHORS:** Clarice Poon, Jingwei Liang, Carola Schoenlieb

**HIGHLIGHT:** In this paper, we present a local convergence analysis for a class of stochastic optimisation methods: the proximal variance reduced stochastic gradient methods, and mainly focus on SAGA (Defazio et al., 2014) and Prox-SVRG (Xiao & Zhang, 2014).

425, **TITLE:** Equivalence of Multicategory SVM and Simplex Cone SVM: Fast Computations and Statistical Theory

<http://proceedings.mlr.press/v80/pouliot18a.html>

**AUTHORS:** Guillaume Pouliot

**HIGHLIGHT:** We show that MSVM and SCSVM are in fact exactly equivalent, and provide a bijection between their tuning parameters.

426, **TITLE:** Learning Dynamics of Linear Denoising Autoencoders

<http://proceedings.mlr.press/v80/pretorius18a.html>

**AUTHORS:** Arnu Pretorius, Steve Kroon, Herman Kamper

**HIGHLIGHT:** Here we develop theory for how noise influences learning in DAEs.

427, **TITLE:** JointGAN: Multi-Domain Joint Distribution Learning with Generative Adversarial Nets

<http://proceedings.mlr.press/v80/pu18a.html>

**AUTHORS:** Yunchen Pu, Shuyang Dai, Zhe Gan, Weiyao Wang, Guoyin Wang, Yizhe Zhang, Ricardo Henao, Lawrence Carin Duke

**HIGHLIGHT:** A new generative adversarial network is developed for joint distribution matching. Distinct from most existing approaches, that only learn conditional distributions, the proposed model aims to learn a joint distribution of multiple random variables (domains).

428, **TITLE:** Selecting Representative Examples for Program Synthesis

<http://proceedings.mlr.press/v80/pu18b.html>

**AUTHORS:** Yewen Pu, Zachery Miranda, Armando Solar-Lezama, Leslie Kaelbling

**HIGHLIGHT:** We describe a method to discover a subset of examples that is both small and representative: the subset is constructed iteratively, using a neural network to predict the probability of unchosen examples conditioned on the chosen examples in the subset, and greedily adding the least probable example.

429, **TITLE:** Generalized Earley Parser: Bridging Symbolic Grammars and Sequence Data for Future Prediction

<http://proceedings.mlr.press/v80/qi18a.html>

**AUTHORS:** Siyuan Qi, Baoxiong Jia, Song-Chun Zhu

**HIGHLIGHT:** In this paper, we generalize the Earley parser to parse sequence data which is neither segmented nor labeled.

430, **TITLE:** Do Outliers Ruin Collaboration?

<http://proceedings.mlr.press/v80/qiao18a.html>

**AUTHORS:** Mingda Qiao

**HIGHLIGHT:** We present an algorithm that achieves an  $O(\eta n + \ln n)$  overhead, which is proved to be worst-case optimal.

431, **TITLE:** Gradually Updated Neural Networks for Large-Scale Image Recognition

<http://proceedings.mlr.press/v80/qiao18b.html>

**AUTHORS:** Siyuan Qiao, Zhishuai Zhang, Wei Shen, Bo Wang, Alan Yuille

**HIGHLIGHT:** In this paper, we present an alternative method to increase the depth.

432, **TITLE:** DCFNet: Deep Neural Network with Decomposed Convolutional Filters

<http://proceedings.mlr.press/v80/qiu18a.html>

**AUTHORS:** Qiang Qiu, Xiuyuan Cheng, Calderbank, Guillermo Sapiro

**HIGHLIGHT:** In this paper, we suggest to decompose convolutional filters in CNN as a truncated expansion with pre-fixed bases, namely the Decomposed Convolutional Filters network (DCFNet), where the expansion coefficients remain learned from data.

433, TITLE: Non-convex Conditional Gradient Sliding  
<http://proceedings.mlr.press/v80/qu18a.html>  
AUTHORS: Chao Qu, Yan Li, Huan Xu  
HIGHLIGHT: In this paper, we propose the non-convex conditional gradient sliding (NCGS) methods and analyze their convergence properties.

434, TITLE: Machine Theory of Mind  
<http://proceedings.mlr.press/v80/rabinowitz18a.html>  
AUTHORS: Neil Rabinowitz, Frank Perbet, Francis Song, Chiyuan Zhang, S. M. Ali Eslami, Matthew Botvinick  
HIGHLIGHT: We design a Theory of Mind neural network  $\{-\}$  a ToMnet  $\{-\}$  which uses meta-learning to build such models of the agents it encounters.

435, TITLE: Fast Parametric Learning with Activation Memorization  
<http://proceedings.mlr.press/v80/rae18a.html>  
AUTHORS: Jack Rae, Chris Dyer, Peter Dayan, Timothy Lillicrap  
HIGHLIGHT: We explore a simplified architecture where we treat a subset of the model parameters as fast memory stores.

436, TITLE: Can Deep Reinforcement Learning Solve Erdos-Selfridge-Spencer Games?  
<http://proceedings.mlr.press/v80/raghu18a.html>  
AUTHORS: Maithra Raghu, Alex Irpan, Jacob Andreas, Bobby Kleinberg, Quoc Le, Jon Kleinberg  
HIGHLIGHT: Here we consider a family of combinatorial games, arising from work of Erdos, Selfridge, and Spencer, and we propose their use as environments for evaluating and comparing different approaches to reinforcement learning.

437, TITLE: Cut-Pursuit Algorithm for Regularizing Nonsmooth Functionals with Graph Total Variation  
<http://proceedings.mlr.press/v80/raguet18a.html>  
AUTHORS: Hugo Raguet, Loic Landrieu  
HIGHLIGHT: We present an extension of the cut-pursuit algorithm, introduced by Landrieu and Obozinski (2017), to the graph total-variation regularization of functions with a separable nondifferentiable part.

438, TITLE: Modeling Others using Oneself in Multi-Agent Reinforcement Learning  
<http://proceedings.mlr.press/v80/raileanu18a.html>  
AUTHORS: Roberta Raileanu, Emily Denton, Arthur Szlam, Rob Fergus  
HIGHLIGHT: We propose a new approach for learning in these domains: Self Other-Modeling (SOM), in which an agent uses its own policy to predict the other agent's actions and update its belief of their hidden goal in an online manner.

439, TITLE: On Nesting Monte Carlo Estimators  
<http://proceedings.mlr.press/v80/rainforth18a.html>  
AUTHORS: Tom Rainforth, Rob Cornish, Hongseok Yang, Andrew Warrington, Frank Wood  
HIGHLIGHT: We investigate the statistical implications of nesting MC estimators, including cases of multiple levels of nesting, and establish the conditions under which they converge.

440, TITLE: Tighter Variational Bounds are Not Necessarily Better  
<http://proceedings.mlr.press/v80/rainforth18b.html>  
AUTHORS: Tom Rainforth, Adam Kosiorek, Tuan Anh Le, Chris Maddison, Maximilian Igl, Frank Wood, Yee Whye Teh  
HIGHLIGHT: Based on our insights, we introduce three new algorithms: the partially importance weighted auto-encoder (PIWAE), the multiply importance weighted auto-encoder (MIWAE), and the combination importance weighted auto-encoder (CIWAE), each of which includes the standard importance weighted auto-encoder (IWAE) as a special case.

- 441, TITLE: SAFFRON: an Adaptive Algorithm for Online Control of the False Discovery Rate  
<http://proceedings.mlr.press/v80/ramdas18a.html>  
AUTHORS: Aaditya Ramdas, Tijana Zrnic, Martin Wainwright, Michael Jordan  
HIGHLIGHT: In this work, we present a powerful new framework for online FDR control that we refer to as “SAFFRON”.
- 442, TITLE: QMIX: Monotonic Value Function Factorisation for Deep Multi-Agent Reinforcement Learning  
<http://proceedings.mlr.press/v80/rashid18a.html>  
AUTHORS: Tabish Rashid, Mikayel Samvelyan, Christian Schroeder, Gregory Farquhar, Jakob Foerster, Shimon Whiteson  
HIGHLIGHT: We evaluate QMIX on a challenging set of StarCraft II micromanagement tasks, and show that QMIX significantly outperforms existing value-based multi-agent reinforcement learning methods.
- 443, TITLE: Gradient Coding from Cyclic MDS Codes and Expander Graphs  
<http://proceedings.mlr.press/v80/raviv18a.html>  
AUTHORS: Netanel Raviv, Rashish Tandon, Alex Dimakis, Itzhak Tamo  
HIGHLIGHT: In this paper we design novel gradient codes using tools from classical coding theory, namely, cyclic MDS codes, which compare favourably with existing solutions, both in the applicable range of parameters and in the complexity of the involved algorithms.
- 444, TITLE: Learning Implicit Generative Models with the Method of Learned Moments  
<http://proceedings.mlr.press/v80/ravuri18a.html>  
AUTHORS: Suman Ravuri, Shakir Mohamed, Mihaela Rosca, Oriol Vinyals  
HIGHLIGHT: We propose a method of moments (MoM) algorithm for training large-scale implicit generative models.
- 445, TITLE: Weightless: Lossy weight encoding for deep neural network compression  
<http://proceedings.mlr.press/v80/reagan18a.html>  
AUTHORS: Brandon Reagan, Udit Gupta, Bob Adolf, Michael Mitzenmacher, Alexander Rush, Gu-Yeon Wei, David Brooks  
HIGHLIGHT: In this paper, we present a novel scheme for lossy weight encoding co-designed with weight simplification techniques.
- 446, TITLE: Learning to Reweight Examples for Robust Deep Learning  
<http://proceedings.mlr.press/v80/ren18a.html>  
AUTHORS: Mengye Ren, Wenyuan Zeng, Bin Yang, Raquel Urtasun  
HIGHLIGHT: In contrast to past reweighting methods, which typically consist of functions of the cost value of each example, in this work we propose a novel meta-learning algorithm that learns to assign weights to training examples based on their gradient directions.
- 447, TITLE: Learning by Playing Solving Sparse Reward Tasks from Scratch  
<http://proceedings.mlr.press/v80/riedmiller18a.html>  
AUTHORS: Martin Riedmiller, Roland Hafner, Thomas Lampe, Michael Neunert, Jonas Degraeve, Tom Wiele, Vlad Mnih, Nicolas Heess, Jost Tobias Springenberg  
HIGHLIGHT: We propose Scheduled Auxiliary Control (SAC-X), a new learning paradigm in the context of Reinforcement Learning (RL).
- 448, TITLE: Been There, Done That: Meta-Learning with Episodic Recall  
<http://proceedings.mlr.press/v80/ritter18a.html>  
AUTHORS: Samuel Ritter, Jane Wang, Zeb Kurth-Nelson, Siddhant Jayakumar, Charles Blundell, Razvan Pascanu, Matthew Botvinick  
HIGHLIGHT: We propose a formalism for generating open-ended yet repetitious environments, then develop a meta-learning architecture for solving these environments.

- 449, TITLE: A Hierarchical Latent Vector Model for Learning Long-Term Structure in Music  
<http://proceedings.mlr.press/v80/roberts18a.html>  
AUTHORS: Adam Roberts, Jesse Engel, Colin Raffel, Curtis Hawthorne, Douglas Eck  
HIGHLIGHT: To address this issue, we propose the use of a hierarchical decoder, which first outputs embeddings for subsequences of the input and then uses these embeddings to generate each subsequence independently.
- 450, TITLE: Learning to Optimize Combinatorial Functions  
<http://proceedings.mlr.press/v80/rosenfeld18a.html>  
AUTHORS: Nir Rosenfeld, Eric Balkanski, Amir Globerson, Yaron Singer  
HIGHLIGHT: Our goal in this paper is to highlight the source of this hardness, and propose an alternative criterion for optimizing general combinatorial functions from sampled data.
- 451, TITLE: Fast Information-theoretic Bayesian Optimisation  
<http://proceedings.mlr.press/v80/ru18a.html>  
AUTHORS: Binxin Ru, Michael A. Osborne, Mark Mcleod, Diego Granziol  
HIGHLIGHT: We develop a fast information-theoretic Bayesian Optimisation method, FITBO, that avoids the need for sampling the global minimiser, thus significantly reducing computational overhead.
- 452, TITLE: Deep One-Class Classification  
<http://proceedings.mlr.press/v80/ruffl8a.html>  
AUTHORS: Lukas Ruff, Robert Vandermeulen, Nico Goernitz, Lucas Deecke, Shoaib Ahmed Siddiqui, Alexander Binder, Emmanuel M?ller, Marius Kloft  
HIGHLIGHT: In this paper we introduce a new anomaly detection method—Deep Support Vector Data Description—, which is trained on an anomaly detection based objective.
- 453, TITLE: Augment and Reduce: Stochastic Inference for Large Categorical Distributions  
<http://proceedings.mlr.press/v80/ruiz18a.html>  
AUTHORS: Francisco Ruiz, Michalis Titsias, Adji Bousso Dieng, David Blei  
HIGHLIGHT: To address this problem, we propose augment and reduce (A&R), a method to alleviate the computational complexity.
- 454, TITLE: Probabilistic Boolean Tensor Decomposition  
<http://proceedings.mlr.press/v80/rukat18a.html>  
AUTHORS: Tammo Rukat, Chris Holmes, Christopher Yau  
HIGHLIGHT: Here, we present its first probabilistic treatment.
- 455, TITLE: Black-Box Variational Inference for Stochastic Differential Equations  
<http://proceedings.mlr.press/v80/ryder18a.html>  
AUTHORS: Tom Ryder, Andrew Golightly, A. Stephen McGough, Dennis Prangle  
HIGHLIGHT: We use a standard mean-field variational approximation of the parameter posterior, and introduce a recurrent neural network to approximate the posterior for the diffusion paths conditional on the parameters.
- 456, TITLE: Spurious Local Minima are Common in Two-Layer ReLU Neural Networks  
<http://proceedings.mlr.press/v80/safran18a.html>  
AUTHORS: Itay Safran, Ohad Shamir  
HIGHLIGHT: We consider the optimization problem associated with training simple ReLU neural networks of the form  $\mathbf{x} \mapsto \sum_{i=1}^k \max\{0, \mathbf{w}_i^\top \mathbf{x}\}$  with respect to the squared loss.
- 457, TITLE: Learning Equations for Extrapolation and Control  
<http://proceedings.mlr.press/v80/sahoo18a.html>  
AUTHORS: Subham Sahoo, Christoph Lampert, Georg Martius

**HIGHLIGHT:** We present an approach to identify concise equations from data using a shallow neural network approach.

458, **TITLE:** Tempered Adversarial Networks

<http://proceedings.mlr.press/v80/sajjadi18a.html>

**AUTHORS:** Mehdi S. M. Sajjadi, Giambattista Parascandolo, Arash Mehrjou, Bernhard Schölkopf

**HIGHLIGHT:** We propose a simple modification that gives the generator control over the real samples which leads to a tempered learning process for both generator and discriminator.

459, **TITLE:** Representation Tradeoffs for Hyperbolic Embeddings

<http://proceedings.mlr.press/v80/sala18a.html>

**AUTHORS:** Frederic Sala, Chris De Sa, Albert Gu, Christopher Re

**HIGHLIGHT:** To embed general metric spaces, we propose a hyperbolic generalization of multidimensional scaling (h-MDS).

460, **TITLE:** Graph Networks as Learnable Physics Engines for Inference and Control

<http://proceedings.mlr.press/v80/sanchez-gonzalez18a.html>

**AUTHORS:** Alvaro Sanchez-Gonzalez, Nicolas Heess, Jost Tobias Springenberg, Josh Merel, Martin Riedmiller, Raia Hadsell, Peter Battaglia

**HIGHLIGHT:** Here we introduce a new class of learnable models—based on graph networks—which implement an inductive bias for object- and relation-centric representations of complex, dynamical systems.

461, **TITLE:** A Classification-Based Study of Covariate Shift in GAN Distributions

<http://proceedings.mlr.press/v80/santurkar18a.html>

**AUTHORS:** Shibani Santurkar, Ludwig Schmidt, Aleksander Madry

**HIGHLIGHT:** In this paper, we develop quantitative and scalable tools for assessing the diversity of GAN distributions.

462, **TITLE:** TAPAS: Tricks to Accelerate (encrypted) Prediction As a Service

<http://proceedings.mlr.press/v80/sanyal18a.html>

**AUTHORS:** Amartya Sanyal, Matt Kusner, Adria Gascon, Varun Kanade

**HIGHLIGHT:** Machine learning methods are widely used for a variety of prediction problems.

463, **TITLE:** Tight Regret Bounds for Bayesian Optimization in One Dimension

<http://proceedings.mlr.press/v80/scarlett18a.html>

**AUTHORS:** Jonathan Scarlett

**HIGHLIGHT:** We consider the problem of Bayesian optimization (BO) in one dimension, under a Gaussian process prior and Gaussian sampling noise.

464, **TITLE:** Learning with Abandonment

<http://proceedings.mlr.press/v80/schmit18a.html>

**AUTHORS:** Sven Schmit, Ramesh Johari

**HIGHLIGHT:** We propose a general thresholded learning model for scenarios like this, and discuss the structure of optimal policies.

465, **TITLE:** Not to Cry Wolf: Distantly Supervised Multitask Learning in Critical Care

<http://proceedings.mlr.press/v80/schwab18a.html>

**AUTHORS:** Patrick Schwab, Emanuela Keller, Carl Muroi, David J. Mack, Christian Strussle, Walter Karlen

**HIGHLIGHT:** We frame the problem of false alarm reduction from multivariate time series as a machine-learning task and address it with a novel multitask network architecture that utilises distant supervision through multiple related auxiliary tasks in order to reduce the number of expensive labels required for training.

466, **TITLE:** Progress & Compress: A scalable framework for continual learning

<http://proceedings.mlr.press/v80/schwarz18a.html>

AUTHORS: Jonathan Schwarz, Wojciech Czarnecki, Jelena Luketina, Agnieszka Grabska-Barwinska, Yee Whye Teh, Razvan Pascanu, Raia Hadsell

HIGHLIGHT: We introduce a conceptually simple and scalable framework for continual learning domains where tasks are learned sequentially.

467, TITLE: Multi-Fidelity Black-Box Optimization with Hierarchical Partitions

<http://proceedings.mlr.press/v80/sen18a.html>

AUTHORS: Rajat Sen, Kirthevasan Kandasamy, Sanjay Shakkottai

HIGHLIGHT: Motivated by settings such as hyper-parameter tuning and physical simulations, we consider the problem of black-box optimization of a function.

468, TITLE: Overcoming Catastrophic Forgetting with Hard Attention to the Task

<http://proceedings.mlr.press/v80/serra18a.html>

AUTHORS: Joan Serra, Didac Suris, Marius Miron, Alexandros Karatzoglou

HIGHLIGHT: In this paper, we propose a task-based hard attention mechanism that preserves previous tasks' information without affecting the current task's learning.

469, TITLE: Bounding and Counting Linear Regions of Deep Neural Networks

<http://proceedings.mlr.press/v80/serra18b.html>

AUTHORS: Thiago Serra, Christian Tjandraatmadja, Srikumar Ramalingam

HIGHLIGHT: We present (i) tighter upper and lower bounds for the maximum number of linear regions on rectifier networks, which are exact for inputs of dimension one; (ii) a first upper bound for multi-layer maxout networks; and (iii) a first method to perform exact enumeration or counting of the number of regions by modeling the DNN with a mixed-integer linear formulation.

470, TITLE: First Order Generative Adversarial Networks

<http://proceedings.mlr.press/v80/seward18a.html>

AUTHORS: Calvin Seward, Thomas Unterthiner, Urs Bergmann, Nikolay Jetchev, Sepp Hochreiter

HIGHLIGHT: To formally describe an optimal update direction, we introduce a theoretical framework which allows the derivation of requirements on both the divergence and corresponding method for determining an update direction, with these requirements guaranteeing unbiased mini-batch updates in the direction of steepest descent.

471, TITLE: Finding Influential Training Samples for Gradient Boosted Decision Trees

<http://proceedings.mlr.press/v80/sharchilev18a.html>

AUTHORS: Boris Sharchilev, Yury Ustinovskiy, Pavel Serdyukov, Maarten Rijke

HIGHLIGHT: We propose several ways of extending this framework to non-parametric GBDT ensembles under the assumption that tree structures remain fixed.

472, TITLE: Solving Partial Assignment Problems using Random Clique Complexes

<http://proceedings.mlr.press/v80/sharma18a.html>

AUTHORS: Charu Sharma, Deepak Nathani, Manohar Kaul

HIGHLIGHT: We present an alternate formulation of the partial assignment problem as matching random clique complexes, that are higher-order analogues of random graphs, designed to provide a set of invariants that better detect higher-order structure.

473, TITLE: Adafactor: Adaptive Learning Rates with Sublinear Memory Cost

<http://proceedings.mlr.press/v80/shazeer18a.html>

AUTHORS: Noam Shazeer, Mitchell Stern

HIGHLIGHT: We propose update clipping and a gradually increasing decay rate scheme as remedies.

474, TITLE: Locally Private Hypothesis Testing

<http://proceedings.mlr.press/v80/sheffet18a.html>

AUTHORS: Or Sheffet



**HIGHLIGHT:** We initiate the study of differentially private hypothesis testing in the local-model, under both the standard (symmetric) randomized-response mechanism (Warner 1965, Kasiviswanathan et al, 2008) and the newer (non-symmetric) mechanisms (Bassily & Smith, 2015, Bassily et al, 2017).

475, **TITLE:** Learning in Integer Latent Variable Models with Nested Automatic Differentiation  
<http://proceedings.mlr.press/v80/sheldon18a.html>

**AUTHORS:** Daniel Sheldon, Kevin Winner, Debora Sujono

**HIGHLIGHT:** We develop nested automatic differentiation (AD) algorithms for exact inference and learning in integer latent variable models.

476, **TITLE:** Towards More Efficient Stochastic Decentralized Learning: Faster Convergence and Sparse Communication

<http://proceedings.mlr.press/v80/shen18a.html>

**AUTHORS:** Zebang Shen, Aryan Mokhtari, Tengfei Zhou, Peilin Zhao, Hui Qian

**HIGHLIGHT:** In this paper, we generalize the decentralized optimization problem to a monotone operator root finding problem, and propose a stochastic algorithm named DSBA that (1) converges geometrically with a rate linearly depending on the problem condition number, and (2) can be implemented using sparse communication only.

477, **TITLE:** An Algorithmic Framework of Variable Metric Over-Relaxed Hybrid Proximal Extra-Gradient Method

<http://proceedings.mlr.press/v80/shen18b.html>

**AUTHORS:** Li Shen, Peng Sun, Yitong Wang, Wei Liu, Tong Zhang

**HIGHLIGHT:** We propose a novel algorithmic framework of Variable Metric Over-Relaxed Hybrid Proximal Extra-gradient (VMOR-HPE) method with a global convergence guarantee for the maximal monotone operator inclusion problem.

478, **TITLE:** A Spectral Approach to Gradient Estimation for Implicit Distributions

<http://proceedings.mlr.press/v80/shi18a.html>

**AUTHORS:** Jiaxin Shi, Shengyang Sun, Jun Zhu

**HIGHLIGHT:** To this end, we develop a gradient estimator for implicit distributions based on Stein's identity and a spectral decomposition of kernel operators, where the eigenfunctions are approximated by the Nyström method.

479, **TITLE:** TACO: Learning Task Decomposition via Temporal Alignment for Control

<http://proceedings.mlr.press/v80/shiarlis18a.html>

**AUTHORS:** Kyriacos Shiarlis, Markus Wulfmeier, Sasha Salter, Shimon Whiteson, Ingmar Posner

**HIGHLIGHT:** In contrast, we propose a weakly supervised, domain-agnostic approach based on task sketches, which include only the sequence of sub-tasks performed in each demonstration. By reusing the corresponding sub-policies within and between tasks, we can provide training data for each policy from different high-level tasks and compose them to perform novel ones.

480, **TITLE:** CRAFTML, an Efficient Clustering-based Random Forest for Extreme Multi-label Learning

<http://proceedings.mlr.press/v80/siblini18a.html>

**AUTHORS:** Wissam Siblini, Pascale Kuntz, Frank Meyer

**HIGHLIGHT:** To overcome these limits, we here introduce a new random forest based algorithm with a very fast partitioning approach called CRAFTML.

481, **TITLE:** Asynchronous Stochastic Quasi-Newton MCMC for Non-Convex Optimization

<http://proceedings.mlr.press/v80/simsekli18a.html>

**AUTHORS:** Umut Simsekli, Cagatay Yildiz, Than Huy Nguyen, Taylan Cemgil, Gael Richard

**HIGHLIGHT:** By building up on this recent theory, in this study, we develop an asynchronous-parallel stochastic L-BFGS algorithm for non-convex optimization.

482, **TITLE:** K-means clustering using random matrix sparsification

<http://proceedings.mlr.press/v80/sinha18a.html>

- AUTHORS: Kaushik Sinha  
HIGHLIGHT: In this paper we show that we can randomly sparsify the original data matrix resulting in a sparse data matrix which can significantly speed up the above mentioned matrix vector multiplication step without significantly affecting cluster quality.
- 483, TITLE: Towards End-to-End Prosody Transfer for Expressive Speech Synthesis with Tacotron  
<http://proceedings.mlr.press/v80/skerry-ryan18a.html>  
AUTHORS: RJ Skerry-Ryan, Eric Battenberg, Ying Xiao, Yuxuan Wang, Daisy Stanton, Joel Shor, Ron Weiss, Rob Clark, Rif A. Saurous  
HIGHLIGHT: We present an extension to the Tacotron speech synthesis architecture that learns a latent embedding space of prosody, derived from a reference acoustic representation containing the desired prosody.
- 484, TITLE: An Inference-Based Policy Gradient Method for Learning Options  
<http://proceedings.mlr.press/v80/smith18a.html>  
AUTHORS: Matthew Smith, Herke Hoof, Joelle Pineau  
HIGHLIGHT: In this work we develop a novel policy gradient method for the automatic learning of policies with options.
- 485, TITLE: Accelerating Natural Gradient with Higher-Order Invariance  
<http://proceedings.mlr.press/v80/song18a.html>  
AUTHORS: Yang Song, Jiaming Song, Stefano Ermon  
HIGHLIGHT: In this paper, we study invariance properties from a combined perspective of Riemannian geometry and numerical differential equation solving.
- 486, TITLE: Knowledge Transfer with Jacobian Matching  
<http://proceedings.mlr.press/v80/srinivas18a.html>  
AUTHORS: Suraj Srinivas, Francois Fleuret  
HIGHLIGHT: Knowledge Transfer with Jacobian Matching
- 487, TITLE: Universal Planning Networks: Learning Generalizable Representations for Visuomotor Control  
<http://proceedings.mlr.press/v80/srinivas18b.html>  
AUTHORS: Aravind Srinivas, Allan Jabri, Pieter Abbeel, Sergey Levine, Chelsea Finn  
HIGHLIGHT: To this end, we introduce universal planning networks (UPN).
- 488, TITLE: Structured Control Nets for Deep Reinforcement Learning  
<http://proceedings.mlr.press/v80/srouji18a.html>  
AUTHORS: Mario Srouji, Jian Zhang, Ruslan Salakhutdinov  
HIGHLIGHT: In this work, we propose a new neural network architecture for the policy network representation that is simple yet effective.
- 489, TITLE: Approximation Algorithms for Cascading Prediction Models  
<http://proceedings.mlr.press/v80/streeter18a.html>  
AUTHORS: Matthew Streeter  
HIGHLIGHT: We present an approximation algorithm that takes a pool of pre-trained models as input and produces from it a cascaded model with similar accuracy but lower average-case cost.
- 490, TITLE: Learning Low-Dimensional Temporal Representations  
<http://proceedings.mlr.press/v80/su18a.html>  
AUTHORS: Bing Su, Ying Wu  
HIGHLIGHT: We propose a DR method called LT-LDA to learn low-dimensional temporal representations.
- 491, TITLE: Exploiting the Potential of Standard Convolutional Autoencoders for Image Restoration by Evolutionary Search

<http://proceedings.mlr.press/v80/suganuma18a.html>

AUTHORS: Masanori Suganuma, Mete Ozay, Takayuki Okatani

HIGHLIGHT: Researchers have applied deep neural networks to image restoration tasks, in which they proposed various network architectures, loss functions, and training methods.

492, TITLE: Stagewise Safe Bayesian Optimization with Gaussian Processes

<http://proceedings.mlr.press/v80/sui18a.html>

AUTHORS: Yanan Sui, Zhuang, Joel Burdick, Yisong Yue

HIGHLIGHT: We consider this problem of optimizing an unknown utility function with absolute feedback or preference feedback subject to unknown safety constraints.

493, TITLE: Neural Program Synthesis from Diverse Demonstration Videos

<http://proceedings.mlr.press/v80/sun18a.html>

AUTHORS: Shao-Hua Sun, Hyeonwoo Noh, Sriram Somasundaram, Joseph Lim

HIGHLIGHT: To empower machines with this ability, we propose a neural program synthesizer that is able to explicitly synthesize underlying programs from behaviorally diverse and visually complicated demonstration videos.

494, TITLE: Scalable approximate Bayesian inference for particle tracking data

<http://proceedings.mlr.press/v80/sun18b.html>

AUTHORS: Ruoxi Sun, Liam Paninski

HIGHLIGHT: Here we propose a flexible and scalable amortized approach for Bayesian inference on this task.

495, TITLE: Graphical Nonconvex Optimization via an Adaptive Convex Relaxation

<http://proceedings.mlr.press/v80/sun18c.html>

AUTHORS: Qiang Sun, Kean Ming Tan, Han Liu, Tong Zhang

HIGHLIGHT: In this paper, we propose the graphical nonconvex optimization for optimal estimation in Gaussian graphical models, which is then approximated by a sequence of convex programs.

496, TITLE: Convolutional Imputation of Matrix Networks

<http://proceedings.mlr.press/v80/sun18d.html>

AUTHORS: Qingyun Sun, Mengyuan Yan, David Donoho, Boyd

HIGHLIGHT: To recover the matrix network, we propose a structural assumption that the matrices are low-rank after the graph Fourier transform on the network.

497, TITLE: Differentiable Compositional Kernel Learning for Gaussian Processes

<http://proceedings.mlr.press/v80/sun18e.html>

AUTHORS: Shengyang Sun, Guodong Zhang, Chaoqi Wang, Wenyuan Zeng, Jiaman Li, Roger Grosse

HIGHLIGHT: We present the Neural Kernel Network (NKN), a flexible family of kernels represented by a neural network.

498, TITLE: Learning the Reward Function for a Misspecified Model

<http://proceedings.mlr.press/v80/talvitie18a.html>

AUTHORS: Erik Talvitie

HIGHLIGHT: This paper presents a novel error bound that accounts for the reward model's behavior in states sampled from the model.

499, TITLE:  $\mathcal{D}^2$ : Decentralized Training over Decentralized Data

<http://proceedings.mlr.press/v80/tang18a.html>

AUTHORS: Hanlin Tang, Xiangru Lian, Ming Yan, Ce Zhang, Ji Liu

HIGHLIGHT: In this paper, we ask the question: Can we design a decentralized parallel stochastic gradient descent algorithm that is less sensitive to the data variance across workers?

500, TITLE: Neural Inverse Rendering for General Reflectance Photometric Stereo

<http://proceedings.mlr.press/v80/taniai18a.html>

AUTHORS: Tatsunori Taniai, Takanori Maehara

HIGHLIGHT: We present a novel convolutional neural network architecture for photometric stereo (Woodham, 1980), a problem of recovering 3D object surface normals from multiple images observed under varying illuminations.

501, TITLE: Black Box FDR

<http://proceedings.mlr.press/v80/tansey18a.html>

AUTHORS: Wesley Tansey, Yixin Wang, David Blei, Raul Rabadan

HIGHLIGHT: We present Black Box FDR (BB-FDR), an empirical-Bayes method for analyzing multi-experiment studies when many covariates are gathered per experiment.

502, TITLE: Best Arm Identification in Linear Bandits with Linear Dimension Dependency

<http://proceedings.mlr.press/v80/tao18a.html>

AUTHORS: Chao Tao, Sa?l Blanco, Yuan Zhou

HIGHLIGHT: Using this estimator, we describe an algorithm whose sample complexity depends linearly on the dimension  $d$ , as well as an algorithm with sample complexity dependent on the reward gaps of the best  $d$  arms, matching the lower bound arising from the ordinary top-arm identification problem.

503, TITLE: Chi-square Generative Adversarial Network

<http://proceedings.mlr.press/v80/tao18b.html>

AUTHORS: Chenyang Tao, Liqun Chen, Ricardo Henao, Jianfeng Feng, Lawrence Carin Duke

HIGHLIGHT: We elucidate the theoretical connections between these three popular GAN training criteria and propose a novel procedure, called  $\chi^2$  (Chi-square) GAN, that is conceptually simple, stable at training and resistant to mode collapse.

504, TITLE: Lyapunov Functions for First-Order Methods: Tight Automated Convergence Guarantees

<http://proceedings.mlr.press/v80/taylor18a.html>

AUTHORS: Adrien Taylor, Bryan Van Scoy, Laurent Lessard

HIGHLIGHT: We present a novel way of generating Lyapunov functions for proving linear convergence rates of first-order optimization methods.

505, TITLE: Bayesian Uncertainty Estimation for Batch Normalized Deep Networks

<http://proceedings.mlr.press/v80/teye18a.html>

AUTHORS: Mattias Teye, Hossein Azizpour, Kevin Smith

HIGHLIGHT: We show that training a deep network using batch normalization is equivalent to approximate inference in Bayesian models.

506, TITLE: Decoupling Gradient-Like Learning Rules from Representations

<http://proceedings.mlr.press/v80/thomas18a.html>

AUTHORS: Philip Thomas, Christoph Dann, Emma Brunskill

HIGHLIGHT: In this paper we focus on gradient-like learning rules, wherein these two decisions are coupled in a subtle (and often unintentional) way.

507, TITLE: CoVeR: Learning Covariate-Specific Vector Representations with Tensor Decompositions

<http://proceedings.mlr.press/v80/tian18a.html>

AUTHORS: Kevin Tian, Teng Zhang, James Zou

HIGHLIGHT: We propose CoVeR, a new tensor decomposition model for vector embeddings with covariates.

508, TITLE: Importance Weighted Transfer of Samples in Reinforcement Learning

<http://proceedings.mlr.press/v80/tirinzone18a.html>

AUTHORS: Andrea Tirinzoni, Andrea Sessa, Matteo Pirota, Marcello Restelli

HIGHLIGHT: In this paper, we propose a model-based technique that automatically estimates the relevance (importance weight) of each source sample for solving the target task.

- 509, TITLE: Adversarial Regression with Multiple Learners  
<http://proceedings.mlr.press/v80/tong18a.html>  
AUTHORS: Liang Tong, Sixie Yu, Scott Alfeld, vorobeychik  
HIGHLIGHT: We present an algorithm for computing this equilibrium, and show through extensive experiments that equilibrium models are significantly more robust than conventional regularized linear regression.
- 510, TITLE: Convergent Tree Backup and Retrace with Function Approximation  
<http://proceedings.mlr.press/v80/touati18a.html>  
AUTHORS: Ahmed Touati, Pierre-Luc Bacon, Doina Precup, Pascal Vincent  
HIGHLIGHT: In this work, we show that the Tree Backup and Retrace algorithms are unstable with linear function approximation, both in theory and in practice with specific examples.
- 511, TITLE: Learning Longer-term Dependencies in RNNs with Auxiliary Losses  
<http://proceedings.mlr.press/v80/trinh18a.html>  
AUTHORS: Trieu Trinh, Andrew Dai, Thang Luong, Quoc Le  
HIGHLIGHT: This paper proposes a simple method that improves the ability to capture long term dependencies in RNNs by adding an unsupervised auxiliary loss to the original objective.
- 512, TITLE: Theoretical Analysis of Sparse Subspace Clustering with Missing Entries  
<http://proceedings.mlr.press/v80/tsakiris18a.html>  
AUTHORS: Manolis Tsakiris, Rene Vidal  
HIGHLIGHT: In this paper we give theoretical guarantees for SSC with incomplete data, and provide theoretical evidence that projecting the zero-filled data onto the observation pattern of the point being expressed can lead to substantial improvement in performance; a phenomenon already known experimentally.
- 513, TITLE: StrassenNets: Deep Learning with a Multiplication Budget  
<http://proceedings.mlr.press/v80/tschannen18a.html>  
AUTHORS: Michael Tschannen, Aran Khanna, Animashree Anandkumar  
HIGHLIGHT: We perform end-to-end learning of low-cost approximations of matrix multiplications in DNN layers by casting matrix multiplications as 2-layer sum-product networks (SPNs) (arithmetic circuits) and learning their (ternary) edge weights from data.
- 514, TITLE: Invariance of Weight Distributions in Rectified MLPs  
<http://proceedings.mlr.press/v80/tsuchida18a.html>  
AUTHORS: Russell Tsuchida, Fred Roosta, Marcus Gallagher  
HIGHLIGHT: We derive the equivalent kernels of MLPs with ReLU or Leaky ReLU activations for all rotationally-invariant weight distributions, generalizing a previous result that required Gaussian weight distributions.
- 515, TITLE: Least-Squares Temporal Difference Learning for the Linear Quadratic Regulator  
<http://proceedings.mlr.press/v80/tu18a.html>  
AUTHORS: Stephen Tu, Benjamin Recht  
HIGHLIGHT: We give the first finite-time analysis of the number of samples needed to estimate the value function for a fixed static state-feedback policy to within epsilon-relative error.
- 516, TITLE: The Mirage of Action-Dependent Baselines in Reinforcement Learning  
<http://proceedings.mlr.press/v80/tucker18a.html>  
AUTHORS: George Tucker, Surya Bhupatiraju, Shixiang Gu, Richard Turner, Zoubin Ghahramani, Sergey Levine  
HIGHLIGHT: We confirm this unexpected result by reviewing the open-source code accompanying these prior papers, and show that subtle implementation decisions cause deviations from the methods presented in the papers and explain the source of the previously observed empirical gains.
- 517, TITLE: Adversarial Risk and the Dangers of Evaluating Against Weak Attacks

<http://proceedings.mlr.press/v80/uesato18a.html>

AUTHORS: Jonathan Uesato, Brendan O'Donoghue, Pushmeet Kohli, Aaron Oord

HIGHLIGHT: This paper investigates recently proposed approaches for defending against adversarial examples and evaluating adversarial robustness.

518, TITLE: DVAE++: Discrete Variational Autoencoders with Overlapping Transformations

<http://proceedings.mlr.press/v80/vahdat18a.html>

AUTHORS: Arash Vahdat, William Macready, Zhengbing Bian, Amir Khoshaman, Evgeny Andriyash

HIGHLIGHT: We propose a new class of smoothing transformations based on a mixture of two overlapping distributions, and show that the proposed transformation can be used for training binary latent models with either directed or undirected priors.

519, TITLE: Programmatically Interpretable Reinforcement Learning

<http://proceedings.mlr.press/v80/verma18a.html>

AUTHORS: Abhinav Verma, Vijayaraghavan Murali, Rishabh Singh, Pushmeet Kohli, Swarat Chaudhuri

HIGHLIGHT: We propose a new method, called Neurally Directed Program Search (NDPS), for solving the challenging nonsmooth optimization problem of finding a programmatic policy with maximal reward.

520, TITLE: Clustering Semi-Random Mixtures of Gaussians

<http://proceedings.mlr.press/v80/vijayaraghavan18a.html>

AUTHORS: Aravindan Vijayaraghavan, Pranjal Awasthi

HIGHLIGHT: In this paper, we propose a natural robust model for k-means clustering that generalizes the Gaussian mixture model, and that we believe will be useful in identifying robust algorithms.

521, TITLE: A Probabilistic Theory of Supervised Similarity Learning for Pointwise ROC Curve Optimization

<http://proceedings.mlr.press/v80/vogel18a.html>

AUTHORS: Robin Vogel, Aurélien Bellet, Stéphane Clémens

HIGHLIGHT: In this paper, similarity learning is investigated from the perspective of pairwise bipartite ranking, where the goal is to rank the elements of a database by decreasing order of the probability that they share the same label with some query data point, based on the similarity scores.

522, TITLE: Hierarchical Multi-Label Classification Networks

<http://proceedings.mlr.press/v80/wehrmann18a.html>

AUTHORS: Jonas Wehrmann, Ricardo Cerri, Rodrigo Barros

HIGHLIGHT: In this paper, we propose novel neural network architectures for HMC called HMCN, capable of simultaneously optimizing local and global loss functions for discovering local hierarchical class-relationships and global information from the entire class hierarchy while penalizing hierarchical violations.

523, TITLE: Transfer Learning via Learning to Transfer

<http://proceedings.mlr.press/v80/wei18a.html>

AUTHORS: Ying WEI, Yu Zhang, Junzhou Huang, Qiang Yang

HIGHLIGHT: Motivated by this, we propose a novel transfer learning framework known as Learning to Transfer (L2T) to automatically determine what and how to transfer are the best by leveraging previous transfer learning experiences.

524, TITLE: Semi-Supervised Learning on Data Streams via Temporal Label Propagation

<http://proceedings.mlr.press/v80/wagner18a.html>

AUTHORS: Tal Wagner, Sudipto Guha, Shiva Kasiviswanathan, Nina Mishra

HIGHLIGHT: We present a semi-supervised learning algorithm for this task.

525, TITLE: Neural Dynamic Programming for Musical Self Similarity

<http://proceedings.mlr.press/v80/walder18a.html>

AUTHORS: Christian Walder, Dongwoo Kim

HIGHLIGHT: We present a neural sequence model designed specifically for symbolic music.

- 526, TITLE: Thompson Sampling for Combinatorial Semi-Bandits  
<http://proceedings.mlr.press/v80/wang18a.html>  
AUTHORS: Siwei Wang, Wei Chen  
HIGHLIGHT: We analyze the standard TS algorithm for the general CMAB, and obtain the first distribution-dependent regret bound of  $O(m \log T / \Delta_{\min})$  for TS under general CMAB, where  $m$  is the number of arms,  $T$  is the time horizon, and  $\Delta_{\min}$  is the minimum gap between the expected reward of the optimal solution and any non-optimal solution.
- 527, TITLE: PredRNN++: Towards A Resolution of the Deep-in-Time Dilemma in Spatiotemporal Predictive Learning  
<http://proceedings.mlr.press/v80/wang18b.html>  
AUTHORS: Yunbo Wang, Zhifeng Gao, Mingsheng Long, Jianmin Wang, Philip S Yu  
HIGHLIGHT: We present PredRNN++, a recurrent network for spatiotemporal predictive learning.
- 528, TITLE: Analyzing the Robustness of Nearest Neighbors to Adversarial Examples  
<http://proceedings.mlr.press/v80/wang18c.html>  
AUTHORS: Yizhen Wang, Somesh Jha, Kamalika Chaudhuri  
HIGHLIGHT: In this work, we introduce a theoretical framework analogous to bias-variance theory for understanding these effects.
- 529, TITLE: Competitive Multi-agent Inverse Reinforcement Learning with Sub-optimal Demonstrations  
<http://proceedings.mlr.press/v80/wang18d.html>  
AUTHORS: Xingyu Wang, Diego Klabjan  
HIGHLIGHT: Compared to previous works that decouple agents in the game by assuming optimality in expert policies, we introduce a new objective function that directly pits experts against Nash Equilibrium policies, and we design an algorithm to solve for the reward function in the context of inverse reinforcement learning with deep neural networks as model approximations.
- 530, TITLE: Coded Sparse Matrix Multiplication  
<http://proceedings.mlr.press/v80/wang18e.html>  
AUTHORS: Sinong Wang, Jiashang Liu, Ness Shroff  
HIGHLIGHT: In this paper, we develop a new coded computation strategy, we call sparse code, which achieves near optimal recovery threshold, low computation overhead, and linear decoding time  $O(\text{nnz}(C))$ .
- 531, TITLE: A Fast and Scalable Joint Estimator for Integrating Additional Knowledge in Learning Multiple Related Sparse Gaussian Graphical Models  
<http://proceedings.mlr.press/v80/wang18f.html>  
AUTHORS: Beilun Wang, Arshdeep Sekhon, Yanjun Qi  
HIGHLIGHT: In this paper, we propose a novel  $\text{Joint Elementary Estimator}$  incorporating additional  $\text{Knowledge}$  (JEEK) to infer multiple related sparse Gaussian Graphical models from large-scale heterogeneous data.
- 532, TITLE: Provable Variable Selection for Streaming Features  
<http://proceedings.mlr.press/v80/wang18g.html>  
AUTHORS: Jing Wang, Jie Shen, Ping Li  
HIGHLIGHT: In this work, we propose a provable online feature selection algorithm that utilizes the online leverage score.
- 533, TITLE: Style Tokens: Unsupervised Style Modeling, Control and Transfer in End-to-End Speech Synthesis  
<http://proceedings.mlr.press/v80/wang18h.html>  
AUTHORS: Yuxuan Wang, Daisy Stanton, Yu Zhang, RJ-Skerry Ryan, Eric Battenberg, Joel Shor, Ying Xiao, Ye Jia, Fei Ren, Rif A. Saurous

**HIGHLIGHT:** In this work, we propose “global style tokens” (GSTs), a bank of embeddings that are jointly trained within Tacotron, a state-of-the-art end-to-end speech synthesis system.

534, **TITLE:** Adversarial Distillation of Bayesian Neural Network Posteriors  
<http://proceedings.mlr.press/v80/wang18i.html>  
**AUTHORS:** Kuan-Chieh Wang, Paul Vicol, James Lucas, Li Gu, Roger Grosse, Richard Zemel  
**HIGHLIGHT:** We propose a framework, Adversarial Posterior Distillation, to distill the SGLD samples using a Generative Adversarial Network (GAN).

535, **TITLE:** Minimax Concave Penalized Multi-Armed Bandit Model with High-Dimensional Covariates  
<http://proceedings.mlr.press/v80/wang18j.html>  
**AUTHORS:** Xue Wang, Mingcheng Wei, Tao Yao  
**HIGHLIGHT:** In this paper, we propose a Minimax Concave Penalized Multi-Armed Bandit (MCP-Bandit) algorithm for a decision-maker facing high-dimensional data with latent sparse structure in an online learning and decision-making process.

536, **TITLE:** Online Convolutional Sparse Coding with Sample-Dependent Dictionary  
<http://proceedings.mlr.press/v80/wang18k.html>  
**AUTHORS:** Yaqing Wang, Quanming Yao, James Tin-Yau Kwok, Lionel M. NI  
**HIGHLIGHT:** In this paper, instead of convolving with a dictionary shared by all samples, we propose the use of a sample-dependent dictionary in which each filter is a linear combination of a small set of base filters learned from data.

537, **TITLE:** Stein Variational Message Passing for Continuous Graphical Models  
<http://proceedings.mlr.press/v80/wang18l.html>  
**AUTHORS:** Dilin Wang, Zhe Zeng, Qiang Liu  
**HIGHLIGHT:** We propose a novel distributed inference algorithm for continuous graphical models, by extending Stein variational gradient descent (SVGD) to leverage the Markov dependency structure of the distribution of interest.

538, **TITLE:** Approximate Leave-One-Out for Fast Parameter Tuning in High Dimensions  
<http://proceedings.mlr.press/v80/wang18m.html>  
**AUTHORS:** Shuaiwen Wang, Wenda Zhou, Haihao Lu, Arian Maleki, Vahab Mirrokni  
**HIGHLIGHT:** We propose two frameworks to obtain a computationally efficient approximation ALO of the leave-one-out cross validation (LOOCV) risk for nonsmooth losses and regularizers.

539, **TITLE:** Curriculum Learning by Transfer Learning: Theory and Experiments with Deep Networks  
<http://proceedings.mlr.press/v80/weinshall18a.html>  
**AUTHORS:** Daphna Weinshall, Gad Cohen, Dan Amir  
**HIGHLIGHT:** We describe a method which infers the curriculum by way of transfer learning from another network, pre-trained on a different task.

540, **TITLE:** Extracting Automata from Recurrent Neural Networks Using Queries and Counterexamples  
<http://proceedings.mlr.press/v80/weiss18a.html>  
**AUTHORS:** Gail Weiss, Yoav Goldberg, Eran Yahav  
**HIGHLIGHT:** We present a novel algorithm that uses exact learning and abstraction to extract a deterministic finite automaton describing the state dynamics of a given trained RNN.

541, **TITLE:** LeapsAndBounds: A Method for Approximately Optimal Algorithm Configuration  
<http://proceedings.mlr.press/v80/weisz18a.html>  
**AUTHORS:** Gellert Weisz, Andras Gyorgy, Csaba Szepesvari  
**HIGHLIGHT:** We propose LeapsAndBounds, an algorithm that tests configurations on randomly selected problem instances for longer and longer time.

542, **TITLE:** Deep Predictive Coding Network for Object Recognition



<http://proceedings.mlr.press/v80/wen18a.html>

AUTHORS: Haiguang Wen, Kuan Han, Junxing Shi, Yizhen Zhang, Eugenio Culurciello, Zhongming Liu  
HIGHLIGHT: Based on the predictive coding theory in neuro- science, we designed a bi-directional and recurrent neural net, namely deep predictive coding networks (PCN), that has feedforward, feedback, and recurrent connections.

543, TITLE: Towards Fast Computation of Certified Robustness for ReLU Networks

<http://proceedings.mlr.press/v80/weng18a.html>

AUTHORS: Lily Weng, Huan Zhang, Hongge Chen, Zhao Song, Cho-Jui Hsieh, Luca Daniel, Duane Boning, Inderjit Dhillon

HIGHLIGHT: In this paper, we exploit the special structure of ReLU networks and provide two computationally efficient algorithms (Fast-Lin, Fast-Lip) that are able to certify non-trivial lower bounds of minimum adversarial distortions.

544, TITLE: Provable Defenses against Adversarial Examples via the Convex Outer Adversarial Polytope

<http://proceedings.mlr.press/v80/wong18a.html>

AUTHORS: Eric Wong, Zico Kolter

HIGHLIGHT: We propose a method to learn deep ReLU-based classifiers that are provably robust against norm-bounded adversarial perturbations on the training data.

545, TITLE: Local Density Estimation in High Dimensions

<http://proceedings.mlr.press/v80/wu18a.html>

AUTHORS: Xian Wu, Moses Charikar, Vishnu Natchu

HIGHLIGHT: An important question that arises in the study of high dimensional vector representations learned from data is: given a set  $D$  of vectors and a query  $q$ , estimate the number of points within a specified distance threshold of  $q$ .

546, TITLE: Adaptive Exploration-Exploitation Tradeoff for Opportunistic Bandits

<http://proceedings.mlr.press/v80/wu18b.html>

AUTHORS: Huasen Wu, Xueying Guo, Xin Liu

HIGHLIGHT: In this paper, we propose and study opportunistic bandits - a new variant of bandits where the regret of pulling a suboptimal arm varies under different environmental conditions, such as network load or produce price.

547, TITLE: SQL-Rank: A Listwise Approach to Collaborative Ranking

<http://proceedings.mlr.press/v80/wu18c.html>

AUTHORS: Liwei Wu, Cho-Jui Hsieh, James Sharpnack

HIGHLIGHT: In this paper, we propose a listwise approach for constructing user-specific rankings in recommendation systems in a collaborative fashion.

548, TITLE: Error Compensated Quantized SGD and its Applications to Large-scale Distributed Optimization

<http://proceedings.mlr.press/v80/wu18d.html>

AUTHORS: Jiaxiang Wu, Weidong Huang, Junzhou Huang, Tong Zhang

HIGHLIGHT: In this paper, we propose the error compensated quantized stochastic gradient descent algorithm to improve the training efficiency.

549, TITLE: Reinforcing Adversarial Robustness using Model Confidence Induced by Adversarial Training

<http://proceedings.mlr.press/v80/wu18e.html>

AUTHORS: Xi Wu, Uyeong Jang, Jiefeng Chen, Lingjiao Chen, Somesh Jha

HIGHLIGHT: In this paper we study leveraging confidence information induced by adversarial training to reinforce adversarial robustness of a given adversarially trained model.

550, TITLE: Discrete-Continuous Mixtures in Probabilistic Programming: Generalized Semantics and Inference Algorithms

<http://proceedings.mlr.press/v80/wu18f.html>

AUTHORS: Yi Wu, Siddharth Srivastava, Nicholas Hay, Simon Du, Stuart Russell  
HIGHLIGHT: We develop two new general sampling algorithms that are provably correct under the MTBN framework: the lexicographic likelihood weighting (LLW) for general MTBNs and the lexicographic particle filter (LPF), a specialized algorithm for state-space models.

551, TITLE: Variance Regularized Counterfactual Risk Minimization via Variational Divergence Minimization

<http://proceedings.mlr.press/v80/wu18g.html>

AUTHORS: Hang Wu, May Wang  
HIGHLIGHT: In this work, inspired by learning bounds for importance sampling problems, we present a new counterfactual learning principle for off-policy learning with bandit feedbacks.

552, TITLE: Deep k-Means: Re-Training and Parameter Sharing with Harder Cluster Assignments for Compressing Deep Convolutions

<http://proceedings.mlr.press/v80/wu18h.html>

AUTHORS: Junru Wu, Yue Wang, Zhenyu Wu, Zhangyang Wang, Ashok Veeraraghavan, Yingyan Lin  
HIGHLIGHT: To this end, we proposed a simple yet effective scheme for compressing convolutions though applying k-means clustering on the weights, compression is achieved through weight-sharing, by only recording  $\$K\$$  cluster centers and weight assignment indexes.  
We additionally propose an improved set of metrics to estimate energy consumption of CNN hardware implementations, whose estimation results are verified to be consistent with previously proposed energy estimation tool extrapolated from actual hardware measurements.

553, TITLE: Bayesian Quadrature for Multiple Related Integrals

<http://proceedings.mlr.press/v80/xi18a.html>

AUTHORS: Xiaoyue Xi, Francois-Xavier Briol, Mark Girolami  
HIGHLIGHT: In this paper, we demonstrate that this paradigm can provide additional advantages, such as the possibility of transferring information between several numerical methods.

554, TITLE: Model-Level Dual Learning

<http://proceedings.mlr.press/v80/xia18a.html>

AUTHORS: Yingce Xia, Xu Tan, Fei Tian, Tao Qin, Nenghai Yu, Tie-Yan Liu  
HIGHLIGHT: In this work, we propose a new learning framework, model-level dual learning, which takes duality of tasks into consideration while designing the architectures for the primal/dual models, and ties the model parameters that playing similar roles in the two tasks.

555, TITLE: Dynamical Isometry and a Mean Field Theory of CNNs: How to Train 10,000-Layer Vanilla Convolutional Neural Networks

<http://proceedings.mlr.press/v80/xiao18a.html>

AUTHORS: Lechao Xiao, Yasaman Bahri, Jascha Sohl-Dickstein, Samuel Schoenholz, Jeffrey Pennington  
HIGHLIGHT: In this work, we demonstrate that it is possible to train vanilla CNNs with ten thousand layers or more simply by using an appropriate initialization scheme.

556, TITLE: Orthogonality-Promoting Distance Metric Learning: Convex Relaxation and Theoretical Analysis

<http://proceedings.mlr.press/v80/xie18a.html>

AUTHORS: Pengtao Xie, Wei Wu, Yichen Zhu, Eric Xing  
HIGHLIGHT: In this paper, we address these three issues by (1) seeking convex relaxations of the original nonconvex problems so that the global optimal is guaranteed to be achievable; (2) providing a formal analysis on OPR's capability of promoting balancedness; (3) providing a theoretical analysis that directly reveals the relationship between OPR and generalization performance.

557, TITLE: Nonoverlap-Promoting Variable Selection

<http://proceedings.mlr.press/v80/xie18b.html>

AUTHORS: Pengtao Xie, Hongbao Zhang, Yichen Zhu, Eric Xing

**HIGHLIGHT:** In this paper, we consider variable selection for models where multiple responses are to be predicted based on the same set of covariates.

558, **TITLE:** Learning Semantic Representations for Unsupervised Domain Adaptation  
<http://proceedings.mlr.press/v80/xie18c.html>  
**AUTHORS:** Shaoan Xie, Zibin Zheng, Liang Chen, Chuan Chen  
**HIGHLIGHT:** In this paper, we present moving semantic transfer network, which learn semantic representations for unlabeled target samples by aligning labeled source centroid and pseudo-labeled target centroid.

559, **TITLE:** Rates of Convergence of Spectral Methods for Graphon Estimation  
<http://proceedings.mlr.press/v80/xu18a.html>  
**AUTHORS:** Jiaming Xu  
**HIGHLIGHT:** This paper studies the problem of estimating the graphon function – a generative mechanism for a class of random graphs that are useful approximations to real networks.

560, **TITLE:** Learning Registered Point Processes from Idiosyncratic Observations  
<http://proceedings.mlr.press/v80/xu18b.html>  
**AUTHORS:** Hongteng Xu, Lawrence Carin, Hongyuan Zha  
**HIGHLIGHT:** An alternating optimization method is proposed to learn a “registered” point process that accounts for shared structure, as well as “warping” functions that characterize idiosyncratic aspects of each observed sequence.

561, **TITLE:** Representation Learning on Graphs with Jumping Knowledge Networks  
<http://proceedings.mlr.press/v80/xu18c.html>  
**AUTHORS:** Keyulu Xu, Chengtao Li, Yonglong Tian, Tomohiro Sonobe, Ken-ichi Kawarabayashi, Stefanie Jegelka  
**HIGHLIGHT:** We analyze some important properties of these models, and propose a strategy to overcome those.

562, **TITLE:** Learning to Explore via Meta-Policy Gradient  
<http://proceedings.mlr.press/v80/xu18d.html>  
**AUTHORS:** Tianbing Xu, Qiang Liu, Liang Zhao, Jian Peng  
**HIGHLIGHT:** In this work, we develop a simple meta-policy gradient algorithm that allows us to adaptively learn the exploration policy in DDPG.

563, **TITLE:** Nonparametric Regression with Comparisons: Escaping the Curse of Dimensionality with Ordinal Information  
<http://proceedings.mlr.press/v80/xu18e.html>  
**AUTHORS:** Yichong Xu, Hariank Muthakana, Sivaraman Balakrishnan, Aarti Singh, Artur Dubrawski  
**HIGHLIGHT:** We develop an algorithm called Ranking-Regression (RR) and analyze its accuracy as a function of size of the labeled and unlabeled datasets and various noise parameters.

564, **TITLE:** Optimal Tuning for Divide-and-conquer Kernel Ridge Regression with Massive Data  
<http://proceedings.mlr.press/v80/xu18f.html>  
**AUTHORS:** Ganggang Xu, Zuofeng Shang, Guang Cheng  
**HIGHLIGHT:** In this paper, we propose a data-driven procedure based on divide-and-conquer for selecting the tuning parameters in kernel ridge regression by modifying the popular Generalized Cross-validation (GCV, Wahba, 1990).

565, **TITLE:** Continuous and Discrete-time Accelerated Stochastic Mirror Descent for Strongly Convex Functions  
<http://proceedings.mlr.press/v80/xu18g.html>  
**AUTHORS:** Pan Xu, Tianhao Wang, Quanquan Gu  
**HIGHLIGHT:** Our results suggest that the only existing ASMD algorithm, namely, AC-SA proposed in Ghadimi & Lan (2012) is one instance of its kind, and we can actually derive new instances of ASMD with fewer tuning parameters.

- 566, TITLE: A Semantic Loss Function for Deep Learning with Symbolic Knowledge  
<http://proceedings.mlr.press/v80/xu18h.html>  
AUTHORS: Jingyi Xu, Zilu Zhang, Tal Friedman, Yitao Liang, Guy Broeck  
HIGHLIGHT: This paper develops a novel methodology for using symbolic knowledge in deep learning.
- 567, TITLE: Causal Bandits with Propagating Inference  
<http://proceedings.mlr.press/v80/yabel18a.html>  
AUTHORS: Akihiro Yabe, Daisuke Hatano, Hanna Sumita, Shinji Ito, Naonori Kakimura, Takuro Fukunaga, Ken-ichi Kawarabayashi  
HIGHLIGHT: Recently, a bandit framework over a causal graph was introduced, where the structure of the causal graph is available as side-information and the arms are identified with interventions on the causal graph.
- 568, TITLE: Active Learning with Logged Data  
<http://proceedings.mlr.press/v80/yan18a.html>  
AUTHORS: Songbai Yan, Kamalika Chaudhuri, Tara Javidi  
HIGHLIGHT: In this paper, we combine both approaches to provide an algorithm that uses logged data to bootstrap and inform experimentation, thus achieving the best of both worlds.
- 569, TITLE: Binary Classification with Karmic, Threshold-Quasi-Concave Metrics  
<http://proceedings.mlr.press/v80/yan18b.html>  
AUTHORS: Bowei Yan, Sanmi Koyejo, Kai Zhong, Pradeep Ravikumar  
HIGHLIGHT: In this paper, we advance this understanding of binary classification for complex performance measures by identifying two key properties: a so-called Karmic property, and a more technical threshold-quasi-concavity property, which we show is milder than existing structural assumptions imposed on performance measures.
- 570, TITLE: Characterizing and Learning Equivalence Classes of Causal DAGs under Interventions  
<http://proceedings.mlr.press/v80/yang18a.html>  
AUTHORS: Karren Yang, Abigail Katcoff, Caroline Uhler  
HIGHLIGHT: In this paper, we extend these identifiability results to general interventions, which may modify the dependencies between targeted variables and their causes without eliminating them.
- 571, TITLE: Dependent Relational Gamma Process Models for Longitudinal Networks  
<http://proceedings.mlr.press/v80/yang18b.html>  
AUTHORS: Sikun Yang, Heinz Koepl  
HIGHLIGHT: A probabilistic framework based on the covariate-dependent relational gamma process is developed to analyze relational data arising from longitudinal networks.
- 572, TITLE: Goodness-of-Fit Testing for Discrete Distributions via Stein Discrepancy  
<http://proceedings.mlr.press/v80/yang18c.html>  
AUTHORS: Jiasen Yang, Qiang Liu, Vinayak Rao, Jennifer Neville  
HIGHLIGHT: In this work, we introduce a kernelized Stein discrepancy measure for discrete spaces, and develop a nonparametric goodness-of-fit test for discrete distributions with intractable normalization constants.
- 573, TITLE: Mean Field Multi-Agent Reinforcement Learning  
<http://proceedings.mlr.press/v80/yang18d.html>  
AUTHORS: Yaodong Yang, Rui Luo, Minne Li, Ming Zhou, Weinan Zhang, Jun Wang  
HIGHLIGHT: In this paper, we present Mean Field Reinforcement Learning where the interactions within the population of agents are approximated by those between a single agent and the average effect from the overall population or neighboring agents; the interplay between the two entities is mutually reinforced: the learning of the individual agent's optimal policy depends on the dynamics of the population, while the dynamics of the population change according to the collective patterns of the individual policies.

- 574, TITLE: Yes, but Did It Work?: Evaluating Variational Inference  
<http://proceedings.mlr.press/v80/yao18a.html>  
AUTHORS: Yuling Yao, Aki Vehtari, Daniel Simpson, Andrew Gelman  
HIGHLIGHT: We propose two diagnostic algorithms to alleviate this problem.
- 575, TITLE: Hierarchical Text Generation and Planning for Strategic Dialogue  
<http://proceedings.mlr.press/v80/yarats18a.html>  
AUTHORS: Denis Yarats, Mike Lewis  
HIGHLIGHT: We introduce an approach to learning representations of messages in dialogues by maximizing the likelihood of subsequent sentences and actions, which decouples the semantics of the dialogue utterance from its linguistic realization.
- 576, TITLE: Massively Parallel Algorithms and Hardness for Single-Linkage Clustering under  $\ell_p$  Distances  
<http://proceedings.mlr.press/v80/yaroslavtsev18a.html>  
AUTHORS: Grigory Yaroslavtsev, Adithya Vadapalli  
HIGHLIGHT: We present first massively parallel (MPC) algorithms and hardness of approximation results for computing Single-Linkage Clustering of  $n$  input  $d$ -dimensional vectors under Hamming,  $\ell_1$ ,  $\ell_2$  and  $\ell_\infty$  distances.
- 577, TITLE: Communication-Computation Efficient Gradient Coding  
<http://proceedings.mlr.press/v80/ye18a.html>  
AUTHORS: Min Ye, Emmanuel Abbe  
HIGHLIGHT: This paper develops coding techniques to reduce the running time of distributed learning tasks.
- 578, TITLE: Variable Selection via Penalized Neural Network: a Drop-Out-One Loss Approach  
<http://proceedings.mlr.press/v80/ye18b.html>  
AUTHORS: Mao Ye, Yan Sun  
HIGHLIGHT: We propose a variable selection method for high dimensional regression models, which allows for complex, nonlinear, and high-order interactions among variables.
- 579, TITLE: Loss Decomposition for Fast Learning in Large Output Spaces  
<http://proceedings.mlr.press/v80/yen18a.html>  
AUTHORS: Ian En-Hsu Yen, Satyen Kale, Felix Yu, Daniel Holtmann-Rice, Sanjiv Kumar, Pradeep Ravikumar  
HIGHLIGHT: In this work, we propose a novel technique to reduce the intractable high dimensional search problem to several much more tractable lower dimensional ones via dual decomposition of the loss function.
- 580, TITLE: Byzantine-Robust Distributed Learning: Towards Optimal Statistical Rates  
<http://proceedings.mlr.press/v80/yin18a.html>  
AUTHORS: Dong Yin, Yudong Chen, Ramchandran Kannan, Peter Bartlett  
HIGHLIGHT: In this paper, we develop distributed optimization algorithms that are provably robust against Byzantine failures—arbitrary and potentially adversarial behavior, in distributed computing systems, with a focus on achieving optimal statistical performance.
- 581, TITLE: Semi-Implicit Variational Inference  
<http://proceedings.mlr.press/v80/yin18b.html>  
AUTHORS: Mingzhang Yin, Mingyuan Zhou  
HIGHLIGHT: Semi-Implicit Variational Inference
- 582, TITLE: Disentangled Sequential Autoencoder  
<http://proceedings.mlr.press/v80/yingzhen18a.html>  
AUTHORS: Li Yingzhen, Stephan Mandt  
HIGHLIGHT: We present a VAE architecture for encoding and generating high dimensional sequential data, such as video or audio.

- 583, TITLE: Probably Approximately Metric-Fair Learning  
<http://proceedings.mlr.press/v80/yona18a.html>  
AUTHORS: Gal Yona, Guy Rothblum  
HIGHLIGHT: The seminal work of Dwork et al. [ITCS 2012] introduced a metric-based notion of individual fairness: given a task-specific similarity metric, their notion required that every pair of similar individuals should be treated similarly.
- 584, TITLE: GAIN: Missing Data Imputation using Generative Adversarial Nets  
<http://proceedings.mlr.press/v80/yoon18a.html>  
AUTHORS: Jinsung Yoon, James Jordon, Mihaela Schaar  
HIGHLIGHT: We propose a novel method for imputing missing data by adapting the well-known Generative Adversarial Nets (GAN) framework.
- 585, TITLE: RadialGAN: Leveraging multiple datasets to improve target-specific predictive models using Generative Adversarial Networks  
<http://proceedings.mlr.press/v80/yoon18b.html>  
AUTHORS: Jinsung Yoon, James Jordon, Mihaela Schaar  
HIGHLIGHT: In this paper we propose a novel approach to the problem in which we use multiple GAN architectures to learn to translate from one dataset to another, thereby allowing us to effectively enlarge the target dataset, and therefore learn better predictive models than if we simply used the target dataset.
- 586, TITLE: GraphRNN: Generating Realistic Graphs with Deep Auto-regressive Models  
<http://proceedings.mlr.press/v80/you18a.html>  
AUTHORS: Jiaxuan You, Rex Ying, Xiang Ren, William Hamilton, Jure Leskovec  
HIGHLIGHT: Here we propose GraphRNN, a deep autoregressive model that addresses the above challenges and approximates any distribution of graphs with minimal assumptions about their structure.
- 587, TITLE: An Efficient Semismooth Newton based Algorithm for Convex Clustering  
<http://proceedings.mlr.press/v80/yuan18a.html>  
AUTHORS: Yancheng Yuan, Defeng Sun, Kim-Chuan Toh  
HIGHLIGHT: In this paper, we propose a semismooth Newton based augmented Lagrangian method for large-scale convex clustering problems.
- 588, TITLE: A Conditional Gradient Framework for Composite Convex Minimization with Applications to Semidefinite Programming  
<http://proceedings.mlr.press/v80/yurtsever18a.html>  
AUTHORS: Alp Yurtsever, Olivier Fercoq, Francesco Locatello, Volkan Cevher  
HIGHLIGHT: We propose a conditional gradient framework for a composite convex minimization template with broad applications.
- 589, TITLE: Policy Optimization as Wasserstein Gradient Flows  
<http://proceedings.mlr.press/v80/zhang18a.html>  
AUTHORS: Ruiyi Zhang, Changyou Chen, Chunyuan Li, Lawrence Carin  
HIGHLIGHT: To make optimization feasible, we develop efficient algorithms by numerically solving the corresponding discrete gradient flows.
- 590, TITLE: Problem Dependent Reinforcement Learning Bounds Which Can Identify Bandit Structure in MDPs  
<http://proceedings.mlr.press/v80/zanette18a.html>  
AUTHORS: Andrea Zanette, Emma Brunskill  
HIGHLIGHT: In this paper, we study whether there exist algorithms for the more general framework (MDP) which automatically provide the best performance bounds for the specific problem at hand without user intervention and without modifying the algorithm.

591, TITLE: Fast and Sample Efficient Inductive Matrix Completion via Multi-Phase Procrustes Flow  
<http://proceedings.mlr.press/v80/zhang18b.html>  
AUTHORS: Xiao Zhang, Simon Du, Quanquan Gu  
HIGHLIGHT: We present and analyze a new gradient-based non-convex optimization algorithm that converges to the true underlying matrix at a linear rate with sample complexity only linearly depending on  $n$  and logarithmically depending on  $d$ .

592, TITLE: Large-Scale Sparse Inverse Covariance Estimation via Thresholding and Max-Det Matrix Completion  
<http://proceedings.mlr.press/v80/zhang18c.html>  
AUTHORS: Richard Zhang, Salar Fattahi, Somayeh Sojoudi  
HIGHLIGHT: This paper proves an extension of this result, and describes a Newton-CG algorithm to efficiently solve the MDMC problem.

593, TITLE: High Performance Zero-Memory Overhead Direct Convolutions  
<http://proceedings.mlr.press/v80/zhang18d.html>  
AUTHORS: Jiyuan Zhang, Franz Franchetti, Tze Meng Low  
HIGHLIGHT: In this paper, we demonstrate that direct convolution, when implemented correctly, eliminates all memory overhead, and yields performance that is between 10% to 400% times better than existing high performance implementations of convolution layers on conventional and embedded CPU architectures.

594, TITLE: Safe Element Screening for Submodular Function Minimization  
<http://proceedings.mlr.press/v80/zhang18e.html>  
AUTHORS: Weizhong Zhang, Bin Hong, Lin Ma, Wei Liu, Tong Zhang  
HIGHLIGHT: Relying on this study, we subsequently propose a novel safe screening method to quickly identify the elements guaranteed to be included (we refer to them as active) or excluded (inactive) in the final optimal solution of SFM during the optimization process.

595, TITLE: Improving the Privacy and Accuracy of ADMM-Based Distributed Algorithms  
<http://proceedings.mlr.press/v80/zhang18f.html>  
AUTHORS: Xueru Zhang, Mohammad Mahdi Khalili, Mingyan Liu  
HIGHLIGHT: We propose a perturbation method for ADMM where the perturbed term is correlated with the penalty parameters; this is shown to improve the utility and privacy simultaneously.

596, TITLE: Stabilizing Gradients for Deep Neural Networks via Efficient SVD Parameterization  
<http://proceedings.mlr.press/v80/zhang18g.html>  
AUTHORS: Jiong Zhang, Qi Lei, Inderjit Dhillon  
HIGHLIGHT: In this paper, we present an efficient parametrization of the transition matrix of an RNN that allows us to stabilize the gradients that arise in its training.

597, TITLE: Learning Long Term Dependencies via Fourier Recurrent Units  
<http://proceedings.mlr.press/v80/zhang18h.html>  
AUTHORS: Jiong Zhang, Yibo Lin, Zhao Song, Inderjit Dhillon  
HIGHLIGHT: In this paper we propose a simple recurrent architecture, the Fourier Recurrent Unit (FRU), that stabilizes the gradients that arise in its training while giving us stronger expressive power.

598, TITLE: Tropical Geometry of Deep Neural Networks  
<http://proceedings.mlr.press/v80/zhang18i.html>  
AUTHORS: Liwen Zhang, Gregory Naitzat, Lek-Heng Lim  
HIGHLIGHT: We establish, for the first time, explicit connections between feedforward neural networks with ReLU activation and tropical geometry — we show that the family of such neural networks is equivalent to the family of tropical rational maps.

- 599, TITLE: Deep Bayesian Nonparametric Tracking  
<http://proceedings.mlr.press/v80/zhang18j.html>  
AUTHORS: Aonan Zhang, John Paisley  
HIGHLIGHT: We propose an integration of Bayesian nonparametric methods within deep neural networks for modeling irregular patterns in time-series data.
- 600, TITLE: Composable Planning with Attributes  
<http://proceedings.mlr.press/v80/zhang18k.html>  
AUTHORS: Amy Zhang, Sainbayar Sukhbaatar, Adam Lerer, Arthur Szlam, Rob Fergus  
HIGHLIGHT: We propose a method that learns a policy for transitioning between “nearby” sets of attributes, and maintains a graph of possible transitions.
- 601, TITLE: Noisy Natural Gradient as Variational Inference  
<http://proceedings.mlr.press/v80/zhang18l.html>  
AUTHORS: Guodong Zhang, Shengyang Sun, David Duvenaud, Roger Grosse  
HIGHLIGHT: We show that natural gradient ascent with adaptive weight noise implicitly fits a variational posterior to maximize the evidence lower bound (ELBO).
- 602, TITLE: A Primal-Dual Analysis of Global Optimality in Nonconvex Low-Rank Matrix Recovery  
<http://proceedings.mlr.press/v80/zhang18m.html>  
AUTHORS: Xiao Zhang, Lingxiao Wang, Yaodong Yu, Quanquan Gu  
HIGHLIGHT: We propose a primal-dual based framework for analyzing the global optimality of nonconvex low-rank matrix recovery.
- 603, TITLE: Fully Decentralized Multi-Agent Reinforcement Learning with Networked Agents  
<http://proceedings.mlr.press/v80/zhang18n.html>  
AUTHORS: Kaiqing Zhang, Zhuoran Yang, Han Liu, Tong Zhang, Tamer Basar  
HIGHLIGHT: To maximize the globally averaged return over the network, we propose two fully decentralized actor-critic algorithms, which are applicable to large-scale MARL problems in an online fashion.
- 604, TITLE: Dynamic Regret of Strongly Adaptive Methods  
<http://proceedings.mlr.press/v80/zhang18o.html>  
AUTHORS: Lijun Zhang, Tianbao Yang, jin, Zhi-Hua Zhou  
HIGHLIGHT: In this paper, we illustrate an intrinsic connection between these two concepts by showing that the dynamic regret can be expressed in terms of the adaptive regret and the functional variation.
- 605, TITLE: Inter and Intra Topic Structure Learning with Word Embeddings  
<http://proceedings.mlr.press/v80/zhao18a.html>  
AUTHORS: He Zhao, Lan Du, Wray Buntine, Mingyuan Zhou  
HIGHLIGHT: In this paper, we propose a novel topic model with a deep structure that explores both inter-topic and intra-topic structures informed by word embeddings.
- 606, TITLE: Adversarially Regularized Autoencoders  
<http://proceedings.mlr.press/v80/zhao18b.html>  
AUTHORS: Junbo Zhao, Yoon Kim, Kelly Zhang, Alexander Rush, Yann LeCun  
HIGHLIGHT: In this work, we propose a more flexible method for training deep latent variable models of discrete structures.
- 607, TITLE: MSplit LBI: Realizing Feature Selection and Dense Estimation Simultaneously in Few-shot and Zero-shot Learning  
<http://proceedings.mlr.press/v80/zhao18c.html>  
AUTHORS: Bo Zhao, Xinwei Sun, Yanwei Fu, Yuan Yao, Yizhou Wang



**HIGHLIGHT:** In this paper, we propose the idea that the features consist of three orthogonal parts, namely sparse strong signals, dense weak signals and random noise, in which both strong and weak signals contribute to the fitting of data.

608, **TITLE:** Composite Marginal Likelihood Methods for Random Utility Models

<http://proceedings.mlr.press/v80/zhao18d.html>

**AUTHORS:** Zhibing Zhao, Lirong Xia

**HIGHLIGHT:** We propose a novel and flexible rank-breaking-then-composite-marginal-likelihood (RBCML) framework for learning random utility models (RUMs), which include the Plackett-Luce model.

609, **TITLE:** Lightweight Stochastic Optimization for Minimizing Finite Sums with Infinite Data

<http://proceedings.mlr.press/v80/zheng18a.html>

**AUTHORS:** Shuai Zheng, James Tin-Yau Kwok

**HIGHLIGHT:** In this paper, we propose two SGD-like algorithms for expected risk minimization with random perturbation, namely, stochastic sample average gradient (SSAG) and stochastic SAGA (S-SAGA).

610, **TITLE:** A Robust Approach to Sequential Information Theoretic Planning

<http://proceedings.mlr.press/v80/zheng18b.html>

**AUTHORS:** Sue Zheng, Jason Pacheco, John Fisher

**HIGHLIGHT:** Through the use of robust M-estimators we obtain bounds on absolute deviation of estimated MI.

611, **TITLE:** Revealing Common Statistical Behaviors in Heterogeneous Populations

<http://proceedings.mlr.press/v80/zhitnikov18a.html>

**AUTHORS:** Andrey Zhitnikov, Rotem Mulayoff, Tomer Michaeli

**HIGHLIGHT:** In this paper, we propose nonparametric algorithms for estimating the common covariance matrix and the common density function of several variables in a heterogeneous group of subjects.

612, **TITLE:** Understanding Generalization and Optimization Performance of Deep CNNs

<http://proceedings.mlr.press/v80/zhou18a.html>

**AUTHORS:** Pan Zhou, Jiashi Feng

**HIGHLIGHT:** This work aims to provide understandings on the remarkable success of deep convolutional neural networks (CNNs) by theoretically analyzing their generalization performance and establishing optimization guarantees for gradient descent based training algorithms.

613, **TITLE:** Distributed Asynchronous Optimization with Unbounded Delays: How Slow Can You Go?

<http://proceedings.mlr.press/v80/zhou18b.html>

**AUTHORS:** Zhengyuan Zhou, Panayotis Mertikopoulos, Nicholas Bambos, Peter Glynn, Yinyu Ye, Li-Jia Li, Li Fei-Fei

**HIGHLIGHT:** Nevertheless, by using a judiciously chosen quasilinear step-size sequence, we show that it is possible to amortize these delays and achieve global convergence with probability 1, even when the delays grow at a polynomial rate.

614, **TITLE:** A Simple Stochastic Variance Reduced Algorithm with Fast Convergence Rates

<http://proceedings.mlr.press/v80/zhou18c.html>

**AUTHORS:** Kaiwen Zhou, Fanhua Shang, James Cheng

**HIGHLIGHT:** In this paper, we introduce a simple stochastic variance reduced algorithm (MiG), which enjoys the best-known convergence rates for both strongly convex and non-strongly convex problems.

615, **TITLE:** Stochastic Variance-Reduced Cubic Regularized Newton Methods

<http://proceedings.mlr.press/v80/zhou18d.html>

**AUTHORS:** Dongruo Zhou, Pan Xu, Quanquan Gu

**HIGHLIGHT:** We propose a stochastic variance-reduced cubic regularized Newton method (SVRC) for non-convex optimization.

616, TITLE: Racing Thompson: an Efficient Algorithm for Thompson Sampling with Non-conjugate Priors  
<http://proceedings.mlr.press/v80/zhou18e.html>

AUTHORS: Yichi Zhou, Jun Zhu, Jingwei Zhuo

HIGHLIGHT: In this paper, we propose a novel algorithm for Thompson sampling which only requires to draw samples from a tractable proposal distribution.

After that we construct a set of random variables and our goal is to identify the one with highest mean which is an instance of best arm identification problems.

617, TITLE: Distributed Nonparametric Regression under Communication Constraints

<http://proceedings.mlr.press/v80/zhu18a.html>

AUTHORS: Yuancheng Zhu, John Lafferty

HIGHLIGHT: This paper studies the problem of nonparametric estimation of a smooth function with data distributed across multiple machines.

618, TITLE: Message Passing Stein Variational Gradient Descent

<http://proceedings.mlr.press/v80/zhuo18a.html>

AUTHORS: Jingwei Zhuo, Chang Liu, Jiaxin Shi, Jun Zhu, Ning Chen, Bo Zhang

HIGHLIGHT: We propose Message Passing SVGD (MP-SVGD) to solve this problem.

619, TITLE: Stochastic Variance-Reduced Hamilton Monte Carlo Methods

<http://proceedings.mlr.press/v80/zou18a.html>

AUTHORS: Difan Zou, Pan Xu, Quanquan Gu

HIGHLIGHT: We propose a fast stochastic Hamilton Monte Carlo (HMC) method, for sampling from a smooth and strongly log-concave distribution.

620, TITLE: Rectify Heterogeneous Models with Semantic Mapping

<http://proceedings.mlr.press/v80/ye18c.html>

AUTHORS: Han-Jia Ye, De-Chuan Zhan, Yuan Jiang, Zhi-Hua Zhou

HIGHLIGHT: Learnware (Zhou; 2016) describes a novel perspective, and claims that learning models should have reusable and evolvable properties.

621, TITLE: Hierarchical Long-term Video Prediction without Supervision

<http://proceedings.mlr.press/v80/wichers18a.html>

AUTHORS: Nevan wickers, Ruben Villegas, Dumitru Erhan, Honglak Lee

HIGHLIGHT: Our method can predict about 20 seconds into the future and provides better results compared to Denton and Fergus (2018) and Finn et al. (2016) on the Human 3.6M dataset.