

- 1, TITLE: Uncovering Causality from Multivariate Hawkes Integrated Cumulants
<http://proceedings.mlr.press/v70/achab17a.html>
AUTHORS: Massil Achab, Emmanuel Bacry, Stéphane Gaiffas, Iacopo Mastromatteo, Jean-François Muzy
HIGHLIGHT: For that purpose, we introduce a moment matching method that fits the second-order and the third-order integrated cumulants of the process.

- 2, TITLE: A Unified Maximum Likelihood Approach for Estimating Symmetric Properties of Discrete Distributions
<http://proceedings.mlr.press/v70/acharya17a.html>
AUTHORS: Jayadev Acharya, Hirakendu Das, Alon Orlitsky, Ananda Theertha Suresh
HIGHLIGHT: We show that a single, simple, plug-in estimator—profile maximum likelihood (PML)—is sample competitive for all symmetric properties, and in particular is asymptotically sample-optimal for all the above properties.

- 3, TITLE: Constrained Policy Optimization
<http://proceedings.mlr.press/v70/achiam17a.html>
AUTHORS: Joshua Achiam, David Held, Aviv Tamar, Pieter Abbeel
HIGHLIGHT: We propose Constrained Policy Optimization (CPO), the first general-purpose policy search algorithm for constrained reinforcement learning with guarantees for near-constraint satisfaction at each iteration.

- 4, TITLE: The Price of Differential Privacy for Online Learning
<http://proceedings.mlr.press/v70/agarwal17a.html>
AUTHORS: Naman Agarwal, Karan Singh
HIGHLIGHT: We design differentially private algorithms for the problem of online linear optimization in the full information and bandit settings with optimal $O(T^{0.5})$ regret bounds.

- 5, TITLE: Local Bayesian Optimization of Motor Skills
<http://proceedings.mlr.press/v70/akrou17a.html>
AUTHORS: Riad Akrou, Dmitry Sorokin, Jan Peters, Gerhard Neumann
HIGHLIGHT: To scale to higher dimensional problems, we leverage the sample efficiency of Bayesian optimization in a local context.

- 6, TITLE: Connected Subgraph Detection with Mirror Descent on SDPs
<http://proceedings.mlr.press/v70/aksoylar17a.html>
AUTHORS: Cem Aksoylar, Lorenzo Orecchia, Venkatesh Saligrama
HIGHLIGHT: We propose a novel, computationally efficient mirror-descent based optimization framework for subgraph detection in graph-structured data.

- 7, TITLE: Learning from Clinical Judgments: Semi-Markov-Modulated Marked Hawkes Processes for Risk Prognosis
<http://proceedings.mlr.press/v70/alaa17a.html>
AUTHORS: Ahmed M. Alaa, Scott Hu, Mihaela Schaar
HIGHLIGHT: To allow for accurate prognosis of deteriorating patients, we develop a novel continuous-time probabilistic model for a monitored patient's temporal sequence of physiological data.

- 8, TITLE: A Semismooth Newton Method for Fast, Generic Convex Programming
<http://proceedings.mlr.press/v70/ali17a.html>
AUTHORS: Alnur Ali, Eric Wong, J. Zico Kolter
HIGHLIGHT: We introduce Newton-ADMM, a method for fast conic optimization.

- 9, TITLE: Learning Continuous Semantic Representations of Symbolic Expressions
<http://proceedings.mlr.press/v70/allamanis17a.html>
AUTHORS: Miltiadis Allamanis, Pankajan Chanthirasegaran, Pushmeet Kohli, Charles Sutton
HIGHLIGHT: As a step in this direction, we propose a new architecture, called neural equivalence network, for the problem of learning continuous semantic representations of algebraic and logical expressions.

- 10, TITLE: Natasha: Faster Non-Convex Stochastic Optimization via Strongly Non-Convex Parameter
<http://proceedings.mlr.press/v70/allen-zhu17a.html>
AUTHORS: Zeyuan Allen-Zhu
HIGHLIGHT: Given a non-convex function $f(x)$ that is an average of n smooth functions, we design stochastic first-order methods to find its approximate stationary points.
- 11, TITLE: Doubly Accelerated Methods for Faster CCA and Generalized Eigendecomposition
<http://proceedings.mlr.press/v70/allen-zhu17b.html>
AUTHORS: Zeyuan Allen-Zhu, Yuanzhi Li
HIGHLIGHT: We propose algorithms LazyEV and LazyCCA to solve the two problems with running times linearly dependent on the input size and on k . Furthermore, our algorithms are doubly-accelerated: our running times depend only on the square root of the matrix condition number, and on the square root of the eigengap.
- 12, TITLE: Faster Principal Component Regression and Stable Matrix Chebyshev Approximation
<http://proceedings.mlr.press/v70/allen-zhu17c.html>
AUTHORS: Zeyuan Allen-Zhu, Yuanzhi Li
HIGHLIGHT: We solve principal component regression (PCR), up to a multiplicative accuracy $1+\gamma$, by reducing the problem to $\tilde{O}(\gamma^{-1})$ black-box calls of ridge regression.
- 13, TITLE: Follow the Compressed Leader: Faster Online Learning of Eigenvectors and Faster MMWU
<http://proceedings.mlr.press/v70/allen-zhu17d.html>
AUTHORS: Zeyuan Allen-Zhu, Yuanzhi Li
HIGHLIGHT: We propose a follow-the-compressed-leader framework which, not only matches the optimal regret of MMWU (up to polylog factors), but runs no slower than FTPL.
- 14, TITLE: Near-Optimal Design of Experiments via Regret Minimization
<http://proceedings.mlr.press/v70/allen-zhu17e.html>
AUTHORS: Zeyuan Allen-Zhu, Yuanzhi Li, Aarti Singh, Yining Wang
HIGHLIGHT: We consider computationally tractable methods for the experimental design problem, where k out of n design points of dimension p are selected so that certain optimality criteria are approximately satisfied.
- 15, TITLE: OptNet: Differentiable Optimization as a Layer in Neural Networks
<http://proceedings.mlr.press/v70/amos17a.html>
AUTHORS: Brandon Amos, J. Zico Kolter
HIGHLIGHT: In this paper, we explore the foundations for such an architecture: we show how techniques from sensitivity analysis, bilevel optimization, and implicit differentiation can be used to exactly differentiate through these layers and with respect to layer parameters; we develop a highly efficient solver for these layers that exploits fast GPU-based batch solves within a primal-dual interior point method, and which provides backpropagation gradients with virtually no additional cost on top of the solve; and we highlight the application of these approaches in several problems.
- 16, TITLE: Input Convex Neural Networks
<http://proceedings.mlr.press/v70/amos17b.html>
AUTHORS: Brandon Amos, Lei Xu, J. Zico Kolter
HIGHLIGHT: In this paper we lay the basic groundwork for these models, proposing methods for inference, optimization and learning, and analyze their representational power.
- 17, TITLE: An Efficient, Sparsity-Preserving, Online Algorithm for Low-Rank Approximation
<http://proceedings.mlr.press/v70/anderson17a.html>
AUTHORS: David Anderson, Ming Gu
HIGHLIGHT: In this work, we present a novel truncated LU factorization called Spectrum-Revealing LU (SRLU) for effective low-rank matrix approximation, and develop a fast algorithm to compute an SRLU factorization.

- 18, TITLE: Modular Multitask Reinforcement Learning with Policy Sketches
<http://proceedings.mlr.press/v70/andreas17a.html>
AUTHORS: Jacob Andreas, Dan Klein, Sergey Levine
HIGHLIGHT: We describe a framework for multitask deep reinforcement learning guided by policy sketches.
- 19, TITLE: Averaged-DQN: Variance Reduction and Stabilization for Deep Reinforcement Learning
<http://proceedings.mlr.press/v70/anschel17a.html>
AUTHORS: Oron Ansel, Nir Baram, Nahum Shimkin
HIGHLIGHT: To understand the effect of the algorithm, we examine the source of value function estimation errors and provide an analytical comparison within a simplified model.
- 20, TITLE: A Simple Multi-Class Boosting Framework with Theoretical Guarantees and Empirical Proficiency
<http://proceedings.mlr.press/v70/appel17a.html>
AUTHORS: Ron Appel, Pietro Perona
HIGHLIGHT: We evaluate our method on MNIST and standard UCI datasets against other state-of-the-art methods, showing the empirical proficiency of our method.
- 21, TITLE: Deep Voice: Real-time Neural Text-to-Speech
<http://proceedings.mlr.press/v70/arik17a.html>
AUTHORS: Sercan ?. Arik, Mike Chrzanowski, Adam Coates, Gregory Diamos, Andrew Gibiansky, Yongguo Kang, Xian Li, John Miller, Andrew Ng, Jonathan Raiman, Shubho Sengupta, Mohammad Shoeybi
HIGHLIGHT: For the segmentation model, we propose a novel way of performing phoneme boundary detection with deep neural networks using connectionist temporal classification (CTC) loss.
- 22, TITLE: Oracle Complexity of Second-Order Methods for Finite-Sum Problems
<http://proceedings.mlr.press/v70/arjevani17a.html>
AUTHORS: Yossi Arjevani, Ohad Shamir
HIGHLIGHT: In this paper, we provide evidence that the answer – perhaps surprisingly – is negative, at least in terms of worst-case guarantees.
- 23, TITLE: Wasserstein Generative Adversarial Networks
<http://proceedings.mlr.press/v70/arjovsky17a.html>
AUTHORS: Martin Arjovsky, Soumith Chintala, Léon Bottou
HIGHLIGHT: We introduce a new algorithm named WGAN, an alternative to traditional GAN training.
- 24, TITLE: Generalization and Equilibrium in Generative Adversarial Nets (GANs)
<http://proceedings.mlr.press/v70/arora17a.html>
AUTHORS: Sanjeev Arora, Rong Ge, Yingyu Liang, Tengyu Ma, Yi Zhang
HIGHLIGHT: Generalization and Equilibrium in Generative Adversarial Nets (GANs)
- 25, TITLE: A Closer Look at Memorization in Deep Networks
<http://proceedings.mlr.press/v70/arpit17a.html>
AUTHORS: Devansh Arpit, Stanislaw Jastrzebski, Nicolas Ballas, David Krueger, Emmanuel Bengio, Maxinder S. Kanwal, Tegan Maharaj, Asja Fischer, Aaron Courville, Yoshua Bengio, Simon Lacoste-Julien
HIGHLIGHT: We examine the role of memorization in deep learning, drawing connections to capacity, generalization, and adversarial robustness.
- 26, TITLE: An Alternative Softmax Operator for Reinforcement Learning
<http://proceedings.mlr.press/v70/asadi17a.html>
AUTHORS: Kavosh Asadi, Michael L. Littman
HIGHLIGHT: In this work, we study a differentiable softmax operator that, among other properties, is a non-expansion ensuring a convergent behavior in learning and planning.

27, TITLE: Random Fourier Features for Kernel Ridge Regression: Approximation Bounds and Statistical Guarantees
http://proceedings.mlr.press/v70/avron17a.html
AUTHORS: Haim Avron, Michael Kapralov, Cameron Musco, Christopher Musco, Ameya Velingker, Amir Zandieh
HIGHLIGHT: In this paper we take steps toward filling this gap.

28, TITLE: Minimax Regret Bounds for Reinforcement Learning
http://proceedings.mlr.press/v70/azar17a.html
AUTHORS: Mohammad Gheshlaghi Azar, Ian Osband, Rami Munos
HIGHLIGHT: We consider the problem of provably optimal exploration in reinforcement learning for finite horizon MDPs.

29, TITLE: Learning the Structure of Generative Models without Labeled Data
http://proceedings.mlr.press/v70/bach17a.html
AUTHORS: Stephen H. Bach, Bryan He, Alexander Ratner, Christopher R?
HIGHLIGHT: We propose a structure estimation method that maximizes the ℓ_1 -regularized marginal pseudolikelihood of the observed data.

30, TITLE: Uniform Deviation Bounds for k-Means Clustering
http://proceedings.mlr.press/v70/bachem17a.html
AUTHORS: Olivier Bachem, Mario Lucic, S. Hamed Hassani, Andreas Krause
HIGHLIGHT: In this paper, we provide a novel framework to obtain uniform deviation bounds for loss functions which are unbounded.

31, TITLE: Distributed and Provably Good Seedings for k-Means in Constant Rounds
http://proceedings.mlr.press/v70/bachem17b.html
AUTHORS: Olivier Bachem, Mario Lucic, Andreas Krause
HIGHLIGHT: In this paper, we provide a novel analysis of the k-Means algorithm that bounds the expected solution quality for any number of rounds and oversampling factors greater than k , the two parameters one needs to choose in practice.

32, TITLE: Learning Algorithms for Active Learning
http://proceedings.mlr.press/v70/bachman17a.html
AUTHORS: Philip Bachman, Alessandro Sordani, Adam Trischler
HIGHLIGHT: We introduce a model that learns active learning algorithms via metalearning.

33, TITLE: Improving Viterbi is Hard: Better Runtimes Imply Faster Clique Algorithms
http://proceedings.mlr.press/v70/backurs17a.html
AUTHORS: Arturs Backurs, Christos Tzamos
HIGHLIGHT: In this paper, we explain this difficulty by providing matching conditional lower bounds.

34, TITLE: Differentially Private Clustering in High-Dimensional Euclidean Spaces
http://proceedings.mlr.press/v70/balcan17a.html
AUTHORS: Maria-Florina Balcan, Travis Dick, Yingyu Liang, Wenlong Mou, Hongyang Zhang
HIGHLIGHT: In this work, we give differentially private and efficient algorithms achieving strong guarantees for k -means and k -median clustering when $d = \Omega(\text{polylog}(n))$.

35, TITLE: Strongly-Typed Agents are Guaranteed to Interact Safely
http://proceedings.mlr.press/v70/balduzzi17a.html
AUTHORS: David Balduzzi
HIGHLIGHT: In this paper, we formalize a common-sense notion of when algorithms are well-behaved: an algorithm is safe if it does no harm.

- 36, TITLE: The Shattered Gradients Problem: If resnets are the answer, then what is the question?
<http://proceedings.mlr.press/v70/balduzzi17b.html>
AUTHORS: David Balduzzi, Marcus Frean, Lennox Leary, J. P. Lewis, Kurt Wan-Duo Ma, Brian McWilliams
HIGHLIGHT: In this paper, we identify the shattered gradients problem.
- 37, TITLE: Neural Taylor Approximations: Convergence and Exploration in Rectifier Networks
<http://proceedings.mlr.press/v70/balduzzi17c.html>
AUTHORS: David Balduzzi, Brian McWilliams, Tony Butler-Yeoman
HIGHLIGHT: Nevertheless, methods from convex optimization such as gradient descent and Adam are widely used as building blocks for deep learning algorithms.
- 38, TITLE: Spectral Learning from a Single Trajectory under Finite-State Policies
<http://proceedings.mlr.press/v70/balle17a.html>
AUTHORS: Borja Balle, Odalric-Ambrym Maillard
HIGHLIGHT: We present spectral methods of moments for learning sequential models from a single trajectory, in stark contrast with the classical literature that assumes the availability of multiple i.i.d. trajectories.
- 39, TITLE: Lost Relatives of the Gumbel Trick
<http://proceedings.mlr.press/v70/balog17a.html>
AUTHORS: Matej Balog, Nilesh Tripuraneni, Zoubin Ghahramani, Adrian Weller
HIGHLIGHT: We derive an entire family of related methods, of which the Gumbel trick is one member, and show that the new methods have superior properties in several settings with minimal additional computational cost.
- 40, TITLE: Dynamic Word Embeddings
<http://proceedings.mlr.press/v70/bamler17a.html>
AUTHORS: Robert Bamler, Stephan Mandt
HIGHLIGHT: We present a probabilistic language model for time-stamped text data which tracks the semantic evolution of individual words over time.
- 41, TITLE: End-to-End Differentiable Adversarial Imitation Learning
<http://proceedings.mlr.press/v70/baram17a.html>
AUTHORS: Nir Baram, Oron Anshel, Itai Caspi, Shie Mannor
HIGHLIGHT: In this paper, we introduce the Model-based Generative Adversarial Imitation Learning (MGAIL) algorithm.
- 42, TITLE: Emulating the Expert: Inverse Optimization through Online Learning
<http://proceedings.mlr.press/v70/barmann17a.html>
AUTHORS: Andreas Barmann, Sebastian Pokutta, Oskar Schneider
HIGHLIGHT: In this paper, we demonstrate how to learn the objective function of a decision maker while only observing the problem input data and the decision maker's corresponding decisions over multiple rounds.
- 43, TITLE: Unimodal Probability Distributions for Deep Ordinal Classification
<http://proceedings.mlr.press/v70/beckham17a.html>
AUTHORS: Christopher Beckham, Christopher Pal
HIGHLIGHT: We propose a straightforward technique to constrain discrete ordinal probability distributions to be unimodal via the use of the Poisson and binomial probability distributions.
- 44, TITLE: Globally Induced Forest: A Prepruning Compression Scheme
<http://proceedings.mlr.press/v70/begon17a.html>
AUTHORS: Jean-Michel Begon, Arnaud Joly, Pierre Geurts
HIGHLIGHT: In this paper, we propose the Globally Induced Forest (GIF) to remedy this problem.

45, TITLE: End-to-End Learning for Structured Prediction Energy Networks
<http://proceedings.mlr.press/v70/belanger17a.html>
AUTHORS: David Belanger, Bishan Yang, Andrew McCallum
HIGHLIGHT: This paper presents end-to-end learning for SPENs, where the energy function is discriminatively trained by back-propagating through gradient-based prediction.

46, TITLE: Learning to Discover Sparse Graphical Models
<http://proceedings.mlr.press/v70/belilovsky17a.html>
AUTHORS: Eugene Belilovsky, Kyle Kastner, Gael Varoquaux, Matthew B. Blaschko
HIGHLIGHT: We propose here to leverage this latter source of information as training data to learn a function, parametrized by a neural network, that maps empirical covariance matrices to estimated graph structures.

47, TITLE: A Distributional Perspective on Reinforcement Learning
<http://proceedings.mlr.press/v70/bellemare17a.html>
AUTHORS: Marc G. Bellemare, Will Dabney, R?mi Munos
HIGHLIGHT: In this paper we argue for the fundamental importance of the value distribution: the distribution of the random return received by a reinforcement learning agent.

48, TITLE: Neural Optimizer Search with Reinforcement Learning
<http://proceedings.mlr.press/v70/bello17a.html>
AUTHORS: Irwan Bello, Barret Zoph, Vijay Vasudevan, Quoc V. Le
HIGHLIGHT: We present an approach to automate the process of discovering optimization methods, with a focus on deep learning architectures.

49, TITLE: Learning Texture Manifolds with the Periodic Spatial GAN
<http://proceedings.mlr.press/v70/bergmann17a.html>
AUTHORS: Urs Bergmann, Nikolay Jetchev, Roland Vollgraf
HIGHLIGHT: This paper introduces a novel approach to texture synthesis based on generative adversarial networks (GAN) (Goodfellow et al., 2014), and call this technique Periodic Spatial GAN (PSGAN).

50, TITLE: Differentially Private Learning of Undirected Graphical Models Using Collective Graphical Models
<http://proceedings.mlr.press/v70/bernstein17a.html>
AUTHORS: Garrett Bernstein, Ryan McKenna, Tao Sun, Daniel Sheldon, Michael Hay, Gerome Miklau
HIGHLIGHT: We investigate the problem of learning discrete graphical models in a differentially private way.

51, TITLE: Efficient Online Bandit Multiclass Learning with $\tilde{O}(\sqrt{T})$ Regret
<http://proceedings.mlr.press/v70/beygelzimer17a.html>
AUTHORS: Alina Beygelzimer, Francesco Orabona, Chicheng Zhang
HIGHLIGHT: We present an efficient second-order algorithm with $\tilde{O}(1/\eta \sqrt{T})$ regret for the bandit online multiclass problem.

52, TITLE: Guarantees for Greedy Maximization of Non-submodular Functions with Applications
<http://proceedings.mlr.press/v70/bian17a.html>
AUTHORS: Andrew An Bian, Joachim M. Buhmann, Andreas Krause, Sebastian Tschiatschek
HIGHLIGHT: We investigate the performance of the standard Greedy algorithm for cardinality constrained maximization of non-submodular nondecreasing set functions.

53, TITLE: Robust Submodular Maximization: A Non-Uniform Partitioning Approach
<http://proceedings.mlr.press/v70/bogunovic17a.html>
AUTHORS: Ilija Bogunovic, Slobodan Mitrovic, Jonathan Scarlett, Volkan Cevher
HIGHLIGHT: In this paper, we solve a key open problem raised therein, presenting a new Partitioned Robust (PRo) submodular maximization algorithm that achieves the same guarantee for more general $\tau = o(k)$.

- 54, TITLE: Unsupervised Learning by Predicting Noise
<http://proceedings.mlr.press/v70/bojanowski17a.html>
AUTHORS: Piotr Bojanowski, Armand Joulin
HIGHLIGHT: We propose to fix a set of target representations, called Noise As Targets (NAT), and to constrain the deep features to align to them.
- 55, TITLE: Adaptive Neural Networks for Efficient Inference
<http://proceedings.mlr.press/v70/bolukbasi17a.html>
AUTHORS: Tolga Bolukbasi, Joseph Wang, Ofer Dekel, Venkatesh Saligrama
HIGHLIGHT: We present an approach to adaptively utilize deep neural networks in order to reduce the evaluation time on new examples without loss of accuracy.
- 56, TITLE: Compressed Sensing using Generative Models
<http://proceedings.mlr.press/v70/bora17a.html>
AUTHORS: Ashish Bora, Ajil Jalal, Eric Price, Alexandros G. Dimakis
HIGHLIGHT: We show how to achieve guarantees similar to standard compressed sensing but without employing sparsity at all.
- 57, TITLE: Programming with a Differentiable Forth Interpreter
<http://proceedings.mlr.press/v70/bosnjak17a.html>
AUTHORS: Matko Bošnjak, Tim Rocktäschel, Jason Naradowsky, Sebastian Riedel
HIGHLIGHT: In this paper, we consider the case of prior procedural knowledge for neural networks, such as knowing how a program should traverse a sequence, but not what local actions should be performed at each step.
- 58, TITLE: Practical Gauss-Newton Optimisation for Deep Learning
<http://proceedings.mlr.press/v70/botev17a.html>
AUTHORS: Aleksandar Botev, Hippolyt Ritter, David Barber
HIGHLIGHT: We present an efficient block-diagonal approximation to the Gauss-Newton matrix for feedforward neural networks.
- 59, TITLE: Lazifying Conditional Gradient Algorithms
<http://proceedings.mlr.press/v70/braun17a.html>
AUTHORS: Gabor Braun, Sebastian Pokutta, Daniel Zink
HIGHLIGHT: We show a general method to lazify various conditional gradient algorithms, which in actual computations leads to several orders of magnitude of speedup in wall-clock time.
- 60, TITLE: Clustering High Dimensional Dynamic Data Streams
<http://proceedings.mlr.press/v70/braverman17a.html>
AUTHORS: Vladimir Braverman, Gereon Frahling, Harry Lang, Christian Sohler, Lin F. Yang
HIGHLIGHT: We present data streaming algorithms for the k -median problem in high-dimensional dynamic geometric data streams, i.e. streams allowing both insertions and deletions of points from a discrete Euclidean space \mathbb{R}^d .
- 61, TITLE: On the Sampling Problem for Kernel Quadrature
<http://proceedings.mlr.press/v70/briol17a.html>
AUTHORS: François-Xavier Briol, Chris J. Oates, Jon Cockayne, Wilson Ye Chen, Mark Girolami
HIGHLIGHT: This paper argues that the practical choice of sampling distribution is an important open problem.
- 62, TITLE: Reduced Space and Faster Convergence in Imperfect-Information Games via Pruning
<http://proceedings.mlr.press/v70/brown17a.html>
AUTHORS: Noam Brown, Tuomas Sandholm

HIGHLIGHT: In this paper we introduce Best-Response Pruning (BRP), an improvement to iterative algorithms such as CFR that allows poorly-performing actions to be temporarily pruned.

63, **TITLE:** Globally Optimal Gradient Descent for a ConvNet with Gaussian Inputs
<http://proceedings.mlr.press/v70/brutzkus17a.html>
AUTHORS: Alon Brutzkus, Amir Globerson
HIGHLIGHT: Deep learning models are often successfully trained using gradient descent, despite the worst case hardness of the underlying non-convex optimization problem.

64, **TITLE:** Deep Tensor Convolution on Multicores
<http://proceedings.mlr.press/v70/budden17a.html>
AUTHORS: David Budden, Alexander Matveev, Shibani Santurkar, Shraman Ray Chaudhuri, Nir Shavit
HIGHLIGHT: Here we extend and optimize the faster Winograd-class of convolutional algorithms to the N^2 -dimensional case and specifically for CPU hardware.

65, **TITLE:** Multi-objective Bandits: Optimizing the Generalized Gini Index
<http://proceedings.mlr.press/v70/busa-fekete17a.html>
AUTHORS: R?bert Busa-Fekete, Bal?zs Sz?r?nyi, Paul Weng, Shie Mannor
HIGHLIGHT: We propose an online gradient descent algorithm which exploits the convexity of the GGI aggregation function, and controls the exploration in a careful way achieving a distribution-free regret $\tilde{O}(T^{1/2})$ with high probability.

66, **TITLE:** Priv?IT: Private and Sample Efficient Identity Testing
<http://proceedings.mlr.press/v70/cai17a.html>
AUTHORS: Bryan Cai, Constantinos Daskalakis, Gautam Kamath
HIGHLIGHT: We develop differentially private hypothesis testing methods for the small sample regime.

67, **TITLE:** Second-Order Kernel Online Convex Optimization with Adaptive Sketching
<http://proceedings.mlr.press/v70/calandriello17a.html>
AUTHORS: Daniele Calandriello, Alessandro Lazaric, Michal Valko
HIGHLIGHT: In this paper, we introduce kernel online Newton step (KONS), a new second-order KOCO method that also achieves $O(\text{deff}\log T)$ regret.

68, **TITLE:** ?Convex Until Proven Guilty?: Dimension-Free Acceleration of Gradient Descent on Non-Convex Functions
<http://proceedings.mlr.press/v70/carmon17a.html>
AUTHORS: Yair Carmon, John C. Duchi, Oliver Hinder, Aaron Sidford
HIGHLIGHT: We develop and analyze a variant of Nesterov's accelerated gradient descent (AGD) for minimization of smooth non-convex functions.

69, **TITLE:** Sliced Wasserstein Kernel for Persistence Diagrams
<http://proceedings.mlr.press/v70/carriere17a.html>
AUTHORS: Mathieu Carri?re, Marco Cuturi, Steve Oudot
HIGHLIGHT: In this article, we use the Sliced Wasserstein approximation of the Wasserstein distance to define a new kernel for PDs, which is not only provably stable but also provably discriminative w.r.t. the Wasserstein distance W^1_{∞} between PDs.

70, **TITLE:** Multiple Clustering Views from Multiple Uncertain Experts
<http://proceedings.mlr.press/v70/chang17a.html>
AUTHORS: Yale Chang, Junxiang Chen, Michael H. Cho, Peter J. Castaldi, Edwin K. Silverman, Jennifer G. Dy
HIGHLIGHT: In this paper, we address the problem on how to automatically discover multiple ways to cluster data given potentially diverse inputs from multiple uncertain experts.

- 71, TITLE: Uncertainty Assessment and False Discovery Rate Control in High-Dimensional Granger Causal Inference
<http://proceedings.mlr.press/v70/chaudhry17a.html>
AUTHORS: Aditya Chaudhry, Pan Xu, Quanquan Gu
HIGHLIGHT: We make two contributions in this work.
- 72, TITLE: Active Heteroscedastic Regression
<http://proceedings.mlr.press/v70/chaudhuri17a.html>
AUTHORS: Kamalika Chaudhuri, Prateek Jain, Nagarajan Natarajan
HIGHLIGHT: In this work, we consider a theoretical analysis of the label requirement of active learning for regression under a heteroscedastic noise model, where the noise depends on the instance.
- 73, TITLE: Combining Model-Based and Model-Free Updates for Trajectory-Centric Reinforcement Learning
<http://proceedings.mlr.press/v70/chebotar17a.html>
AUTHORS: Yevgen Chebotar, Karol Hausman, Marvin Zhang, Gaurav Sukhatme, Stefan Schaal, Sergey Levine
HIGHLIGHT: In this work, we aim to combine the advantages of these approaches.
- 74, TITLE: Robust Structured Estimation with Single-Index Models
<http://proceedings.mlr.press/v70/chen17a.html>
AUTHORS: Sheng Chen, Arindam Banerjee
HIGHLIGHT: In this paper, we investigate general single-index models (SIMs) in high dimensions.
- 75, TITLE: Adaptive Multiple-Arm Identification
<http://proceedings.mlr.press/v70/chen17b.html>
AUTHORS: Jiecao Chen, Xi Chen, Qin Zhang, Yuan Zhou
HIGHLIGHT: Our goal is to develop a PAC algorithm, which, with probability at least $1-\delta$, identifies a set of K arms with the aggregate regret at most ϵ .
- 76, TITLE: Dueling Bandits with Weak Regret
<http://proceedings.mlr.press/v70/chen17c.html>
AUTHORS: Bangrui Chen, Peter I. Frazier
HIGHLIGHT: We propose a new algorithm for this problem, Winner Stays (WS), with variations for each kind of regret: WS for weak regret (WS-W) has expected cumulative weak regret that is $O(N^2)$, and $O(N\log(N))$ if arms have a total order; WS for strong regret (WS-S) has expected cumulative strong regret of $O(N^2 + N\log(T))$, and $O(N\log(N)+N\log(T))$ if arms have a total order.
- 77, TITLE: Strong NP-Hardness for Sparse Optimization with Concave Penalty Functions
<http://proceedings.mlr.press/v70/chen17d.html>
AUTHORS: Yichen Chen, Dongdong Ge, Mengdi Wang, Zizhuo Wang, Yinyu Ye, Hao Yin
HIGHLIGHT: We prove that finding an $\{O\}(n^{c_1}d^{c_2})$ -optimal solution to the regularized sparse optimization problem is strongly NP-hard for any $c_1, c_2 \in [0,1)$ such that $c_1+c_2 \geq 1$.
- 78, TITLE: Learning to Learn without Gradient Descent by Gradient Descent
<http://proceedings.mlr.press/v70/chen17e.html>
AUTHORS: Yutian Chen, Matthew W. Hoffman, Sergio Gomez Colmenarejo, Misha Denil, Timothy P. Lillicrap, Matt Botvinick, Nando Freitas
HIGHLIGHT: We learn recurrent neural network optimizers trained on simple synthetic functions by gradient descent.
- 79, TITLE: Identification and Model Testing in Linear Structural Equation Models using Auxiliary Variables
<http://proceedings.mlr.press/v70/chen17f.html>
AUTHORS: Bryant Chen, Daniel Kumor, Elias Bareinboim

HIGHLIGHT: In this paper, we provide an algorithm for the identification of causal parameters in linear structural models that subsumes previous state-of-the-art methods.

80, **TITLE:** Toward Efficient and Accurate Covariance Matrix Estimation on Compressed Data
<http://proceedings.mlr.press/v70/chen17g.html>

AUTHORS: Xixian Chen, Michael R. Lyu, Irwin King

HIGHLIGHT: To tackle the challenges of extensive communication costs, large storage capacity requirements, and high processing time complexity when handling massive high-dimensional and distributed data, we propose an efficient and accurate covariance matrix estimation method via data compression.

81, **TITLE:** Online Partial Least Square Optimization: Dropping Convexity for Better Efficiency and Scalability

<http://proceedings.mlr.press/v70/chen17h.html>

AUTHORS: Zhehui Chen, Lin F. Yang, Chris Junchi Li, Tuo Zhao

HIGHLIGHT: Many existing approaches formulate the multiview representation learning as convex optimization problems, where global optima can be obtained by certain algorithms in polynomial time.

82, **TITLE:** Learning to Aggregate Ordinal Labels by Maximizing Separating Width

<http://proceedings.mlr.press/v70/chen17i.html>

AUTHORS: Guangyong Chen, Shengyu Zhang, Di Lin, Hui Huang, Pheng Ann Heng

HIGHLIGHT: Based on a technique of sampling estimated label from the posterior distribution, we define a novel separating width among the labeled observations to characterize the quality of sampled labels, and develop an efficient algorithm to optimize it through solving multiple linear decision boundaries and adjusting prior distributions.

83, **TITLE:** Nearly Optimal Robust Matrix Completion

<http://proceedings.mlr.press/v70/cherapanamjeri17a.html>

AUTHORS: Yeshwanth Cherapanamjeri, Kartik Gupta, Prateek Jain

HIGHLIGHT: In this paper, we consider the problem of Robust Matrix Completion (RMC) where the goal is to recover a low-rank matrix by observing a small number of its entries out of which a few can be arbitrarily corrupted.

84, **TITLE:** Algorithms for ℓ_p Low-Rank Approximation

<http://proceedings.mlr.press/v70/chierichetti17a.html>

AUTHORS: Flavio Chierichetti, Sreenivas Gollapudi, Ravi Kumar, Silvio Lattanzi, Rina Panigrahy, David P. Woodruff

HIGHLIGHT: We consider the problem of approximating a given matrix by a low-rank matrix so as to minimize the entrywise ℓ_p -approximation error, for any $p \geq 1$; the case $p = 2$ is the classical SVD problem.

85, **TITLE:** MEC: Memory-efficient Convolution for Deep Neural Network

<http://proceedings.mlr.press/v70/cho17a.html>

AUTHORS: Minsik Cho, Daniel Brand

HIGHLIGHT: In this work, we propose a memory-efficient convolution or MEC with compact lowering, which reduces memory overhead substantially and accelerates convolution process.

86, **TITLE:** On Relaxing Determinism in Arithmetic Circuits

<http://proceedings.mlr.press/v70/choi17a.html>

AUTHORS: Arthur Choi, Adnan Darwiche

HIGHLIGHT: In this paper, we provide a formal basis under which variants on ACs can be compared, and where the precise roles and semantics of their various properties can be made more transparent.

87, **TITLE:** Improving Stochastic Policy Gradients in Continuous Control with Deep Reinforcement Learning using the Beta Distribution

<http://proceedings.mlr.press/v70/chou17a.html>

AUTHORS: Po-Wei Chou, Daniel Maturana, Sebastian Scherer

HIGHLIGHT: In this work, we propose to use the Beta distribution as an alternative and analyze the bias and variance of the policy gradients of both policies.

88, **TITLE:** On Kernelized Multi-armed Bandits
<http://proceedings.mlr.press/v70/chowdhury17a.html>
AUTHORS: Sayak Ray Chowdhury, Aditya Gopalan
HIGHLIGHT: We consider the stochastic bandit problem with a continuous set of arms, with the expected reward function over the arms assumed to be fixed but unknown.

89, **TITLE:** Parseval Networks: Improving Robustness to Adversarial Examples
<http://proceedings.mlr.press/v70/cisse17a.html>
AUTHORS: Moustapha Cisse, Piotr Bojanowski, Edouard Grave, Yann Dauphin, Nicolas Usunier
HIGHLIGHT: We introduce Parseval networks, a form of deep neural networks in which the Lipschitz constant of linear, convolutional and aggregation layers is constrained to be smaller than $\$1\$$.

90, **TITLE:** Deep Latent Dirichlet Allocation with Topic-Layer-Adaptive Stochastic Gradient Riemannian MCMC
<http://proceedings.mlr.press/v70/cong17a.html>
AUTHORS: Yulai Cong, Bo Chen, Hongwei Liu, Mingyuan Zhou
HIGHLIGHT: Exploiting that Fisher information matrix with stochastic gradient MCMC, we present topic-layer-adaptive stochastic gradient Riemannian (TLASGR) MCMC that jointly learns simplex-constrained global parameters across all layers and topics, with topic and layer specific learning rates.

91, **TITLE:** AdaNet: Adaptive Structural Learning of Artificial Neural Networks
<http://proceedings.mlr.press/v70/cortes17a.html>
AUTHORS: Corinna Cortes, Xavier Gonzalvo, Vitaly Kuznetsov, Mehryar Mohri, Scott Yang
HIGHLIGHT: We present a new framework for analyzing and learning artificial neural networks.

92, **TITLE:** Random Feature Expansions for Deep Gaussian Processes
<http://proceedings.mlr.press/v70/cutajar17a.html>
AUTHORS: Kurt Cutajar, Edwin V. Bonilla, Pietro Michiardi, Maurizio Filippone
HIGHLIGHT: In this work we introduce a novel formulation of DGPs based on random feature expansions that we train using stochastic variational inference.

93, **TITLE:** Soft-DTW: a Differentiable Loss Function for Time-Series
<http://proceedings.mlr.press/v70/cuturi17a.html>
AUTHORS: Marco Cuturi, Mathieu Blondel
HIGHLIGHT: We propose in this paper a differentiable learning loss between time series, building upon the celebrated dynamic time warping (DTW) discrepancy.

94, **TITLE:** Understanding Synthetic Gradients and Decoupled Neural Interfaces
<http://proceedings.mlr.press/v70/czarnecki17a.html>
AUTHORS: Wojciech Marian Czarnecki, Grzegorz Swirszcz, Max Jaderberg, Simon Osindero, Oriol Vinyals, Koray Kavukcuoglu
HIGHLIGHT: In this paper, we study DNIs through the use of synthetic gradients on feed-forward networks to better understand their behaviour and elucidate their effect on optimisation.

95, **TITLE:** Stochastic Generative Hashing
<http://proceedings.mlr.press/v70/dai17a.html>
AUTHORS: Bo Dai, Ruiqi Guo, Sanjiv Kumar, Niao He, Le Song
HIGHLIGHT: In this paper, we propose a novel generative approach to learn hash functions through Minimum Description Length principle such that the learned hash codes maximally compress the dataset and can also be used to regenerate the inputs.

- 96, TITLE: Logarithmic Time One-Against-Some
<http://proceedings.mlr.press/v70/daume17a.html>
AUTHORS: Hal Daum[?] III, Nikos Karampatziakis, John Langford, Paul Mineiro
HIGHLIGHT: We create a new online reduction of multiclass classification to binary classification for which training and prediction time scale logarithmically with the number of classes.
- 97, TITLE: Language Modeling with Gated Convolutional Networks
<http://proceedings.mlr.press/v70/dauphin17a.html>
AUTHORS: Yann N. Dauphin, Angela Fan, Michael Auli, David Grangier
HIGHLIGHT: In this paper we develop a finite context approach through stacked convolutions, which can be more efficient since they allow parallelization over sequential tokens.
- 98, TITLE: An Infinite Hidden Markov Model With Similarity-Biased Transitions
<http://proceedings.mlr.press/v70/dawson17a.html>
AUTHORS: Colin Reimer Dawson, Chaofan Huang, Clayton T. Morrison
HIGHLIGHT: We describe a generalization of the Hierarchical Dirichlet Process Hidden Markov Model (HDP-HMM) which is able to encode prior information that state transitions are more likely between “nearby” states.
- 99, TITLE: Distributed Batch Gaussian Process Optimization
<http://proceedings.mlr.press/v70/daxberger17a.html>
AUTHORS: Erik A. Daxberger, Bryan Kian Hsiang Low
HIGHLIGHT: This paper presents a novel distributed batch Gaussian process upper confidence bound (DB-GP-UCB) algorithm for performing batch Bayesian optimization (BO) of highly complex, costly-to-evaluate black-box objective functions.
- 100, TITLE: Consistency Analysis for Binary Classification Revisited
<http://proceedings.mlr.press/v70/dembczynski17a.html>
AUTHORS: Krzysztof Dembczynski, Wojciech Kotlowski, Oluwasanmi Koyejo, Nagarajan Natarajan
HIGHLIGHT: In this manuscript we analyze both settings, from statistical and algorithmic points of view, to explore the connections and to highlight differences between them for a wide range of metrics.
- 101, TITLE: iSurvive: An Interpretable, Event-time Prediction Model for mHealth
<http://proceedings.mlr.press/v70/dempsey17a.html>
AUTHORS: Walter H. Dempsey, Alexander Moreno, Christy K. Scott, Michael L. Dennis, David H. Gustafson, Susan A. Murphy, James M. Rehg
HIGHLIGHT: We present a parameter learning method for GLM emissions and survival model fitting, and present promising results on both synthetic data and an mHealth drug use dataset.
- 102, TITLE: Image-to-Markup Generation with Coarse-to-Fine Attention
<http://proceedings.mlr.press/v70/deng17a.html>
AUTHORS: Yuntian Deng, Anssi Kanervisto, Jeffrey Ling, Alexander M. Rush
HIGHLIGHT: We present a neural encoder-decoder model to convert images into presentational markup based on a scalable coarse-to-fine attention mechanism.
Our method is evaluated in the context of image-to-LaTeX generation, and we introduce a new dataset of real-world rendered mathematical expressions paired with LaTeX markup.
- 103, TITLE: RobustFill: Neural Program Learning under Noisy I/O
<http://proceedings.mlr.press/v70/devlin17a.html>
AUTHORS: Jacob Devlin, Jonathan Uesato, Surya Bhupatiraju, Rishabh Singh, Abdel-rahman Mohamed, Pushmeet Kohli
HIGHLIGHT: Here, for the first time, we directly compare both approaches on a large-scale, real-world learning task and we additionally contrast to rule-based program synthesis, which uses hand-crafted semantics to guide the program generation.

- 104, TITLE: Being Robust (in High Dimensions) Can Be Practical
<http://proceedings.mlr.press/v70/diakonikolas17a.html>
AUTHORS: Ilias Diakonikolas, Gautam Kamath, Daniel M. Kane, Jerry Li, Ankur Moitra, Alistair Stewart
HIGHLIGHT: In this work, we address both of these issues by establishing sample complexity bounds that are optimal, up to logarithmic factors, as well as giving various refinements that allow the algorithms to tolerate a much larger fraction of corruptions.
- 105, TITLE: Probabilistic Path Hamiltonian Monte Carlo
<http://proceedings.mlr.press/v70/dinh17a.html>
AUTHORS: Vu Dinh, Arman Bilge, Cheng Zhang, Frederick A. Matsen IV
HIGHLIGHT: In this paper, we develop Probabilistic Path HMC (PPHMC) as a first step to sampling distributions on spaces with intricate combinatorial structure.
- 106, TITLE: Sharp Minima Can Generalize For Deep Nets
<http://proceedings.mlr.press/v70/dinh17b.html>
AUTHORS: Laurent Dinh, Razvan Pascanu, Samy Bengio, Yoshua Bengio
HIGHLIGHT: Specifically, when focusing on deep networks with rectifier units, we can exploit the particular geometry of parameter space induced by the inherent symmetries that these architectures exhibit to build equivalent models corresponding to arbitrarily sharper minima.
- 107, TITLE: A Divergence Bound for Hybrids of MCMC and Variational Inference and an Application to Langevin Dynamics and SGVI
<http://proceedings.mlr.press/v70/domke17a.html>
AUTHORS: Justin Domke
HIGHLIGHT: This paper derives a distribution over variational parameters, designed to minimize a bound on the divergence between the resulting marginal distribution and the target, and gives an example of how to sample from this distribution in a way that interpolates between the behavior of existing methods based on Langevin dynamics and stochastic gradient variational inference (SGVI).
- 108, TITLE: Dance Dance Convolution
<http://proceedings.mlr.press/v70/donahue17a.html>
AUTHORS: Chris Donahue, Zachary C. Lipton, Julian McAuley
HIGHLIGHT: For step selection, we present a conditional LSTM generative model that substantially outperforms n-gram and fixed-window approaches.
- 109, TITLE: Stochastic Variance Reduction Methods for Policy Evaluation
<http://proceedings.mlr.press/v70/du17a.html>
AUTHORS: Simon S. Du, Jianshu Chen, Lihong Li, Lin Xiao, Dengyong Zhou
HIGHLIGHT: In this paper, we focus on policy evaluation with linear function approximation over a fixed dataset.
- 110, TITLE: Rule-Enhanced Penalized Regression by Column Generation using Rectangular Maximum Agreement
<http://proceedings.mlr.press/v70/eckstein17a.html>
AUTHORS: Jonathan Eckstein, Noam Goldberg, Ai Kagawa
HIGHLIGHT: We describe a learning procedure enhancing L1-penalized regression by adding dynamically generated rules describing multidimensional “box” sets.
- 111, TITLE: Neural Audio Synthesis of Musical Notes with WaveNet Autoencoders
<http://proceedings.mlr.press/v70/engel17a.html>
AUTHORS: Jesse Engel, Cinjon Resnick, Adam Roberts, Sander Dieleman, Mohammad Norouzi, Douglas Eck, Karen Simonyan
HIGHLIGHT: In this paper, we offer contributions in both these areas to enable similar progress in audio modeling.

112, TITLE: Statistical Inference for Incomplete Ranking Data: The Case of Rank-Dependent Coarsening
<http://proceedings.mlr.press/v70/fahandar17a.html>
AUTHORS: Mohsen Ahmadi Fahandar, Eyke H?llermeier, In?s Couso
HIGHLIGHT: To this end, we propose the concept of rank-dependent coarsening, which assumes that incomplete rankings are produced by projecting a full ranking to a random subset of ranks.

113, TITLE: Maximum Selection and Ranking under Noisy Comparisons
<http://proceedings.mlr.press/v70/falahatgar17a.html>
AUTHORS: Moein Falahatgar, Alon Orlitsky, Venkatadheeraj Pichapati, Ananda Theertha Suresh
HIGHLIGHT: Modifying the popular knockout tournament, we propose a simple maximum-selection algorithm that uses $\mathcal{O}\left(\frac{n}{\epsilon^2} \left(1 + \log \frac{1}{\delta}\right)\right)$ comparisons, optimal up to a constant factor.

114, TITLE: Fake News Mitigation via Point Process Based Intervention
<http://proceedings.mlr.press/v70/farajtabar17a.html>
AUTHORS: Mehrdad Farajtabar, Jiachen Yang, Xiaojing Ye, Huan Xu, Rakshit Trivedi, Elias Khalil, Shuang Li, Le Song, Hongyuan Zha
HIGHLIGHT: We propose the first multistage intervention framework that tackles fake news in social networks by combining reinforcement learning with a point process network activity model.

115, TITLE: Regret Minimization in Behaviorally-Constrained Zero-Sum Games
<http://proceedings.mlr.press/v70/farina17a.html>
AUTHORS: Gabriele Farina, Christian Kroer, Tuomas Sandholm
HIGHLIGHT: We use an instantiation of the CFR framework to develop algorithms for solving behaviorally-constrained (and, as a special case, perturbed in the Selten sense) extensive-form games, which allows us to compute approximate Nash equilibrium refinements.

116, TITLE: Coresets for Vector Summarization with Applications to Network Graphs
<http://proceedings.mlr.press/v70/feldman17a.html>
AUTHORS: Dan Feldman, Sedat Ozer, Daniela Rus
HIGHLIGHT: We provide a deterministic data summarization algorithm that approximates the mean $\bar{p} = \frac{1}{n} \sum_{p \in P} p$ of a set P of n vectors in \mathbb{R}^d , by a weighted mean \tilde{p} of a subset of $O(1/\epsilon)$ vectors, i.e., independent of both n and d .

117, TITLE: Model-Agnostic Meta-Learning for Fast Adaptation of Deep Networks
<http://proceedings.mlr.press/v70/finn17a.html>
AUTHORS: Chelsea Finn, Pieter Abbeel, Sergey Levine
HIGHLIGHT: We propose an algorithm for meta-learning that is model-agnostic, in the sense that it is compatible with any model trained with gradient descent and applicable to a variety of different learning problems, including classification, regression, and reinforcement learning.

118, TITLE: Input Switched Affine Networks: An RNN Architecture Designed for Interpretability
<http://proceedings.mlr.press/v70/foerster17a.html>
AUTHORS: Jakob N. Foerster, Justin Gilmer, Jascha Sohl-Dickstein, Jan Chorowski, David Sussillo
HIGHLIGHT: Here we introduce a recurrent architecture composed of input-switched affine transformations – in other words an RNN without any explicit nonlinearities, but with input-dependent recurrent weights.

119, TITLE: Stabilising Experience Replay for Deep Multi-Agent Reinforcement Learning
<http://proceedings.mlr.press/v70/foerster17b.html>
AUTHORS: Jakob Foerster, Nantas Nardelli, Gregory Farquhar, Triantafyllos Afouras, Philip H. S. Torr, Pushmeet Kohli, Shimon Whiteson

HIGHLIGHT: This paper proposes two methods that address this problem: 1) using a multi-agent variant of importance sampling to naturally decay obsolete data and 2) conditioning each agent's value function on a fingerprint that disambiguates the age of the data sampled from the replay memory.

120, **TITLE:** Counterfactual Data-Fusion for Online Reinforcement Learners
<http://proceedings.mlr.press/v70/forney17a.html>

AUTHORS: Andrew Forney, Judea Pearl, Elias Bareinboim

HIGHLIGHT: In this paper, we show how counterfactual-based decision-making circumvents these problems and leads to a coherent fusion of observational and experimental data.

121, **TITLE:** Forward and Reverse Gradient-Based Hyperparameter Optimization
<http://proceedings.mlr.press/v70/franceschi17a.html>

AUTHORS: Luca Franceschi, Michele Donini, Paolo Frasconi, Massimiliano Pontil

HIGHLIGHT: We present a series of experiments on image and phone classification tasks.

122, **TITLE:** Learning to Detect Sepsis with a Multitask Gaussian Process RNN Classifier
<http://proceedings.mlr.press/v70/futoma17a.html>

AUTHORS: Joseph Futoma, Sanjay Hariharan, Katherine Heller

HIGHLIGHT: We present a scalable end-to-end classifier that uses streaming physiological and medication data to accurately predict the onset of sepsis, a life-threatening complication from infections that has high mortality and morbidity.

123, **TITLE:** Deep Bayesian Active Learning with Image Data
<http://proceedings.mlr.press/v70/gal17a.html>

AUTHORS: Yarin Gal, Riashat Islam, Zoubin Ghahramani

HIGHLIGHT: In this paper we combine recent advances in Bayesian deep learning into the active learning framework in a practical way.

124, **TITLE:** Local-to-Global Bayesian Network Structure Learning
<http://proceedings.mlr.press/v70/gao17a.html>

AUTHORS: Tian Gao, Kshitij Fadnis, Murray Campbell

HIGHLIGHT: We introduce a new local-to-global structure learning algorithm, called graph growing structure learning (GGSL), to learn Bayesian network (BN) structures.

125, **TITLE:** Communication-efficient Algorithms for Distributed Stochastic Principal Component Analysis
<http://proceedings.mlr.press/v70/garber17a.html>

AUTHORS: Dan Garber, Ohad Shamir, Nathan Srebro

HIGHLIGHT: We study the fundamental problem of Principal Component Analysis in a statistical distributed setting in which each machine out of m stores a sample of n points sampled i.i.d. from a single unknown distribution.

126, **TITLE:** Differentiable Programs with Neural Libraries
<http://proceedings.mlr.press/v70/gaunt17a.html>

AUTHORS: Alexander L. Gaunt, Marc Brockschmidt, Nate Kushman, Daniel Tarlow

HIGHLIGHT: We develop a framework for combining differentiable programming languages with neural networks.

127, **TITLE:** Zonotope Hit-and-run for Efficient Sampling from Projection DPPs
<http://proceedings.mlr.press/v70/gautier17a.html>

AUTHORS: Guillaume Gautier, R?mi Bardenet, Michal Valko

HIGHLIGHT: We build a novel MCMC sampler that combines ideas from combinatorial geometry, linear programming, and Monte Carlo methods to sample from DPPs with a fixed sample cardinality, also called projection DPPs.

- 128, TITLE: No Spurious Local Minima in Nonconvex Low Rank Problems: A Unified Geometric Analysis
<http://proceedings.mlr.press/v70/ge17a.html>
AUTHORS: Rong Ge, Chi Jin, Yi Zheng
HIGHLIGHT: In this paper we develop a new framework that captures the common landscape underlying the common non-convex low-rank matrix problems including matrix sensing, matrix completion and robust PCA.
- 129, TITLE: Convolutional Sequence to Sequence Learning
<http://proceedings.mlr.press/v70/gehring17a.html>
AUTHORS: Jonas Gehring, Michael Auli, David Grangier, Denis Yarats, Yann N. Dauphin
HIGHLIGHT: We introduce an architecture based entirely on convolutional neural networks.
- 130, TITLE: On Context-Dependent Clustering of Bandits
<http://proceedings.mlr.press/v70/gentile17a.html>
AUTHORS: Claudio Gentile, Shuai Li, Purushottam Kar, Alexandros Karatzoglou, Giovanni Zappella, Evans Etrue
HIGHLIGHT: We investigate a novel cluster-of-bandit algorithm CAB for collaborative recommendation tasks that implements the underlying feedback sharing mechanism by estimating user neighborhoods in a context-dependent manner.
- 131, TITLE: Neural Message Passing for Quantum Chemistry
<http://proceedings.mlr.press/v70/gilmer17a.html>
AUTHORS: Justin Gilmer, Samuel S. Schoenholz, Patrick F. Riley, Oriol Vinyals, George E. Dahl
HIGHLIGHT: In this paper, we reformulate existing models into a single common framework we call Message Passing Neural Networks (MPNNs) and explore additional novel variations within this framework.
- 132, TITLE: Convex Phase Retrieval without Lifting via PhaseMax
<http://proceedings.mlr.press/v70/goldstein17a.html>
AUTHORS: Tom Goldstein, Christoph Studer
HIGHLIGHT: We compare our approach to other phase retrieval methods and demonstrate that our theory accurately predicts the success of PhaseMax.
- 133, TITLE: Preferential Bayesian Optimization
<http://proceedings.mlr.press/v70/gonzalez17a.html>
AUTHORS: Javier Gonzalez, Zhenwen Dai, Andreas Damianou, Neil D. Lawrence
HIGHLIGHT: We present a new framework for this scenario that we call Preferential Bayesian Optimization (PBO) and that allows to find the optimum of a latent function that can only be queried through pairwise comparisons, so-called duels.
- 134, TITLE: Measuring Sample Quality with Kernels
<http://proceedings.mlr.press/v70/gorham17a.html>
AUTHORS: Jackson Gorham, Lester Mackey
HIGHLIGHT: We develop a theory of weak convergence for KSDs based on Stein's method, demonstrate that commonly used KSDs fail to detect non-convergence even for Gaussian targets, and show that kernels with slowly decaying tails provably determine convergence for a large class of target distributions.
- 135, TITLE: Efficient softmax approximation for GPUs
<http://proceedings.mlr.press/v70/grave17a.html>
AUTHORS: Grave, Armand Joulin, Moustapha Cissé, David Grangier, Hervé Jégou
HIGHLIGHT: We propose an approximate strategy to efficiently train neural network based language models over very large vocabularies.
- 136, TITLE: Automated Curriculum Learning for Neural Networks
<http://proceedings.mlr.press/v70/graves17a.html>
AUTHORS: Alex Graves, Marc G. Bellemare, Jacob Menick, Rami Munos, Koray Kavukcuoglu

HIGHLIGHT: We introduce a method for automatically selecting the path, or syllabus, that a neural network follows through a curriculum so as to maximise learning efficiency.

137, **TITLE:** On Calibration of Modern Neural Networks
<http://proceedings.mlr.press/v70/guo17a.html>
AUTHORS: Chuan Guo, Geoff Pleiss, Yu Sun, Kilian Q. Weinberger
HIGHLIGHT: We discover that modern neural networks, unlike those from a decade ago, are poorly calibrated.

138, **TITLE:** ProtoNN: Compressed and Accurate kNN for Resource-scarce Devices
<http://proceedings.mlr.press/v70/gupta17a.html>
AUTHORS: Chirag Gupta, Arun Sai Suggala, Ankit Goyal, Harsha Vardhan Simhadri, Bhargavi Paranjape, Ashish Kumar, Saurabh Goyal, Raghavendra Udupa, Manik Varma, Prateek Jain
HIGHLIGHT: In this work, we propose ProtoNN, a novel algorithm that addresses the problem of real-time and accurate prediction on resource-scarce devices.

139, **TITLE:** Deep Value Networks Learn to Evaluate and Iteratively Refine Structured Outputs
<http://proceedings.mlr.press/v70/gygli17a.html>
AUTHORS: Michael Gygli, Mohammad Norouzi, Anelia Angelova
HIGHLIGHT: We approach structured output prediction by optimizing a deep value network (DVN) to precisely estimate the task loss on different output configurations for a given input.

140, **TITLE:** Reinforcement Learning with Deep Energy-Based Policies
<http://proceedings.mlr.press/v70/haarnojal17a.html>
AUTHORS: Tuomas Haarnoja, Haoran Tang, Pieter Abbeel, Sergey Levine
HIGHLIGHT: We propose a method for learning expressive energy-based policies for continuous states and actions, which has been feasible only in tabular domains before.

141, **TITLE:** DeepBach: a Steerable Model for Bach Chorales Generation
<http://proceedings.mlr.press/v70/hadjeres17a.html>
AUTHORS: Ga?tan Hadjeres, Fran?ois Pachet, Frank Nielsen
HIGHLIGHT: This paper introduces DeepBach, a graphical model aimed at modeling polyphonic music and specifically hymn-like pieces.

142, **TITLE:** Consistent On-Line Off-Policy Evaluation
<http://proceedings.mlr.press/v70/hallak17a.html>
AUTHORS: Assaf Hallak, Shie Mannor
HIGHLIGHT: In this paper we propose the Consistent Off-Policy Temporal Difference (COP-TD(λ , β)) algorithm that addresses this issue and reduces this bias at some computational expense.

143, **TITLE:** Faster Greedy MAP Inference for Determinantal Point Processes
<http://proceedings.mlr.press/v70/han17a.html>
AUTHORS: Insu Han, Prabhanjan Kambadur, Kyoungsoo Park, Jinwoo Shin
HIGHLIGHT: In this paper, we develop fast algorithms to find the most likely configuration (MAP) of large-scale DPPs, which is NP-hard in general.

144, **TITLE:** Data-Efficient Policy Evaluation Through Behavior Policy Search
<http://proceedings.mlr.press/v70/hanna17a.html>
AUTHORS: Josiah P. Hanna, Philip S. Thomas, Peter Stone, Scott Niekum
HIGHLIGHT: We present a behavior policy search algorithm and empirically demonstrate its effectiveness in lowering the mean squared error of policy performance estimates.

145, **TITLE:** Joint Dimensionality Reduction and Metric Learning: A Geometric Take
<http://proceedings.mlr.press/v70/harandi17a.html>

AUTHORS: Mehrtash Harandi, Mathieu Salzmann, Richard Hartley
HIGHLIGHT: To address this issue, in this paper, we develop a Riemannian framework to jointly learn a mapping performing dimensionality reduction and a metric in the induced space.

146, TITLE: Deep IV: A Flexible Approach for Counterfactual Prediction
<http://proceedings.mlr.press/v70/hartford17a.html>
AUTHORS: Jason Hartford, Greg Lewis, Kevin Leyton-Brown, Matt Taddy
HIGHLIGHT: This paper provides a recipe for augmenting deep learning methods to accurately characterize such relationships in the presence of instrument variables (IVs) – sources of treatment randomization that are conditionally independent from the outcomes.

147, TITLE: Robust Guarantees of Stochastic Greedy Algorithms
<http://proceedings.mlr.press/v70/hassidim17a.html>
AUTHORS: Avinatan Hassidim, Yaron Singer
HIGHLIGHT: In this paper we analyze the robustness of stochastic variants of the greedy algorithm for submodular maximization.

148, TITLE: Efficient Regret Minimization in Non-Convex Games
<http://proceedings.mlr.press/v70/hazan17a.html>
AUTHORS: Elad Hazan, Karan Singh, Cyril Zhang
HIGHLIGHT: We give gradient-based methods that achieve optimal regret, which in turn guarantee convergence to equilibrium in this framework.

149, TITLE: Kernelized Support Tensor Machines
<http://proceedings.mlr.press/v70/he17a.html>
AUTHORS: Lifang He, Chun-Ta Lu, Guixiang Ma, Shen Wang, Linlin Shen, Philip S. Yu, Ann B. Ragin
HIGHLIGHT: Based on tensor factorization theory and kernel methods, we propose a novel Kernelized Support Tensor Machine (KSTM) which integrates kernelized tensor factorization with maximum-margin criterion.

150, TITLE: The Sample Complexity of Online One-Class Collaborative Filtering
<http://proceedings.mlr.press/v70/heckel17a.html>
AUTHORS: Reinhard Heckel, Kannan Ramchandran
HIGHLIGHT: We introduce a simple probabilistic user model, and analyze the performance of an online user-based CF algorithm.

151, TITLE: Warped Convolutions: Efficient Invariance to Spatial Transformations
<http://proceedings.mlr.press/v70/henriques17a.html>
AUTHORS: Jo?o F. Henriques, Andrea Vedaldi
HIGHLIGHT: We present a construction that is simple and exact, yet has the same computational complexity that standard convolutions enjoy.

152, TITLE: Parallel and Distributed Thompson Sampling for Large-scale Accelerated Exploration of Chemical Space
<http://proceedings.mlr.press/v70/hernandez-lobato17a.html>
AUTHORS: Jos? Miguel Hern?andez-Lobato, James Requeima, Edward O. Pyzer-Knapp, Al?n Aspuru-Guzik
HIGHLIGHT: Here, we propose a scalable solution based on a parallel and distributed implementation of Thompson sampling (PDTS).

153, TITLE: DARLA: Improving Zero-Shot Transfer in Reinforcement Learning
<http://proceedings.mlr.press/v70/higgins17a.html>
AUTHORS: Irina Higgins, Arka Pal, Andrei Rusu, Loic Matthey, Christopher Burgess, Alexander Pritzel, Matthew Botvinick, Charles Blundell, Alexander Lerchner
HIGHLIGHT: We propose a new multi-stage RL agent, DARLA (Disentangled Representation Learning Agent), which learns to see before learning to act.

154, TITLE: SPLICE: Fully Tractable Hierarchical Extension of ICA with Pooling
<http://proceedings.mlr.press/v70/hirayama17a.html>
AUTHORS: Jun-ichiro Hirayama, Aapo Hyv?rinen, Motoaki Kawanabe
HIGHLIGHT: We present a novel probabilistic framework for a hierarchical extension of independent component analysis (ICA), with a particular motivation in neuroscientific data analysis and modeling.

155, TITLE: Multilevel Clustering via Wasserstein Means
<http://proceedings.mlr.press/v70/ho17a.html>
AUTHORS: Nhat Ho, XuanLong Nguyen, Mikhail Yurochkin, Hung Hai Bui, Viet Huynh, Dinh Phung
HIGHLIGHT: We propose a novel approach to the problem of multilevel clustering, which aims to simultaneously partition data in each group and discover grouping patterns among groups in a potentially large hierarchically structured corpus of data.

156, TITLE: Learning Deep Latent Gaussian Models with Markov Chain Monte Carlo
<http://proceedings.mlr.press/v70/hoffman17a.html>
AUTHORS: Matthew D. Hoffman
HIGHLIGHT: In this paper, we propose a different approach: rather than use a variational approximation (which produces biased gradient signals), we use Markov chain Monte Carlo (MCMC, which allows us to trade bias for computation).

157, TITLE: Minimizing Trust Leaks for Robust Sybil Detection
<http://proceedings.mlr.press/v70/honer17a.html>
AUTHORS: J?nos H?ner, Shinichi Nakajima, Alexander Bauer, Klaus-Robert M?ller, Nico G?rnitz
HIGHLIGHT: After that, we formally introduce adversarial settings for the graph-based Sybil detection problem and derive a corresponding optimal attacking strategy by exploitation of trust leaks.

158, TITLE: Prox-PDA: The Proximal Primal-Dual Algorithm for Fast Distributed Nonconvex Optimization and Learning Over Networks
<http://proceedings.mlr.press/v70/hong17a.html>
AUTHORS: Mingyi Hong, Davood Hajinezhad, Ming-Min Zhao
HIGHLIGHT: In this paper we consider nonconvex optimization and learning over a network of distributed nodes.

159, TITLE: Analysis and Optimization of Graph Decompositions by Lifted Multicuts
<http://proceedings.mlr.press/v70/hornakova17a.html>
AUTHORS: Andrea Horn?kov?, Jan-Hendrik Lange, Bjoern Andres
HIGHLIGHT: To find optimal decompositions defined by minimum cost lifted multicuts, we establish some properties of some facets of lifted multicut polytopes, define efficient separation procedures and apply these in a branch-and-cut algorithm.

160, TITLE: Dissipativity Theory for Nesterov?s Accelerated Method
<http://proceedings.mlr.press/v70/hu17a.html>
AUTHORS: Bin Hu, Laurent Lessard
HIGHLIGHT: In this paper, we adapt the control theoretic concept of dissipativity theory to provide a natural understanding of Nesterov?s accelerated method.

161, TITLE: Learning Discrete Representations via Information Maximizing Self-Augmented Training
<http://proceedings.mlr.press/v70/hu17b.html>
AUTHORS: Weihua Hu, Takeru Miyato, Seiya Tokui, Eiichi Matsumoto, Masashi Sugiyama
HIGHLIGHT: To this end, we propose a method called Information Maximizing Self-Augmented Training (IMSAT).

162, TITLE: State-Frequency Memory Recurrent Neural Networks

<http://proceedings.mlr.press/v70/hu17c.html>

AUTHORS: Hao Hu, Guo-Jun Qi

HIGHLIGHT: We propose the State-Frequency Memory (SFM), a novel recurrent architecture that allows to separate dynamic patterns across different frequency components and their impacts on modeling the temporal contexts of input sequences.

163, TITLE: Deep Generative Models for Relational Data with Side Information

<http://proceedings.mlr.press/v70/hu17d.html>

AUTHORS: Changwei Hu, Piyush Rai, Lawrence Carin

HIGHLIGHT: We present a probabilistic framework for overlapping community discovery and link prediction for relational data, given as a graph.

164, TITLE: Toward Controlled Generation of Text

<http://proceedings.mlr.press/v70/hu17e.html>

AUTHORS: Zhiting Hu, Zichao Yang, Xiaodan Liang, Ruslan Salakhutdinov, Eric P. Xing

HIGHLIGHT: We propose a new neural generative model which combines variational auto-encoders (VAEs) and holistic attribute discriminators for effective imposition of semantic structures.

165, TITLE: Tensor Decomposition with Smoothness

<http://proceedings.mlr.press/v70/imaizumi17a.html>

AUTHORS: Masaaki Imaizumi, Kohei Hayashi

HIGHLIGHT: To incorporate the smoothness property, we propose the smoothed Tucker decomposition (STD).

166, TITLE: Variational Inference for Sparse and Undirected Models

<http://proceedings.mlr.press/v70/ingraham17a.html>

AUTHORS: John Ingraham, Debora Marks

HIGHLIGHT: Here, we develop a framework for scalable Bayesian inference of discrete undirected models based on two new methods.

167, TITLE: Fairness in Reinforcement Learning

<http://proceedings.mlr.press/v70/jabbari17a.html>

AUTHORS: Shahin Jabbari, Matthew Joseph, Michael Kearns, Jamie Morgenstern, Aaron Roth

HIGHLIGHT: We initiate the study of fairness in reinforcement learning, where the actions of a learning algorithm may affect its environment and future rewards.

168, TITLE: Decoupled Neural Interfaces using Synthetic Gradients

<http://proceedings.mlr.press/v70/jaderberg17a.html>

AUTHORS: Max Jaderberg, Wojciech Marian Czarnecki, Simon Osindero, Oriol Vinyals, Alex Graves, David Silver, Koray Kavukcuoglu

HIGHLIGHT: In this work we break this constraint by decoupling modules by introducing a model of the future computation of the network graph.

169, TITLE: Scalable Generative Models for Multi-label Learning with Missing Labels

<http://proceedings.mlr.press/v70/jain17a.html>

AUTHORS: Vikas Jain, Nirbhay Modhe, Piyush Rai

HIGHLIGHT: We present a scalable, generative framework for multi-label learning with missing labels.

170, TITLE: Sequence Tutor: Conservative Fine-Tuning of Sequence Generation Models with KL-control

<http://proceedings.mlr.press/v70/jaques17a.html>

AUTHORS: Natasha Jaques, Shixiang Gu, Dzmitry Bahdanau, Jos? Miguel Hernandez-Lobato, Richard E. Turner, Douglas Eck

HIGHLIGHT: This paper proposes a general method for improving the structure and quality of sequences generated by a recurrent neural network (RNN), while maintaining information originally learned from data, as well as sample diversity.

- 171, TITLE: Bayesian Optimization with Tree-structured Dependencies
<http://proceedings.mlr.press/v70/jenatton17a.html>
AUTHORS: Rodolphe Jenatton, Cedric Archambeau, Javier Gonzalez, Matthias Seeger
HIGHLIGHT: In this work, we focus on use cases where this domain exhibits a known dependency structure.
- 172, TITLE: Simultaneous Learning of Trees and Representations for Extreme Classification and Density Estimation
<http://proceedings.mlr.press/v70/jernite17a.html>
AUTHORS: Yacine Jernite, Anna Choromanska, David Sontag
HIGHLIGHT: We provide a novel algorithm to simultaneously perform representation learning for the input data and learning of the hierarchical predictor.
- 173, TITLE: From Patches to Images: A Nonparametric Generative Model
<http://proceedings.mlr.press/v70/ji17a.html>
AUTHORS: Geng Ji, Michael C. Hughes, Erik B. Sudderth
HIGHLIGHT: We propose a hierarchical generative model that captures the self-similar structure of image regions as well as how this structure is shared across image collections.
- 174, TITLE: Density Level Set Estimation on Manifolds with DBSCAN
<http://proceedings.mlr.press/v70/jiang17a.html>
AUTHORS: Heinrich Jiang
HIGHLIGHT: We show that DBSCAN can estimate the connected components of the λ -density level set $\{x : f(x) \geq \lambda\}$ given n i.i.d. samples from an unknown density f .
- 175, TITLE: Uniform Convergence Rates for Kernel Density Estimation
<http://proceedings.mlr.press/v70/jiang17b.html>
AUTHORS: Heinrich Jiang
HIGHLIGHT: We (1) derive finite-sample high-probability density estimation bounds for multivariate KDE under mild density assumptions which hold uniformly in \mathbb{R}^d and bandwidth matrices.
- 176, TITLE: Contextual Decision Processes with low Bellman rank are PAC-Learnable
<http://proceedings.mlr.press/v70/jiang17c.html>
AUTHORS: Nan Jiang, Akshay Krishnamurthy, Alekh Agarwal, John Langford, Robert E. Schapire
HIGHLIGHT: We introduce contextual decision processes (CDPs), that unify most prior RL settings.
- 177, TITLE: Efficient Nonmyopic Active Search
<http://proceedings.mlr.press/v70/jiang17d.html>
AUTHORS: Shali Jiang, Gustavo Malkomes, Geoff Converse, Alyssa Shofner, Benjamin Moseley, Roman Garnett
HIGHLIGHT: We conduct experiments on diverse datasets from several domains: drug discovery, materials science, and a citation network.
- 178, TITLE: How to Escape Saddle Points Efficiently
<http://proceedings.mlr.press/v70/jin17a.html>
AUTHORS: Chi Jin, Rong Ge, Praneeth Netrapalli, Sham M. Kakade, Michael I. Jordan
HIGHLIGHT: This paper shows that a perturbed form of gradient descent converges to a second-order stationary point in a number iterations which depends only poly-logarithmically on dimension (i.e., it is almost “dimension-free”).
- 179, TITLE: Tunable Efficient Unitary Neural Networks (EUNN) and their application to RNNs
<http://proceedings.mlr.press/v70/jing17a.html>
AUTHORS: Li Jing, Yichen Shen, Tena Dubcek, John Peurifoy, Scott Skirlo, Yann LeCun, Max Tegmark, Marin Soljacic

HIGHLIGHT: In this work, we present a new architecture for implementing an Efficient Unitary Neural Network (EUNNs); its main advantages can be summarized as follows.

180, **TITLE:** An Adaptive Test of Independence with Analytic Kernel Embeddings
<http://proceedings.mlr.press/v70/jitkrittum17a.html>
AUTHORS: Wittawat Jitkrittum, Zoltan Szabó, Arthur Gretton
HIGHLIGHT: An Adaptive Test of Independence with Analytic Kernel Embeddings

181, **TITLE:** StingyCD: Safely Avoiding Wasteful Updates in Coordinate Descent
<http://proceedings.mlr.press/v70/johnson17a.html>
AUTHORS: Tyler B. Johnson, Carlos Guestrin
HIGHLIGHT: To address this inefficiency, we propose a modified CD algorithm named “StingyCD.”

182, **TITLE:** Differentially Private Chi-squared Test by Unit Circle Mechanism
<http://proceedings.mlr.press/v70/kakizaki17a.html>
AUTHORS: Kazuya Kakizaki, Kazuto Fukuchi, Jun Sakuma
HIGHLIGHT: Based on the analysis, we present unit circle mechanism: a novel differentially private mechanism based on the geometrical property of the test statistics.

183, **TITLE:** Video Pixel Networks
<http://proceedings.mlr.press/v70/kalchbrenner17a.html>
AUTHORS: Nal Kalchbrenner, Aaron Oord, Karen Simonyan, Ivo Danihelka, Oriol Vinyals, Alex Graves, Koray Kavukcuoglu
HIGHLIGHT: We propose a probabilistic video model, the Video Pixel Network (VPN), that estimates the discrete joint distribution of the raw pixel values in a video.

184, **TITLE:** Adaptive Feature Selection: Computationally Efficient Online Sparse Linear Regression under RIP
<http://proceedings.mlr.press/v70/kale17a.html>
AUTHORS: Satyen Kale, Zohar Karnin, Tengyuan Liang, David P. W.
HIGHLIGHT: In this paper, we make the assumption that data matrix satisfies restricted isometry property, and show that this assumption leads to computationally efficient algorithms with sublinear regret for two variants of the problem.

185, **TITLE:** Recursive Partitioning for Personalization using Observational Data
<http://proceedings.mlr.press/v70/kallus17a.html>
AUTHORS: Nathan Kallus
HIGHLIGHT: By reformulating the problem as a single learning task rather than m separate ones, we propose a new approach based on recursively partitioning the data into regimes where different treatments are optimal.

186, **TITLE:** Multi-fidelity Bayesian Optimisation with Continuous Approximations
<http://proceedings.mlr.press/v70/kandasamy17a.html>
AUTHORS: Kirthevasan Kandasamy, Gautam Dasarathy, Jeff Schneider, Barnabás Póczos
HIGHLIGHT: In this work, we develop a Bayesian optimisation method, BOCA, for this setting.

187, **TITLE:** Schema Networks: Zero-shot Transfer with a Generative Causal Model of Intuitive Physics
<http://proceedings.mlr.press/v70/kansky17a.html>
AUTHORS: Ken Kansky, Tom Silver, David A. McIlroy, Mohamed Eldawy, Miguel Lázaro-Gredilla, Xinghua Lou, Nimrod Dorfman, Szymon Sidor, Scott Phoenix, Dileep George
HIGHLIGHT: In pursuit of efficient and robust generalization, we introduce the Schema Network, an object-oriented generative physics simulator capable of disentangling multiple causes of events and reasoning backward through causes to achieve goals.

188, **TITLE:** Learning in POMDPs with Monte Carlo Tree Search

<http://proceedings.mlr.press/v70/katt17a.html>

AUTHORS: Sammie Katt, Frans A. Oliehoek, Christopher Amato
HIGHLIGHT: In this paper, we extend the Monte-Carlo Tree Search method POMCP to BA-POMDPs and show that the resulting method, which we call BA-POMCP, is able to tackle problems that previous solution methods have been unable to solve.

189, TITLE: Meritocratic Fairness for Cross-Population Selection

<http://proceedings.mlr.press/v70/kearns17a.html>

AUTHORS: Michael Kearns, Aaron Roth, Zhiwei Steven Wu
HIGHLIGHT: We consider the problem of selecting a strong pool of individuals from several populations with incomparable skills (e.g. soccer players, mathematicians, and singers) in a fair manner.

190, TITLE: On Approximation Guarantees for Greedy Low Rank Optimization

<http://proceedings.mlr.press/v70/khanna17a.html>

AUTHORS: Rajiv Khanna, Ethan R. Elenberg, Alexandros G. Dimakis, Joydeep Ghosh, Sahand Negahban
HIGHLIGHT: We provide new approximation guarantees for greedy low rank matrix estimation under standard assumptions of restricted strong convexity and smoothness.

191, TITLE: Graph-based Isometry Invariant Representation Learning

<http://proceedings.mlr.press/v70/khasanova17a.html>

AUTHORS: Renata Khasanova, Pascal Frossard
HIGHLIGHT: In this work we present a novel Transformation Invariant Graph-based Network (TIGraNet), which learns graph-based features that are inherently invariant to isometric transformations such as rotation and translation of input images.

192, TITLE: Learning to Discover Cross-Domain Relations with Generative Adversarial Networks

<http://proceedings.mlr.press/v70/kim17a.html>

AUTHORS: Taeksoo Kim, Moonsu Cha, Hyunsoo Kim, Jung Kwon Lee, Jiwon Kim
HIGHLIGHT: We propose a method based on a generative adversarial network that learns to discover relations between different domains (DiscoGAN).

193, TITLE: SplitNet: Learning to Semantically Split Deep Networks for Parameter Reduction and Model Parallelization

<http://proceedings.mlr.press/v70/kim17b.html>

AUTHORS: Juyong Kim, Yookoon Park, Gunhee Kim, Sung Ju Hwang
HIGHLIGHT: We propose a novel deep neural network that is both lightweight and effectively structured for model parallelization.

194, TITLE: Cost-Optimal Learning of Causal Graphs

<http://proceedings.mlr.press/v70/kocaoglu17a.html>

AUTHORS: Murat Kocaoglu, Alex Dimakis, Sriram Vishwanath
HIGHLIGHT: We consider the problem of learning a causal graph over a set of variables with interventions.

195, TITLE: Understanding Black-box Predictions via Influence Functions

<http://proceedings.mlr.press/v70/koh17a.html>

AUTHORS: Pang Wei Koh, Percy Liang
HIGHLIGHT: In this paper, we use influence functions — a classic technique from robust statistics — to trace a model’s prediction through the learning algorithm and back to its training data, thereby identifying training points most responsible for a given prediction.

196, TITLE: Sub-sampled Cubic Regularization for Non-convex Optimization

<http://proceedings.mlr.press/v70/kohler17a.html>

AUTHORS: Jonas Moritz Kohler, Aurelien Lucchi
HIGHLIGHT: Here, we propose a novel method that uses sub-sampling to lower this computational cost.

- 197, TITLE: PixelCNN Models with Auxiliary Variables for Natural Image Modeling
<http://proceedings.mlr.press/v70/kolesnikov17a.html>
AUTHORS: Alexander Kolesnikov, Christoph H. Lampert
HIGHLIGHT: Subsequently, we describe two new generative image models that exploit different image transformations as auxiliary variables: a quantized grayscale view of the image or a multi-resolution image pyramid.
- 198, TITLE: Active Learning for Cost-Sensitive Classification
<http://proceedings.mlr.press/v70/krishnamurthy17a.html>
AUTHORS: Akshay Krishnamurthy, Alekh Agarwal, Tzu-Kuo Huang, Hal Daum? III, John Langford
HIGHLIGHT: We design an active learning algorithm for cost-sensitive multiclass classification: problems where different errors have different costs.
- 199, TITLE: Evaluating Bayesian Models with Posterior Dispersion Indices
<http://proceedings.mlr.press/v70/kucukelbir17a.html>
AUTHORS: Alp Kucukelbir, Yixin Wang, David M. Blei
HIGHLIGHT: We present a family of posterior dispersion indices (PDI) that capture this idea.
- 200, TITLE: Resource-efficient Machine Learning in 2 KB RAM for the Internet of Things
<http://proceedings.mlr.press/v70/kumar17a.html>
AUTHORS: Ashish Kumar, Saurabh Goyal, Manik Varma
HIGHLIGHT: This paper develops a novel tree-based algorithm, called Bonsai, for efficient prediction on IoT devices – such as those based on the Arduino Uno board having an 8 bit ATmega328P microcontroller operating at 16 MHz with no native floating point support, 2 KB RAM and 32 KB read-only flash.
- 201, TITLE: Grammar Variational Autoencoder
<http://proceedings.mlr.press/v70/kusner17a.html>
AUTHORS: Matt J. Kusner, Brooks Paige, Jos? Miguel Hern?andez-Lobato
HIGHLIGHT: We propose a variational autoencoder which directly encodes from and decodes to these parse trees, ensuring the generated outputs are always syntactically valid.
- 202, TITLE: Co-clustering through Optimal Transport
<http://proceedings.mlr.press/v70/laclau17a.html>
AUTHORS: Charlotte Laclau, Ievgen Redko, Basarab Matei, Youn?s Bennani, Vincent Brault
HIGHLIGHT: In this paper, we present a novel method for co-clustering, an unsupervised learning approach that aims at discovering homogeneous groups of data instances and features by grouping them simultaneously.
- 203, TITLE: Conditional Accelerated Lazy Stochastic Gradient Descent
<http://proceedings.mlr.press/v70/lan17a.html>
AUTHORS: Guanghui Lan, Sebastian Pokutta, Yi Zhou, Daniel Zink
HIGHLIGHT: In this work we introduce a conditional accelerated lazy stochastic gradient descent algorithm with optimal number of calls to a stochastic first-order oracle and convergence rate $O(1/\epsilon^2)$ improving over the projection-free, Online Frank-Wolfe based stochastic gradient descent of (Hazan and Kale, 2012) with convergence rate $O(1/\epsilon^4)$.
- 204, TITLE: Consistent k-Clustering
<http://proceedings.mlr.press/v70/lattanzi17a.html>
AUTHORS: Silvio Lattanzi, Sergei Vassilvitskii
HIGHLIGHT: In this work we formalize this notion and introduce the consistent k-clustering problem.
- 205, TITLE: Deep Spectral Clustering Learning
<http://proceedings.mlr.press/v70/law17a.html>
AUTHORS: Marc T. Law, Raquel Urtasun, Richard S. Zemel

HIGHLIGHT: In this paper, we propose a deep supervised clustering metric learning method that formulates a novel loss function.

206, **TITLE:** Coordinated Multi-Agent Imitation Learning

<http://proceedings.mlr.press/v70/lei17a.html>

AUTHORS: Hoang M. Le, Yisong Yue, Peter Carr, Patrick Lucey

HIGHLIGHT: We propose a joint approach that simultaneously learns a latent coordination model along with the individual policies.

207, **TITLE:** Bayesian inference on random simple graphs with power law degree distributions

<http://proceedings.mlr.press/v70/lee17a.html>

AUTHORS: Juho Lee, Creighton Heaululani, Zoubin Ghahramani, Lancelot F. James, Seungjin Choi

HIGHLIGHT: We present a model for random simple graphs with power law (i.e., heavy-tailed) degree distributions.

208, **TITLE:** Confident Multiple Choice Learning

<http://proceedings.mlr.press/v70/lee17b.html>

AUTHORS: Kimin Lee, Changho Hwang, Kyoungsoo Park, Jinwoo Shin

HIGHLIGHT: In this paper, we propose new ensemble methods specialized for deep neural networks, called confident multiple choice learning (CMCL): it is a variant of multiple choice learning (MCL) via addressing its overconfidence issue. In particular, the proposed major components of CMCL beyond the original MCL scheme are (i) new loss, i.e., confident oracle loss, (ii) new architecture, i.e., feature sharing and (iii) new training method, i.e., stochastic labeling.

209, **TITLE:** Deriving Neural Architectures from Sequence and Graph Kernels

<http://proceedings.mlr.press/v70/lei17a.html>

AUTHORS: Tao Lei, Wengong Jin, Regina Barzilay, Tommi Jaakkola

HIGHLIGHT: In this work, we appeal to kernels over combinatorial structures, such as sequences and graphs, to derive appropriate neural operations.

210, **TITLE:** Doubly Greedy Primal-Dual Coordinate Descent for Sparse Empirical Risk Minimization

<http://proceedings.mlr.press/v70/lei17b.html>

AUTHORS: Qi Lei, Ian En-Hsu Yen, Chao-yuan Wu, Inderjit S. Dhillon, Pradeep Ravikumar

HIGHLIGHT: With a convex-concave saddle point objective reformulation, we propose a Doubly Greedy Primal-Dual Coordinate Descent algorithm that is able to exploit sparsity in both primal and dual variables.

211, **TITLE:** Learning to Align the Source Code to the Compiled Object Code

<http://proceedings.mlr.press/v70/levy17a.html>

AUTHORS: Dor Levy, Lior Wolf

HIGHLIGHT: We propose a new neural network architecture and use it for the task of statement-by-statement alignment of source code and its compiled object code.

212, **TITLE:** Dropout Inference in Bayesian Neural Networks with Alpha-divergences

<http://proceedings.mlr.press/v70/li17a.html>

AUTHORS: Yingzhen Li, Yarin Gal

HIGHLIGHT: We propose a re-parametrisation of the alpha-divergence objectives, deriving a simple inference technique which, together with dropout, can be easily implemented with existing models by simply changing the loss of the model.

213, **TITLE:** Provable Alternating Gradient Descent for Non-negative Matrix Factorization with Strong Correlations

<http://proceedings.mlr.press/v70/li17b.html>

AUTHORS: Yuanzhi Li, Yingyu Liang

HIGHLIGHT: This paper proposes a simple and natural alternating gradient descent based algorithm, and shows that with a mild initialization it provably recovers the ground-truth in the presence of strong correlations.

214, TITLE: Provably Optimal Algorithms for Generalized Linear Contextual Bandits
<http://proceedings.mlr.press/v70/li17c.html>

AUTHORS: Lihong Li, Yu Lu, Dengyong Zhou

HIGHLIGHT: In this work, we propose an upper confidence bound based algorithm for generalized linear contextual bandits, which achieves an $\tilde{O}(\sqrt{dT})$ regret over T rounds with d dimensional feature vectors.

215, TITLE: Fast k -Nearest Neighbour Search via Prioritized DCI

<http://proceedings.mlr.press/v70/li17d.html>

AUTHORS: Ke Li, Jitendra Malik

HIGHLIGHT: In this paper, we propose a variant of DCI, which we call Prioritized DCI, and show a remarkable improvement in the dependence of query time on intrinsic dimensionality.

216, TITLE: Forest-type Regression with General Losses and Robust Forest

<http://proceedings.mlr.press/v70/li17e.html>

AUTHORS: Alexander Hanbo Li, Andrew Martin

HIGHLIGHT: This paper introduces a new general framework for forest-type regression which allows the development of robust forest regressors by selecting from a large family of robust loss functions.

217, TITLE: Stochastic Modified Equations and Adaptive Stochastic Gradient Algorithms

<http://proceedings.mlr.press/v70/li17f.html>

AUTHORS: Qianxiao Li, Cheng Tai, Weinan E

HIGHLIGHT: We develop the method of stochastic modified equations (SME), in which stochastic gradient algorithms are approximated in the weak sense by continuous-time stochastic differential equations.

218, TITLE: Convergence Analysis of Proximal Gradient with Momentum for Nonconvex Optimization

<http://proceedings.mlr.press/v70/li17g.html>

AUTHORS: Qunwei Li, Yi Zhou, Yingbin Liang, Pramod K. Varshney

HIGHLIGHT: In this work, we investigate the accelerated proximal gradient method for nonconvex programming (APGnc).

219, TITLE: Exact MAP Inference by Avoiding Fractional Vertices

<http://proceedings.mlr.press/v70/lindgren17a.html>

AUTHORS: Erik M. Lindgren, Alexandros G. Dimakis, Adam Klivans

HIGHLIGHT: We give a natural condition under which we can provably perform MAP inference in polynomial time—we require that the number of fractional vertices in the LP relaxation exceeding the optimal solution is bounded by a polynomial in the problem size.

220, TITLE: Leveraging Union of Subspace Structure to Improve Constrained Clustering

<http://proceedings.mlr.press/v70/lipor17a.html>

AUTHORS: John Lipor, Laura Balzano

HIGHLIGHT: We present a pairwise-constrained clustering algorithm that actively selects queries based on the union-of-subspaces model.

221, TITLE: Zero-Inflated Exponential Family Embeddings

<http://proceedings.mlr.press/v70/liu17a.html>

AUTHORS: Li-Ping Liu, David M. Blei

HIGHLIGHT: In this paper, we develop zero-inflated embeddings, a new embedding method that is designed to learn from sparse observations.

222, TITLE: Iterative Machine Teaching

<http://proceedings.mlr.press/v70/liu17b.html>

AUTHORS: Weiyang Liu, Bo Dai, Ahmad Humayun, Charlene Tay, Chen Yu, Linda B. Smith, James M. Rehg, Le Song
HIGHLIGHT: In this paper, we consider the problem of machine teaching, the inverse problem of machine learning.

223, TITLE: Algorithmic Stability and Hypothesis Complexity

<http://proceedings.mlr.press/v70/liu17c.html>

AUTHORS: Tongliang Liu, Gabor Lugosi, Gergely Neu, Dacheng Tao
HIGHLIGHT: We introduce a notion of algorithmic stability of learning algorithms—that we term hypothesis stability—that captures stability of the hypothesis output by the learning algorithm in the normed space of functions from which hypotheses are selected.

224, TITLE: Analogical Inference for Multi-relational Embeddings

<http://proceedings.mlr.press/v70/liu17d.html>

AUTHORS: Hanxiao Liu, Yuexin Wu, Yiming Yang
HIGHLIGHT: This paper proposes a novel framework for optimizing the latent representations with respect to the analogical properties of the embedded entities and relations.

225, TITLE: Dual Iterative Hard Thresholding: From Non-convex Sparse Minimization to Non-smooth Concave Maximization

<http://proceedings.mlr.press/v70/liu17e.html>

AUTHORS: Bo Liu, Xiao-Tong Yuan, Lezi Wang, Qingshan Liu, Dimitris N. Metaxas
HIGHLIGHT: In this article, we bridge the gap by establishing a duality theory for sparsity-constrained minimization with ℓ_2 -regularized objective and proposing an IHT-style algorithm for dual maximization.

226, TITLE: Gram-CTC: Automatic Unit Selection and Target Decomposition for Sequence Labelling

<http://proceedings.mlr.press/v70/liu17f.html>

AUTHORS: Hairong Liu, Zhenyao Zhu, Xiangang Li, Sanjeev Sathesh
HIGHLIGHT: In this paper, we extend the popular CTC loss criterion to alleviate these limitations, and propose a new loss function called Gram-CTC.

227, TITLE: Learning Infinite Layer Networks Without the Kernel Trick

<http://proceedings.mlr.press/v70/livni17a.html>

AUTHORS: Roi Livni, Daniel Carmon, Amir Globerson
HIGHLIGHT: In this work we give an online algorithm for ILN, which avoids the kernel trick assumption.

228, TITLE: Deep Transfer Learning with Joint Adaptation Networks

<http://proceedings.mlr.press/v70/long17a.html>

AUTHORS: Mingsheng Long, Han Zhu, Jianmin Wang, Michael I. Jordan
HIGHLIGHT: In this paper, we present joint adaptation networks (JAN), which learn a transfer network by aligning the joint distributions of multiple domain-specific layers across domains based on a joint maximum mean discrepancy (JMMD) criterion.

229, TITLE: Multiplicative Normalizing Flows for Variational Bayesian Neural Networks

<http://proceedings.mlr.press/v70/louizos17a.html>

AUTHORS: Christos Louizos, Max Welling
HIGHLIGHT: We show that through this interpretation it is both efficient and straightforward to improve the approximation by employing normalizing flows while still allowing for local reparametrizations and a tractable lower bound.

230, TITLE: How Close Are the Eigenvectors of the Sample and Actual Covariance Matrices?

<http://proceedings.mlr.press/v70/loukas17a.html>

AUTHORS: Andreas Loukas

HIGHLIGHT: For a wide family of distributions, including distributions with finite second moment and sub-gaussian distributions supported in a centered Euclidean ball, we prove that the inner product between eigenvectors of the sample and actual covariance matrices decreases proportionally to the respective eigenvalue distance and the number of samples.

231, **TITLE:** Learning Deep Architectures via Generalized Whitened Neural Networks

<http://proceedings.mlr.press/v70/luo17a.html>

AUTHORS: Ping Luo

HIGHLIGHT: Unlike WNN that reduced runtime by performing whitening every thousand iterations, which degenerates convergence due to the ill conditioning, we present generalized WNN (GWNN), which has three appealing properties.

232, **TITLE:** Learning Gradient Descent: Better Generalization and Longer Horizons

<http://proceedings.mlr.press/v70/lv17a.html>

AUTHORS: Kaifeng Lv, Shunhua Jiang, Jian Li

HIGHLIGHT: In this paper, we propose a new learning-to-learn model and some useful and practical tricks.

233, **TITLE:** Spherical Structured Feature Maps for Kernel Approximation

<http://proceedings.mlr.press/v70/lyu17a.html>

AUTHORS: Yueming Lyu

HIGHLIGHT: Thus, we propose an efficient coordinate decent method to find a local optimum of the discrete Riesz s -energy for SSF maps construction.

234, **TITLE:** Stochastic Gradient MCMC Methods for Hidden Markov Models

<http://proceedings.mlr.press/v70/ma17a.html>

AUTHORS: Yi-An Ma, Nicholas J. Foti, Emily B. Fox

HIGHLIGHT: We consider a marginal likelihood representation of the HMM and propose an algorithm that harnesses the inherent memory decay of the process.

235, **TITLE:** Self-Paced Co-training

<http://proceedings.mlr.press/v70/ma17b.html>

AUTHORS: Fan Ma, Deyu Meng, Qi Xie, Zina Li, Xuanyi Dong

HIGHLIGHT: To these issues, in this study we design a new co-training algorithm named self-paced cotraining (SPaCo) with a “draw with replacement” learning mode.

236, **TITLE:** Interactive Learning from Policy-Dependent Human Feedback

<http://proceedings.mlr.press/v70/macglashan17a.html>

AUTHORS: James MacGlashan, Mark K. Ho, Robert Loftin, Bei Peng, Guan Wang, David L. Roberts, Matthew E. Taylor, Michael L. Littman

HIGHLIGHT: Based on this insight, we introduce Convergent Actor-Critic by Humans (COACH), an algorithm for learning from policy-dependent feedback that converges to a local optimum.

237, **TITLE:** A Laplacian Framework for Option Discovery in Reinforcement Learning

<http://proceedings.mlr.press/v70/machado17a.html>

AUTHORS: Marlos C. Machado, Marc G. Bellemare, Michael Bowling

HIGHLIGHT: In this paper we address the option discovery problem by showing how PVFs implicitly define options.

238, **TITLE:** Frame-based Data Factorizations

<http://proceedings.mlr.press/v70/mair17a.html>

AUTHORS: Sebastian Mair, Aho?ne Boubekki, Ulf Brefeld

HIGHLIGHT: In this paper, we show that the set of vertices of a convex hull, the so-called frame, can be efficiently computed by a quadratic program.

- 239, TITLE: Global optimization of Lipschitz functions
<http://proceedings.mlr.press/v70/malherbe17a.html>
AUTHORS: Cédric Malherbe, Nicolas Vayatis
HIGHLIGHT: The goal of the paper is to design sequential strategies which lead to efficient optimization of an unknown function under the only assumption that it has a finite Lipschitz constant.
- 240, TITLE: On Mixed Memberships and Symmetric Nonnegative Matrix Factorizations
<http://proceedings.mlr.press/v70/mao17a.html>
AUTHORS: Xueyu Mao, Purnamrita Sarkar, Deepayan Chakrabarti
HIGHLIGHT: We link the two approaches, by (a) establishing sufficient conditions for the symmetric NMF optimization to have a unique solution under MMSB, and (b) proposing a computationally efficient algorithm called GeoNMF that is provably optimal and hence consistent for a broad parameter regime.
- 241, TITLE: Bayesian Models of Data Streams with Hierarchical Power Priors
<http://proceedings.mlr.press/v70/masegosa17a.html>
AUTHORS: Andrés Masegosa, Thomas D. Nielsen, Helge Langseth, Darío Ramos-López, Antonio Salmerón, Anders L. Madsen
HIGHLIGHT: In this paper, we approach these problems from a Bayesian perspective covering general conjugate exponential models.
- 242, TITLE: Just Sort It! A Simple and Effective Approach to Active Preference Learning
<http://proceedings.mlr.press/v70/maystre17a.html>
AUTHORS: Lucas Maystre, Matthias Grossglauser
HIGHLIGHT: Our goal is to recover the ranking accurately but to sample the comparisons sparingly.
- 243, TITLE: ChoiceRank: Identifying Preferences from Node Traffic in Networks
<http://proceedings.mlr.press/v70/maystre17b.html>
AUTHORS: Lucas Maystre, Matthias Grossglauser
HIGHLIGHT: We show how to make the inference problem well-posed regardless of the network's structure, and we present ChoiceRank, an iterative algorithm that scales to networks that contains billions of nodes and edges.
- 244, TITLE: Deciding How to Decide: Dynamic Routing in Artificial Neural Networks
<http://proceedings.mlr.press/v70/mcgill17a.html>
AUTHORS: Mason McGill, Pietro Perona
HIGHLIGHT: We propose and systematically evaluate three strategies for training dynamically-routed artificial neural networks: graphs of learned transformations through which different input signals may take different paths.
- 245, TITLE: Risk Bounds for Transferring Representations With and Without Fine-Tuning
<http://proceedings.mlr.press/v70/mcnamara17a.html>
AUTHORS: Daniel McNamara, Maria-Florina Balcan
HIGHLIGHT: We develop sufficient conditions for the success of this approach.
- 246, TITLE: Nonnegative Matrix Factorization for Time Series Recovery From a Few Temporal Aggregates
<http://proceedings.mlr.press/v70/mei17a.html>
AUTHORS: Jiali Mei, Yohann De Castro, Yannig Goude, Georges H. Braïl
HIGHLIGHT: Motivated by electricity consumption reconstitution, we propose a new matrix recovery method using nonnegative matrix factorization (NMF).
- 247, TITLE: Adversarial Variational Bayes: Unifying Variational Autoencoders and Generative Adversarial Networks
<http://proceedings.mlr.press/v70/mescheder17a.html>
AUTHORS: Lars Mescheder, Sebastian Nowozin, Andreas Geiger

HIGHLIGHT: We introduce Adversarial Variational Bayes (AVB), a technique for training Variational Autoencoders with arbitrarily expressive inference models.

248, **TITLE:** Efficient Orthogonal Parametrisation of Recurrent Neural Networks Using Householder Reflections

<http://proceedings.mlr.press/v70/mhammedi17a.html>

AUTHORS: Zakaria Mhammedi, Andrew Hellicar, Ashfaqur Rahman, James Bailey

HIGHLIGHT: Recent methods have been suggested to solve this problem by constraining the transition matrix to be unitary during training which ensures that its norm is equal to one and prevents exploding gradients.

249, **TITLE:** Discovering Discrete Latent Topics with Neural Variational Inference

<http://proceedings.mlr.press/v70/miao17a.html>

AUTHORS: Yishu Miao, Edward Grefenstette, Phil Blunsom

HIGHLIGHT: This paper presents alternative neural approaches to topic modelling by providing parameterisable distributions over topics which permit training by backpropagation in the framework of neural variational inference.

250, **TITLE:** Variational Boosting: Iteratively Refining Posterior Approximations

<http://proceedings.mlr.press/v70/miller17a.html>

AUTHORS: Andrew C. Miller, Nicholas J. Foti, Ryan P. Adams

HIGHLIGHT: We propose a black-box variational inference method to approximate intractable distributions with an increasingly rich approximating class.

251, **TITLE:** Device Placement Optimization with Reinforcement Learning

<http://proceedings.mlr.press/v70/mirhoseini17a.html>

AUTHORS: Azalia Mirhoseini, Hieu Pham, Quoc V. Le, Benoit Steiner, Rasmus Larsen, Yuefeng Zhou, Naveen Kumar, Mohammad Norouzi, Samy Bengio, Jeff Dean

HIGHLIGHT: In this paper, we propose a method which learns to optimize device placement for TensorFlow computational graphs.

252, **TITLE:** Tight Bounds for Approximate Carathéodory and Beyond

<http://proceedings.mlr.press/v70/mirrokn17a.html>

AUTHORS: Vahab Mirrokni, Renato Paes Leme, Adrian Vladu, Sam Chiu-wai Wong

HIGHLIGHT: We present a deterministic nearly-linear time algorithm for approximating any point inside a convex polytope with a sparse convex combination of the polytope's vertices.

253, **TITLE:** Deletion-Robust Submodular Maximization: Data Summarization with the Right to be Forgotten?

<http://proceedings.mlr.press/v70/mirzsoleiman17a.html>

AUTHORS: Baharan Mirzsoleiman, Amin Karbasi, Andreas Krause

HIGHLIGHT: Motivated by this challenge, we introduce the dynamic deletion-robust submodular maximization problem.

254, **TITLE:** Prediction and Control with Temporal Segment Models

<http://proceedings.mlr.press/v70/mishra17a.html>

AUTHORS: Nikhil Mishra, Pieter Abbeel, Igor Mordatch

HIGHLIGHT: We introduce a method for learning the dynamics of complex nonlinear systems based on deep generative models over temporal segments of states and actions.

255, **TITLE:** Improving Gibbs Sampler Scan Quality with DoGS

<http://proceedings.mlr.press/v70/mitliagkas17a.html>

AUTHORS: Ioannis Mitliagkas, Lester Mackey

HIGHLIGHT: In this work, we use Dobrushin influence as the basis of a practical tool to certify and efficiently improve the quality of a Gibbs sampler.

256, TITLE: Differentially Private Submodular Maximization: Data Summarization in Disguise
<http://proceedings.mlr.press/v70/mitrovic17a.html>
AUTHORS: Marko Mitrovic, Mark Bun, Andreas Krause, Amin Karbasi
HIGHLIGHT: We present privacy-preserving algorithms for both monotone and non-monotone submodular maximization under cardinality, matroid, and p-extendible system constraints, with guarantees that are competitive with optimal.

257, TITLE: Active Learning for Top- k Rank Aggregation from Noisy Comparisons
<http://proceedings.mlr.press/v70/mohajer17a.html>
AUTHORS: Soheil Mohajer, Changho Suh, Adel Elmahdy
HIGHLIGHT: Under a fairly general model which subsumes as special cases various models (e.g., Strong Stochastic Transitivity model, BTL model and uniform noise model), we characterize upper bounds on the sample size required for top- k sorting as well as for top- k partitioning.

258, TITLE: Variational Dropout Sparsifies Deep Neural Networks
<http://proceedings.mlr.press/v70/molchanov17a.html>
AUTHORS: Dmitry Molchanov, Arsenii Ashukha, Dmitry Vetrov
HIGHLIGHT: We extend Variational Dropout to the case when dropout rates are unbounded, propose a way to reduce the variance of the gradient estimator and report first experimental results with individual dropout rates per weight.

259, TITLE: Regularising Non-linear Models Using Feature Side-information
<http://proceedings.mlr.press/v70/mollaysa17a.html>
AUTHORS: Amina Mollaysa, Pablo Strasser, Alexandros Kalousis
HIGHLIGHT: In this paper, we propose a framework that allows for the incorporation of the feature side-information during the learning of very general model families to improve the prediction performance.

260, TITLE: Coupling Distributed and Symbolic Execution for Natural Language Queries
<http://proceedings.mlr.press/v70/mou17a.html>
AUTHORS: Lili Mou, Zhengdong Lu, Hang Li, Zhi Jin
HIGHLIGHT: In this paper, we propose to couple distributed and symbolic execution for natural language queries, where the symbolic executor is pretrained with the distributed executor's intermediate execution results in a step-by-step fashion.

261, TITLE: McGan: Mean and Covariance Feature Matching GAN
<http://proceedings.mlr.press/v70/mroueh17a.html>
AUTHORS: Youssef Mroueh, Tom Sercu, Vaibhava Goel
HIGHLIGHT: We introduce new families of Integral Probability Metrics (IPM) for training Generative Adversarial Networks (GAN).

262, TITLE: Sequence to Better Sequence: Continuous Revision of Combinatorial Structures
<http://proceedings.mlr.press/v70/mueller17a.html>
AUTHORS: Jonas Mueller, David Gifford, Tommi Jaakkola
HIGHLIGHT: We present a model that, after learning on observations of (sequence, outcome) pairs, can be efficiently used to revise a new sequence in order to improve its associated outcome.

263, TITLE: Variants of RMSProp and Adagrad with Logarithmic Regret Bounds
<http://proceedings.mlr.press/v70/mukkamala17a.html>
AUTHORS: Mahesh Chandra Mukkamala, Matthias Hein
HIGHLIGHT: In this paper we have analyzed RMSProp, originally proposed for the training of deep neural networks, in the context of online convex optimization and show \sqrt{T} -type regret bounds.

264, TITLE: Meta Networks
<http://proceedings.mlr.press/v70/munkhdalai17a.html>

AUTHORS: Tsendsuren Munkhdalai, Hong Yu
HIGHLIGHT: In this work, we introduce a novel meta learning method, Meta Networks (MetaNet), that learns a meta-level knowledge across tasks and shifts its inductive biases via fast parameterization for rapid generalization.

265, TITLE: Understanding the Representation and Computation of Multilayer Perceptrons: A Case Study in Speech Recognition

<http://proceedings.mlr.press/v70/nagamine17a.html>

AUTHORS: Tasha Nagamine, Nima Mesgarani

HIGHLIGHT: This study provides an empirical framework to study the encoding properties of node activations in various layers of the network, and to construct the exact function applied to each data point in the form of a linear transform.

266, TITLE: Adaptive Sampling Probabilities for Non-Smooth Optimization

<http://proceedings.mlr.press/v70/namkoong17a.html>

AUTHORS: Hongseok Namkoong, Aman Sinha, Steve Yadlowsky, John C. Duchi

HIGHLIGHT: We present a framework that discovers and leverages such structural properties at a low computational cost.

267, TITLE: Delta Networks for Optimized Recurrent Network Computation

<http://proceedings.mlr.press/v70/neil17a.html>

AUTHORS: Daniel Neil, Jun Haeng Lee, Tobi Delbruck, Shih-Chii Liu

HIGHLIGHT: By capitalizing on this property of natural signals, we propose a Recurrent Neural Network (RNN) architecture called a delta network in which each neuron transmits its value only when the change in its activation exceeds a threshold.

268, TITLE: Post-Inference Prior Swapping

<http://proceedings.mlr.press/v70/neiswanger17a.html>

AUTHORS: Willie Neiswanger, Eric Xing

HIGHLIGHT: In this paper, we investigate the following question: for a given model, is it possible to compute an inference result with any convenient false prior, and afterwards, given any target prior of interest, quickly transform this result into the target posterior?

269, TITLE: The Loss Surface of Deep and Wide Neural Networks

<http://proceedings.mlr.press/v70/nguyen17a.html>

AUTHORS: Quynh Nguyen, Matthias Hein

HIGHLIGHT: We show that this is (almost) true, in fact almost all local minima are globally optimal, for a fully connected network with squared loss and analytic activation function given that the number of hidden units of one layer of the network is larger than the number of training points and the network structure from this layer on is pyramidal.

270, TITLE: SARAH: A Novel Method for Machine Learning Problems Using Stochastic Recursive Gradient

<http://proceedings.mlr.press/v70/nguyen17b.html>

AUTHORS: Lam M. Nguyen, Jie Liu, Katya Scheinberg, Martin Takáč

HIGHLIGHT: In this paper, we propose a Stochastic Recursive Gradient algorithm (SARAH), as well as its practical variant SARAH+, as a novel approach to the finite-sum minimization problems.

271, TITLE: Composing Tree Graphical Models with Persistent Homology Features for Clustering Mixed-Type Data

<http://proceedings.mlr.press/v70/ni17a.html>

AUTHORS: Xiuyan Ni, Novi Quadrianto, Yusu Wang, Chao Chen

HIGHLIGHT: In this paper, we propose a clustering method based on a tree-structured graphical model to describe the generation process of mixed-type data.

272, TITLE: Multichannel End-to-end Speech Recognition

<http://proceedings.mlr.press/v70/ochiai17a.html>

- AUTHORS: Tsubasa Ochiai, Shinji Watanabe, Takaaki Hori, John R. Hershey
HIGHLIGHT: In this paper we extend the end-to-end framework to encompass microphone array signal processing for noise suppression and speech enhancement within the acoustic encoding network.
- 273, TITLE: Conditional Image Synthesis with Auxiliary Classifier GANs
<http://proceedings.mlr.press/v70/odena17a.html>
AUTHORS: Augustus Odena, Christopher Olah, Jonathon Shlens
HIGHLIGHT: In this paper we introduce new methods for the improved training of generative adversarial networks (GANs) for image synthesis.
- 274, TITLE: Nyström Method with Kernel K-means++ Samples as Landmarks
<http://proceedings.mlr.press/v70/oglic17a.html>
AUTHORS: Dino Oglic, Thomas Geertner
HIGHLIGHT: We investigate, theoretically and empirically, the effectiveness of kernel K-means++ samples as landmarks in the Nyström method for low-rank approximation of kernel matrices.
- 275, TITLE: Zero-Shot Task Generalization with Multi-Task Deep Reinforcement Learning
<http://proceedings.mlr.press/v70/oh17a.html>
AUTHORS: Junhyuk Oh, Satinder Singh, Honglak Lee, Pushmeet Kohli
HIGHLIGHT: For generalization over unseen instructions, we propose a new objective which encourages learning correspondences between similar subtasks by making analogies.
As a step towards developing zero-shot task generalization capabilities in reinforcement learning (RL), we introduce a new RL problem where the agent should learn to execute sequences of instructions after learning useful skills that solve subtasks.
- 276, TITLE: The Statistical Recurrent Unit
<http://proceedings.mlr.press/v70/oliva17a.html>
AUTHORS: Junier B. Oliva, Barnabás Póczos, Jeff Schneider
HIGHLIGHT: We show the efficacy of SRUs as compared to LSTMs and GRUs in an unbiased manner by optimizing respective architectures' hyperparameters for both synthetic and real-world tasks.
- 277, TITLE: Deep Decentralized Multi-task Multi-Agent Reinforcement Learning under Partial Observability
<http://proceedings.mlr.press/v70/omidshafiei17a.html>
AUTHORS: Shayegan Omidshafiei, Jason Pazis, Christopher Amato, Jonathan P. How, John Vian
HIGHLIGHT: We introduce a decentralized single-task learning approach that is robust to concurrent interactions of teammates, and present an approach for distilling single-task policies into a unified policy that performs well across multiple related tasks, without explicit provision of task identity.
- 278, TITLE: Algebraic Variety Models for High-Rank Matrix Completion
<http://proceedings.mlr.press/v70/ongie17a.html>
AUTHORS: Greg Ongie, Rebecca Willett, Robert D. Nowak, Laura Balzano
HIGHLIGHT: We propose an efficient matrix completion algorithm that minimizes a convex or non-convex surrogate of the rank of the lifted matrix.
- 279, TITLE: Why is Posterior Sampling Better than Optimism for Reinforcement Learning?
<http://proceedings.mlr.press/v70/osband17a.html>
AUTHORS: Ian Osband, Benjamin Van Roy
HIGHLIGHT: We provide insight into the extent of this performance boost and the phenomenon that drives it.
- 280, TITLE: Bidirectional Learning for Time-series Models with Hidden Units
<http://proceedings.mlr.press/v70/osogami17a.html>
AUTHORS: Takayuki Osogami, Hiroshi Kajino, Taro Sekiyama
HIGHLIGHT: Here we propose a way to learn such a time-series model by training a backward model for the time-reversed time-series, where the backward model has a common set of parameters as the original (forward) model.

- 281, TITLE: Count-Based Exploration with Neural Density Models
<http://proceedings.mlr.press/v70/ostrovski17a.html>
AUTHORS: Georg Ostrovski, Marc G. Bellemare, Aaron Oord, Rami Munos
HIGHLIGHT: We answer the first question by demonstrating the use of PixelCNN, an advanced neural density model for images, to supply a pseudo-count.
- 282, TITLE: Dictionary Learning Based on Sparse Distribution Tomography
<http://proceedings.mlr.press/v70/pad17a.html>
AUTHORS: Pedram Pad, Farnood Salehi, Elisa Celis, Patrick Thiran, Michael Unser
HIGHLIGHT: We propose a new statistical dictionary learning algorithm for sparse signals that is based on an α -stable innovation model.
- 283, TITLE: Stochastic Bouncy Particle Sampler
<http://proceedings.mlr.press/v70/pakman17a.html>
AUTHORS: Ari Pakman, Dar Gilboa, David Carlson, Liam Paninski
HIGHLIGHT: We introduce a simple method that controls this trade-off.
- 284, TITLE: A Birth-Death Process for Feature Allocation
<http://proceedings.mlr.press/v70/palla17a.html>
AUTHORS: Konstantina Palla, David Knowles, Zoubin Ghahramani
HIGHLIGHT: We propose a Bayesian nonparametric prior over feature allocations for sequential data, the birth-death feature allocation process (BDFP).
- 285, TITLE: Prediction under Uncertainty in Sparse Spectrum Gaussian Processes with Applications to Filtering and Control
<http://proceedings.mlr.press/v70/pan17a.html>
AUTHORS: Yunpeng Pan, Xinyan Yan, Evangelos A. Theodorou, Byron Boots
HIGHLIGHT: We address this problem by proposing two analytic moment-based approaches with closed-form expressions for SSGP regression with uncertain inputs.
- 286, TITLE: Clustering by Sum of Norms: Stochastic Incremental Algorithm, Convergence and Cluster Recovery
<http://proceedings.mlr.press/v70/panahi17a.html>
AUTHORS: Ashkan Panahi, Devdatt Dubhashi, Fredrik D. Johansson, Chiranjib Bhattacharyya
HIGHLIGHT: We give a scalable stochastic incremental algorithm based on proximal iterations to solve the SON problem with convergence guarantees.
- 287, TITLE: Curiosity-driven Exploration by Self-supervised Prediction
<http://proceedings.mlr.press/v70/pathak17a.html>
AUTHORS: Deepak Pathak, Pulkit Agrawal, Alexei A. Efros, Trevor Darrell
HIGHLIGHT: We formulate curiosity as the error in an agent's ability to predict the consequence of its own actions in a visual feature space learned by a self-supervised inverse dynamics model.
- 288, TITLE: Asynchronous Distributed Variational Gaussian Process for Regression
<http://proceedings.mlr.press/v70/peng17a.html>
AUTHORS: Hao Peng, Shandian Zhe, Xiao Zhang, Yuan Qi
HIGHLIGHT: To solve this problem, we propose ADVGP, the first Asynchronous Distributed Variational Gaussian Process inference for regression, on the recent large-scale machine learning platform, PARAMETER SERVER.
- 289, TITLE: Geometry of Neural Network Loss Surfaces via Random Matrix Theory
<http://proceedings.mlr.press/v70/pennington17a.html>

- AUTHORS: Jeffrey Pennington, Yasaman Bahri
HIGHLIGHT: In this paper, we study the geometry in terms of the distribution of eigenvalues of the Hessian matrix at critical points of varying energy.
We introduce an analytical framework and a set of tools from random matrix theory that allow us to compute an approximation of this distribution under a set of simplifying assumptions.
- 290, TITLE: Multi-task Learning with Labeled and Unlabeled Tasks
<http://proceedings.mlr.press/v70/pentina17a.html>
AUTHORS: Anastasia Pentina, Christoph H. Lampert
HIGHLIGHT: Focusing on an instance-based transfer method we analyze two variants of this setting: when the set of labeled tasks is fixed, and when it can be actively selected by the learner.
- 291, TITLE: Robust Adversarial Reinforcement Learning
<http://proceedings.mlr.press/v70/pinto17a.html>
AUTHORS: Lerrel Pinto, James Davidson, Rahul Sukthankar, Abhinav Gupta
HIGHLIGHT: This paper proposes the idea of robust adversarial reinforcement learning (RARL), where we train an agent to operate in the presence of a destabilizing adversary that applies disturbance forces to the system.
- 292, TITLE: Neural Episodic Control
<http://proceedings.mlr.press/v70/pritzel17a.html>
AUTHORS: Alexander Pritzel, Benigno Uria, Sriram Srinivasan, Adri? Puigdom?nech Badia, Oriol Vinyals, Demis Hassabis, Daan Wierstra, Charles Blundell
HIGHLIGHT: We propose Neural Episodic Control: a deep reinforcement learning agent that is able to rapidly assimilate new experiences and act upon them.
- 293, TITLE: Online and Linear-Time Attention by Enforcing Monotonic Alignments
<http://proceedings.mlr.press/v70/raffel17a.html>
AUTHORS: Colin Raffel, Minh-Thang Luong, Peter J. Liu, Ron J. Weiss, Douglas Eck
HIGHLIGHT: Based on the insight that the alignment between input and output sequence elements is monotonic in many problems of interest, we propose an end-to-end differentiable method for learning monotonic alignments which, at test time, enables computing attention online and in linear time.
- 294, TITLE: On the Expressive Power of Deep Neural Networks
<http://proceedings.mlr.press/v70/raghu17a.html>
AUTHORS: Maithra Raghu, Ben Poole, Jon Kleinberg, Surya Ganguli, Jascha Sohl-Dickstein
HIGHLIGHT: We propose a new approach to the problem of neural network expressivity, which seeks to characterize how structural properties of a neural network family affect the functions it is able to compute.
- 295, TITLE: Estimating the unseen from multiple populations
<http://proceedings.mlr.press/v70/raghunathan17a.html>
AUTHORS: Aditi Raghunathan, Gregory Valiant, James Zou
HIGHLIGHT: We generalize this extrapolation and related unseen estimation problems to the multiple population setting, where population j has an unknown distribution D_j from which we observe n_j samples.
- 296, TITLE: Coherence Pursuit: Fast, Simple, and Robust Subspace Recovery
<http://proceedings.mlr.press/v70/rahmani17a.html>
AUTHORS: Mostafa Rahmani, George Atia
HIGHLIGHT: This paper presents a remarkably simple, yet powerful, algorithm for robust Principal Component Analysis (PCA).
- 297, TITLE: Innovation Pursuit: A New Approach to the Subspace Clustering Problem
<http://proceedings.mlr.press/v70/rahmani17b.html>
AUTHORS: Mostafa Rahmani, George Atia

HIGHLIGHT: This paper presents a new scalable approach, termed Innovation Pursuit (iPursuit), to the problem of subspace clustering.

298, **TITLE:** High Dimensional Bayesian Optimization with Elastic Gaussian Process

<http://proceedings.mlr.press/v70/rana17a.html>

AUTHORS: Santu Rana, Cheng Li, Sunil Gupta, Vu Nguyen, Svetha Venkatesh

HIGHLIGHT: We propose an algorithm that enables local gradient-dependent algorithms to move through the flat terrain by using a sequence of gross-to-finer Gaussian process priors on the objective function as we leverage two underlying facts - a) there exists a large enough length-scales for which the acquisition function can be made to have a significant gradient at any location in the parameter space, and b) the extrema of the consecutive acquisition functions are close although they are different only due to a small difference in the length-scales.

299, **TITLE:** Equivariance Through Parameter-Sharing

<http://proceedings.mlr.press/v70/ravanbakhsh17a.html>

AUTHORS: Siamak Ravanbakhsh, Jeff Schneider, Barnab's P'ozos

HIGHLIGHT: We propose to study equivariance in deep neural networks through parameter symmetries.

300, **TITLE:** Large-Scale Evolution of Image Classifiers

<http://proceedings.mlr.press/v70/real17a.html>

AUTHORS: Esteban Real, Sherry Moore, Andrew Selle, Saurabh Saxena, Yutaka Leon Suematsu, Jie Tan, Quoc V. Le, Alexey Kurakin

HIGHLIGHT: Our goal is to minimize human participation, so we employ evolutionary algorithms to discover such networks automatically.

301, **TITLE:** Parallel Multiscale Autoregressive Density Estimation

<http://proceedings.mlr.press/v70/reed17a.html>

AUTHORS: Scott Reed, A?ron Oord, Nal Kalchbrenner, Sergio G?mez Colmenarejo, Ziyu Wang, Yutian Chen, Dan Belov, Nando Freitas

HIGHLIGHT: In this work, we propose a parallelized PixelCNN that allows more efficient inference by modeling certain pixel groups as conditionally independent.

302, **TITLE:** Real-Time Adaptive Image Compression

<http://proceedings.mlr.press/v70/rippel17a.html>

AUTHORS: Oren Rippel, Lubomir Bourdev

HIGHLIGHT: We present a machine learning-based approach to lossy image compression which outperforms all existing codecs, while running in real-time.

303, **TITLE:** Active Learning for Accurate Estimation of Linear Models

<http://proceedings.mlr.press/v70/riquelme17a.html>

AUTHORS: Carlos Riquelme, Mohammad Ghavamzadeh, Alessandro Lazaric

HIGHLIGHT: We present Trace-UCB, an adaptive allocation algorithm that learns the noise levels while balancing contexts accordingly across the different linear functions, and derive guarantees for simple regret in both expectation and high-probability.

304, **TITLE:** Cognitive Psychology for Deep Neural Networks: A Shape Bias Case Study

<http://proceedings.mlr.press/v70/ritter17a.html>

AUTHORS: Samuel Ritter, David G. T. Barrett, Adam Santoro, Matt M. Botvinick

HIGHLIGHT: To explore the potential value of these tools, we chose a well-established analysis from developmental psychology that explains how children learn word labels for objects, and applied that analysis to DNNs.

305, **TITLE:** Pain-Free Random Differential Privacy with Sensitivity Sampling

<http://proceedings.mlr.press/v70/rubinstein17a.html>

AUTHORS: Benjamin I. P. Rubinstein, Francesco Ald?

HIGHLIGHT: As an alternative, we propose a straightforward sampler for estimating sensitivity of non-private mechanisms.

306, **TITLE:** Enumerating Distinct Decision Trees
<http://proceedings.mlr.press/v70/ruggieri17a.html>
AUTHORS: Salvatore Ruggieri
HIGHLIGHT: We provide an exact enumeration procedure of the subsets that lead to all and only the distinct decision trees.

307, **TITLE:** Bayesian Boolean Matrix Factorisation
<http://proceedings.mlr.press/v70/rukat17a.html>
AUTHORS: Tammo Rukat, Chris C. Holmes, Michalis K. Titsias, Christopher Yau
HIGHLIGHT: We introduce the OrMachine, a probabilistic generative model for Boolean matrix factorisation and derive a Metropolised Gibbs sampler that facilitates efficient parallel posterior inference.

308, **TITLE:** Depth-Width Tradeoffs in Approximating Natural Functions with Neural Networks
<http://proceedings.mlr.press/v70/safran17a.html>
AUTHORS: Itay Safran, Ohad Shamir
HIGHLIGHT: We provide several new depth-based separation results for feed-forward neural networks, proving that various types of simple and natural functions can be better approximated using deeper networks than shallower ones, even if the shallower networks are much larger.

309, **TITLE:** Asymmetric Tri-training for Unsupervised Domain Adaptation
<http://proceedings.mlr.press/v70/saito17a.html>
AUTHORS: Kuniaki Saito, Yoshitaka Ushiku, Tatsuya Harada
HIGHLIGHT: In this paper, we propose the use of an asymmetric tri-training method for unsupervised domain adaptation, where we assign pseudo-labels to unlabeled samples and train the neural networks as if they are true labels.

310, **TITLE:** Semi-Supervised Classification Based on Classification from Positive and Unlabeled Data
<http://proceedings.mlr.press/v70/sakai17a.html>
AUTHORS: Tomoya Sakai, Marthinus Christoffel Plessis, Gang Niu, Masashi Sugiyama
HIGHLIGHT: In this paper, we extend PU classification to also incorporate negative data and propose a novel semi-supervised learning approach.

311, **TITLE:** Analytical Guarantees on Numerical Precision of Deep Neural Networks
<http://proceedings.mlr.press/v70/sakr17a.html>
AUTHORS: Charbel Sakr, Yongjune Kim, Naresh Shanbhag
HIGHLIGHT: We focus on numerical precision – a key parameter defining the complexity of neural networks.

312, **TITLE:** Hierarchy Through Composition with Multitask LMDPs
<http://proceedings.mlr.press/v70/saxe17a.html>
AUTHORS: Andrew M. Saxe, Adam C. Earle, Benjamin Rosman
HIGHLIGHT: We introduce the Multitask LMDP module, which maintains a parallel distributed representation of tasks and may be stacked to form deep hierarchies abstracted in space and time.

313, **TITLE:** Optimal Algorithms for Smooth and Strongly Convex Distributed Optimization in Networks
<http://proceedings.mlr.press/v70/scaman17a.html>
AUTHORS: Kevin Scaman, Francis Bach, S?bastien Bubeck, Yin Tat Lee, Laurent Massouli?
HIGHLIGHT: In this paper, we determine the optimal convergence rates for strongly convex and smooth distributed optimization in two settings: centralized and decentralized communications over a network.

314, **TITLE:** Adapting Kernel Representations Online Using Submodular Maximization
<http://proceedings.mlr.press/v70/schlegel17a.html>

- AUTHORS: Matthew Schlegel, Yangchen Pan, Jiecao Chen, Martha White
HIGHLIGHT: In this work, we develop an approximately submodular criterion for this setting, and an efficient online greedy submodular maximization algorithm for optimizing the criterion.
- 315, TITLE: Developing Bug-Free Machine Learning Systems With Formal Mathematics
<http://proceedings.mlr.press/v70/selsam17a.html>
AUTHORS: Daniel Selsam, Percy Liang, David L. Dill
HIGHLIGHT: We train a variational autoencoder using Certigrad and find the performance comparable to training the same model in TensorFlow.
- 316, TITLE: Identifying Best Interventions through Online Importance Sampling
<http://proceedings.mlr.press/v70/sen17a.html>
AUTHORS: Rajat Sen, Karthikeyan Shanmugam, Alexandros G. Dimakis, Sanjay Shakkottai
HIGHLIGHT: Motivated by applications in computational advertising and systems biology, we consider the problem of identifying the best out of several possible soft interventions at a source node S in an acyclic causal directed graph, to maximize the expected value of a target node Y (located downstream of S).
- 317, TITLE: Failures of Gradient-Based Deep Learning
<http://proceedings.mlr.press/v70/shalev-shwartz17a.html>
AUTHORS: Shai Shalev-Shwartz, Ohad Shamir, Shaked Shammah
HIGHLIGHT: We describe four types of simple problems, for which the gradient-based algorithms commonly used in deep learning either fail or suffer from significant difficulties.
- 318, TITLE: Estimating individual treatment effect: generalization bounds and algorithms
<http://proceedings.mlr.press/v70/shalit17a.html>
AUTHORS: Uri Shalit, Fredrik D. Johansson, David Sontag
HIGHLIGHT: We give a new theoretical analysis and family of algorithms for predicting individual treatment effect (ITE) from observational data, under the assumption known as strong ignorability.
- 319, TITLE: Online Learning with Local Permutations and Delayed Feedback
<http://proceedings.mlr.press/v70/shamir17a.html>
AUTHORS: Ohad Shamir, Liran Szlak
HIGHLIGHT: In this paper, we consider the applicability of this setting to convex online learning with delayed feedback, in which the feedback on the prediction made in round t arrives with some delay τ_t .
- 320, TITLE: Orthogonalized ALS: A Theoretically Principled Tensor Decomposition Algorithm for Practical Use
<http://proceedings.mlr.press/v70/sharan17a.html>
AUTHORS: Vatsal Sharan, Gregory Valiant
HIGHLIGHT: We propose a modification of the ALS approach that is as efficient as standard ALS, but provably recovers the true factors with random initialization under standard incoherence assumptions on the factors of the tensor.
- 321, TITLE: Differentially Private Ordinary Least Squares
<http://proceedings.mlr.press/v70/sheffet17a.html>
AUTHORS: Or Sheffet
HIGHLIGHT: Our work aims at achieving similar guarantees on data under differentially private estimators.
- 322, TITLE: On the Iteration Complexity of Support Recovery via Hard Thresholding Pursuit
<http://proceedings.mlr.press/v70/shen17a.html>
AUTHORS: Jie Shen, Ping Li
HIGHLIGHT: In this paper, we present a novel analysis for the hard thresholding pursuit (HTP) algorithm, showing that it exactly recovers the support of an arbitrary s -sparse signal within $O(k \log k)$ iterations via a properly chosen proxy function, where k is the condition number of the problem.

- 323, TITLE: GSOS: Gauss-Seidel Operator Splitting Algorithm for Multi-Term Nonsmooth Convex Composite Optimization
http://proceedings.mlr.press/v70/shen17b.html
AUTHORS: Li Shen, Wei Liu, Ganzhao Yuan, Shiqian Ma
HIGHLIGHT: In this paper, we propose a fast Gauss-Seidel Operator Splitting (GSOS) algorithm for addressing multi-term nonsmooth convex composite optimization, which has wide applications in machine learning, signal processing and statistics.
- 324, TITLE: World of Bits: An Open-Domain Platform for Web-Based Agents
http://proceedings.mlr.press/v70/shi17a.html
AUTHORS: Tianlin Shi, Andrej Karpathy, Linxi Fan, Jonathan Hernandez, Percy Liang
HIGHLIGHT: To foster reinforcement learning research in such settings, we introduce the World of Bits (WoB), a platform in which agents complete tasks on the Internet by performing low-level keyboard and mouse actions.
- 325, TITLE: Learning Important Features Through Propagating Activation Differences
http://proceedings.mlr.press/v70/shrikumar17a.html
AUTHORS: Avanti Shrikumar, Peyton Greenside, Anshul Kundaje
HIGHLIGHT: Here we present DeepLIFT (Deep Learning Important Features), a method for decomposing the output prediction of a neural network on a specific input by backpropagating the contributions of all neurons in the network to every feature of the input.
- 326, TITLE: Optimal Densification for Fast and Accurate Minwise Hashing
http://proceedings.mlr.press/v70/shrivastava17a.html
AUTHORS: Anshumali Shrivastava
HIGHLIGHT: In this paper, we provide a novel densification scheme which relies on carefully tailored 2-universal hashes.
- 327, TITLE: Bottleneck Conditional Density Estimation
http://proceedings.mlr.press/v70/shu17a.html
AUTHORS: Rui Shu, Hung H. Bui, Mohammad Ghavamzadeh
HIGHLIGHT: We introduce a new framework for training deep generative models for high-dimensional conditional density estimation.
- 328, TITLE: Attentive Recurrent Comparators
http://proceedings.mlr.press/v70/shyam17a.html
AUTHORS: Pranav Shyam, Shubham Gupta, Ambedkar Dukkipati
HIGHLIGHT: Using the representations extracted by ARCs, we develop a way of approximating a dynamic representation space and use it for one-shot learning.
- 329, TITLE: Gradient Boosted Decision Trees for High Dimensional Sparse Output
http://proceedings.mlr.press/v70/si17a.html
AUTHORS: Si Si, Huan Zhang, S. Sathiya Keerthi, Dhruv Mahajan, Inderjit S. Dhillon, Cho-Jui Hsieh
HIGHLIGHT: In this paper, we study the gradient boosted decision trees (GBDT) when the output space is high dimensional and sparse.
- 330, TITLE: The Predictron: End-To-End Learning and Planning
http://proceedings.mlr.press/v70/silver17a.html
AUTHORS: David Silver, Hado Hasselt, Matteo Hessel, Tom Schaul, Arthur Guez, Tim Harley, Gabriel Dulac-Arnold, David Reichert, Neil Rabinowitz, Andre Barreto, Thomas Degris
HIGHLIGHT: In this document we introduce the predictron architecture.
- 331, TITLE: Fractional Langevin Monte Carlo: Exploring Levy Driven Stochastic Differential Equations for Markov Chain Monte Carlo

<http://proceedings.mlr.press/v70/simsekli17a.html>

AUTHORS: Umut Simsekli

HIGHLIGHT: In this study, we extend classical LMC and develop a novel Fractional LMC (FLMC) framework that is based on a family of heavy-tailed distributions, called alpha-stable Levy distributions.

332, TITLE: Nonparanormal Information Estimation

<http://proceedings.mlr.press/v70/singh17a.html>

AUTHORS: Shashank Singh, Barnab Póczos

HIGHLIGHT: To address this, we propose estimators for mutual information when p is assumed to be a nonparanormal (or Gaussian copula) model, a semiparametric compromise between Gaussian and nonparametric extremes.

333, TITLE: High-Dimensional Structured Quantile Regression

<http://proceedings.mlr.press/v70/sivakumar17a.html>

AUTHORS: Vidyashankar Sivakumar, Arindam Banerjee

HIGHLIGHT: In this work we consider the problem of linear quantile regression in high dimensions where the number of predictor variables is much higher than the number of samples available for parameter estimation.

334, TITLE: Robust Budget Allocation via Continuous Submodular Functions

<http://proceedings.mlr.press/v70/staib17a.html>

AUTHORS: Matthew Staib, Stefanie Jegelka

HIGHLIGHT: We hence revisit a continuous version of the Budget Allocation or Bipartite Influence Maximization problem introduced by Alon et al. (2012) from a robust optimization perspective, where an adversary may choose the least favorable parameters within a confidence set.

335, TITLE: Probabilistic Submodular Maximization in Sub-Linear Time

<http://proceedings.mlr.press/v70/stan17a.html>

AUTHORS: Serban Stan, Morteza Zadimoghaddam, Andreas Krause, Amin Karbasi

HIGHLIGHT: In this paper, we consider optimizing submodular functions that are drawn from some unknown distribution.

As a remedy, we introduce the problem of sublinear time probabilistic submodular maximization: Given training examples of functions (e.g., via user feature vectors), we seek to reduce the ground set so that optimizing new functions drawn from the same distribution will provide almost as much value when restricted to the reduced ground set as when using the full set.

336, TITLE: Approximate Steepest Coordinate Descent

<http://proceedings.mlr.press/v70/stich17a.html>

AUTHORS: Sebastian U. Stich, Anant Raj, Martin Jaggi

HIGHLIGHT: We propose a new selection rule for the coordinate selection in coordinate descent methods for huge-scale optimization.

337, TITLE: Ordinal Graphical Models: A Tale of Two Approaches

<http://proceedings.mlr.press/v70/suggala17a.html>

AUTHORS: Arun Sai Suggala, Eunho Yang, Pradeep Ravikumar

HIGHLIGHT: In this paper, we theoretically investigate two classes of graphical models for ordinal data, corresponding to the two main categories of univariate ordinal distributions.

338, TITLE: Tensor Balancing on Statistical Manifold

<http://proceedings.mlr.press/v70/sugiyama17a.html>

AUTHORS: Mahito Sugiyama, Hiroyuki Nakahara, Koji Tsuda

HIGHLIGHT: We present an efficient balancing algorithm with quadratic convergence using Newton's method and show in numerical experiments that the proposed algorithm is several orders of magnitude faster than existing ones.

339, TITLE: Safety-Aware Algorithms for Adversarial Contextual Bandit

<http://proceedings.mlr.press/v70/sun17a.html>

AUTHORS: Wen Sun, Debadepta Dey, Ashish Kapoor

HIGHLIGHT: In this work we study the safe sequential decision making problem under the setting of adversarial contextual bandits with sequential risk constraints.

340, TITLE: Relative Fisher Information and Natural Gradient for Learning Large Modular Models

<http://proceedings.mlr.press/v70/sun17b.html>

AUTHORS: Ke Sun, Frank Nielsen

HIGHLIGHT: We provide an analysis on a list of commonly used components, and demonstrate how to use this concept to further improve optimization.

341, TITLE: meProp: Sparsified Back Propagation for Accelerated Deep Learning with Reduced Overfitting

<http://proceedings.mlr.press/v70/sun17c.html>

AUTHORS: Xu Sun, Xuancheng Ren, Shuming Ma, Houfeng Wang

HIGHLIGHT: We propose a simple yet effective technique for neural network learning.

342, TITLE: Deeply AggreVaTeD: Differentiable Imitation Learning for Sequential Prediction

<http://proceedings.mlr.press/v70/sun17d.html>

AUTHORS: Wen Sun, Arun Venkatraman, Geoffrey J. Gordon, Byron Boots, J. Andrew Bagnell

HIGHLIGHT: To take advantage of this extra information, we propose AggreVaTeD, an extension of the Imitation Learning (IL) approach of Ross & Bagnell (2014).

343, TITLE: Axiomatic Attribution for Deep Networks

<http://proceedings.mlr.press/v70/sundararajan17a.html>

AUTHORS: Mukund Sundararajan, Ankur Taly, Qiqi Yan

HIGHLIGHT: We study the problem of attributing the prediction of a deep network to its input features, a problem previously studied by several other works.

344, TITLE: Distributed Mean Estimation with Limited Communication

<http://proceedings.mlr.press/v70/suresh17a.html>

AUTHORS: Ananda Theertha Suresh, Felix X. Yu, Sanjiv Kumar, H. Brendan McMahan

HIGHLIGHT: Motivated by the need for distributed learning and optimization algorithms with low communication cost, we study communication efficient algorithms for distributed mean estimation.

345, TITLE: Selective Inference for Sparse High-Order Interaction Models

<http://proceedings.mlr.press/v70/suzumura17a.html>

AUTHORS: Shinya Suzumura, Kazuya Nakagawa, Yuta Umezu, Koji Tsuda, Ichiro Takeuchi

HIGHLIGHT: In this paper we study feature selection and statistical inference for sparse high-order interaction models.

346, TITLE: Coherent Probabilistic Forecasts for Hierarchical Time Series

<http://proceedings.mlr.press/v70/taieb17a.html>

AUTHORS: Souhaib Ben Taieb, James W. Taylor, Rob J. Hyndman

HIGHLIGHT: We consider the situation where probabilistic forecasts are needed for each series in the hierarchy, and propose an algorithm to compute predictive distributions rather than mean forecasts only.

347, TITLE: Partitioned Tensor Factorizations for Learning Mixed Membership Models

<http://proceedings.mlr.press/v70/tan17a.html>

AUTHORS: Zilong Tan, Sayan Mukherjee

HIGHLIGHT: We present an efficient algorithm for learning mixed membership models when the number of variables p is much larger than the number of hidden components k .

348, TITLE: Gradient Coding: Avoiding Stragglers in Distributed Learning

<http://proceedings.mlr.press/v70/tandon17a.html>

AUTHORS: Rashish Tandon, Qi Lei, Alexandros G. Dimakis, Nikos Karampatziakis
HIGHLIGHT: We propose a novel coding theoretic framework for mitigating stragglers in distributed learning.

349, TITLE: Gradient Projection Iterative Sketch for Large-Scale Constrained Least-Squares

<http://proceedings.mlr.press/v70/tang17a.html>

AUTHORS: Junqi Tang, Mohammad Golbabaee, Mike E. Davies
HIGHLIGHT: We propose a randomized first order optimization algorithm Gradient Projection Iterative Sketch (GPIS) and an accelerated variant for efficiently solving large scale constrained Least Squares (LS).

350, TITLE: Neural Networks and Rational Functions

<http://proceedings.mlr.press/v70/telgarsky17a.html>

AUTHORS: Matus Telgarsky
HIGHLIGHT: Neural Networks and Rational Functions

351, TITLE: Stochastic DCA for the Large-sum of Non-convex Functions Problem and its Application to Group Variable Selection in Classification

<http://proceedings.mlr.press/v70/thi17a.html>

AUTHORS: Hoai An Le Thi, Hoai Minh Le, Duy Nhat Phan, Bach Tran
HIGHLIGHT: In this paper, we present a stochastic version of DCA (Difference of Convex functions Algorithm) to solve a class of optimization problems whose objective function is a large sum of non-convex functions and a regularization term.

352, TITLE: An Analytical Formula of Population Gradient for two-layered ReLU network and its Applications in Convergence and Critical Point Analysis

<http://proceedings.mlr.press/v70/tian17a.html>

AUTHORS: Yuandong Tian
HIGHLIGHT: In this paper, we explore theoretical properties of training a two-layered ReLU network $g(\mathbf{x}; \mathbf{w}) = \sum_{j=1}^K \sigma(\mathbf{w}_j^\top \mathbf{x})$ with centered d -dimensional spherical Gaussian input \mathbf{x} ($\sigma = \text{ReLU}$).

353, TITLE: Evaluating the Variance of Likelihood-Ratio Gradient Estimators

<http://proceedings.mlr.press/v70/tokui17a.html>

AUTHORS: Seiya Tokui, Issei Sato
HIGHLIGHT: In this study, we establish a novel framework of gradient estimation that includes most of the common gradient estimators as special cases.

354, TITLE: Accelerating Eulerian Fluid Simulation With Convolutional Networks

<http://proceedings.mlr.press/v70/tompson17a.html>

AUTHORS: Jonathan Tompson, Kristofer Schlachter, Pablo Sprechmann, Ken Perlin
HIGHLIGHT: In this work, we propose a data-driven approach that leverages the approximation power of deep-learning with the precision of standard solvers to obtain fast and highly realistic simulations.

355, TITLE: Boosted Fitted Q-Iteration

<http://proceedings.mlr.press/v70/tosatto17a.html>

AUTHORS: Samuele Tosatto, Matteo Pirota, Carlo D'Eramo, Marcello Restelli
HIGHLIGHT: This paper is about the study of B-FQI, an Approximated Value Iteration (AVI) algorithm that exploits a boosting procedure to estimate the action-value function in reinforcement learning problems.

356, TITLE: Diameter-Based Active Learning

<http://proceedings.mlr.press/v70/tosh17a.html>

AUTHORS: Christopher Tosh, Sanjoy Dasgupta
HIGHLIGHT: We provide, for the first time, an efficient algorithm that is able to realize this upper bound, and we empirically demonstrate its good performance.

- 357, TITLE: Magnetic Hamiltonian Monte Carlo
<http://proceedings.mlr.press/v70/tripuraneni17a.html>
AUTHORS: Nilesh Tripuraneni, Mark Rowland, Zoubin Ghahramani, Richard Turner
HIGHLIGHT: In this paper, we present a generalization of HMC which exploits non-canonical Hamiltonian dynamics.
- 358, TITLE: Know-Evolve: Deep Temporal Reasoning for Dynamic Knowledge Graphs
<http://proceedings.mlr.press/v70/trivedi17a.html>
AUTHORS: Rakshit Trivedi, Hanjun Dai, Yichen Wang, Le Song
HIGHLIGHT: To this end, we present Know-Evolve, a novel deep evolutionary knowledge network that learns non-linearly evolving entity representations over time.
- 359, TITLE: Hyperplane Clustering via Dual Principal Component Pursuit
<http://proceedings.mlr.press/v70/tsakiris17a.html>
AUTHORS: Manolis C. Tsakiris, Ren? Vidal
HIGHLIGHT: In this paper we provide theoretical and algorithmic contributions to the problem of clustering data from a union of hyperplanes, by extending a recent subspace learning method called Dual Principal Component Pursuit (DPCP) to the multi-hyperplane case.
- 360, TITLE: Breaking Locality Accelerates Block Gauss-Seidel
<http://proceedings.mlr.press/v70/tu17a.html>
AUTHORS: Stephen Tu, Shivaram Venkataraman, Ashia C. Wilson, Alex Gittens, Michael I. Jordan, Benjamin Recht
HIGHLIGHT: Motivated by this finding, we analyze the accelerated block Gauss-Seidel algorithm in the random coordinate sampling setting.
- 361, TITLE: Multilabel Classification with Group Testing and Codes
<http://proceedings.mlr.press/v70/ubaru17a.html>
AUTHORS: Shashanka Ubaru, Arya Mazumdar
HIGHLIGHT: In this work, we propose a novel approach based on group testing to solve such large multilabel classification problems with sparse label vectors.
- 362, TITLE: Learning Stable Stochastic Nonlinear Dynamical Systems
<http://proceedings.mlr.press/v70/umlauft17a.html>
AUTHORS: Jonas Umlauft, Sandra Hirche
HIGHLIGHT: Therefore, this paper proposes a framework for learning stable stochastic systems from data.
- 363, TITLE: Learning Determinantal Point Processes with Moments and Cycles
<http://proceedings.mlr.press/v70/urschel17a.html>
AUTHORS: John Urschel, Victor-Emmanuel Brunel, Ankur Moitra, Philippe Rigollet
HIGHLIGHT: Our contribution is twofold: (i) we establish the optimal sample complexity achievable in this problem and show that it is governed by a natural parameter, which we call the cycle sparsity; (ii) we propose a provably fast combinatorial algorithm that implements the method of moments efficiently and achieves optimal sample complexity.
- 364, TITLE: Automatic Discovery of the Statistical Types of Variables in a Dataset
<http://proceedings.mlr.press/v70/valera17a.html>
AUTHORS: Isabel Valera, Zoubin Ghahramani
HIGHLIGHT: In this paper, we fill this gap by proposing a Bayesian method, which accurately discovers the statistical data types in both synthetic and real data.
- 365, TITLE: Model-Independent Online Learning for Influence Maximization

<http://proceedings.mlr.press/v70/vaswani17a.html>

AUTHORS: Sharan Vaswani, Branislav Kveton, Zheng Wen, Mohammad Ghavamzadeh, Laks V. S. Lakshmanan, Mark Schmidt

HIGHLIGHT: For this, we propose a pairwise-influence semi-bandit feedback model and develop a LinUCB-based bandit algorithm.

366, TITLE: FeUdal Networks for Hierarchical Reinforcement Learning

<http://proceedings.mlr.press/v70/vezhnevets17a.html>

AUTHORS: Alexander Sasha Vezhnevets, Simon Osindero, Tom Schaul, Nicolas Heess, Max Jaderberg, David Silver, Koray Kavukcuoglu

HIGHLIGHT: We introduce FeUdal Networks (FuNs): a novel architecture for hierarchical reinforcement learning.

367, TITLE: Scalable Multi-Class Gaussian Process Classification using Expectation Propagation

<http://proceedings.mlr.press/v70/villacampa-calvo17a.html>

AUTHORS: Carlos Villacampa-Calvo, Daniel Hernandez-Lobato

HIGHLIGHT: This paper describes an expectation propagation (EP) method for multi-class classification with Gaussian processes that scales well to very large datasets.

368, TITLE: Learning to Generate Long-term Future via Hierarchical Prediction

<http://proceedings.mlr.press/v70/villegas17a.html>

AUTHORS: Ruben Villegas, Jimei Yang, Yuliang Zou, Sungryull Sohn, Xunyu Lin, Honglak Lee

HIGHLIGHT: We propose a hierarchical approach for making long-term predictions of future frames.

369, TITLE: On orthogonality and learning recurrent networks with long term dependencies

<http://proceedings.mlr.press/v70/vorontsov17a.html>

AUTHORS: Eugene Vorontsov, Chiheb Trabelsi, Samuel Kadoury, Chris Pal

HIGHLIGHT: To perform this analysis, we propose a weight matrix factorization and parameterization strategy through which we can bound matrix norms and therein control the degree of expansivity induced during backpropagation.

370, TITLE: Fast Bayesian Intensity Estimation for the Permanental Process

<http://proceedings.mlr.press/v70/walder17a.html>

AUTHORS: Christian J. Walder, Adrian N. Bishop

HIGHLIGHT: In this paper we present a fast Bayesian inference scheme for the permanental process, a Cox process under which the square root of the intensity is a Gaussian process.

371, TITLE: Optimal and Adaptive Off-policy Evaluation in Contextual Bandits

<http://proceedings.mlr.press/v70/wang17a.html>

AUTHORS: Yu-Xiang Wang, Alekh Agarwal, Miroslav Dudík

HIGHLIGHT: We study the off-policy evaluation problem—estimating the value of a target policy using data collected by another policy—under the contextual bandit model.

372, TITLE: Capacity Releasing Diffusion for Speed and Locality

<http://proceedings.mlr.press/v70/wang17b.html>

AUTHORS: Di Wang, Kimon Fountoulakis, Monika Henzinger, Michael W. Mahoney, Satish Rao

HIGHLIGHT: We introduce a novel Capacity Releasing Diffusion (CRD) Process, which is both faster and stays more local than the classical spectral diffusion process.

373, TITLE: Sketched Ridge Regression: Optimization Perspective, Statistical Perspective, and Model Averaging

<http://proceedings.mlr.press/v70/wang17c.html>

AUTHORS: Shusen Wang, Alex Gittens, Michael W. Mahoney

HIGHLIGHT: We establish theoretically and empirically that model averaging greatly decreases this gap.

- 374, TITLE: Robust Gaussian Graphical Model Estimation with Arbitrary Corruption
<http://proceedings.mlr.press/v70/wang17d.html>
AUTHORS: Lingxiao Wang, Quanquan Gu
HIGHLIGHT: We propose a robust estimator for the sparse precision matrix in the high-dimensional regime.
- 375, TITLE: Max-value Entropy Search for Efficient Bayesian Optimization
<http://proceedings.mlr.press/v70/wang17e.html>
AUTHORS: Zi Wang, Stefanie Jegelka
HIGHLIGHT: We propose a new criterion, Max-value Entropy Search (MES), that instead uses the information about the maximum function value.
- 376, TITLE: Efficient Distributed Learning with Sparsity
<http://proceedings.mlr.press/v70/wang17f.html>
AUTHORS: Jialei Wang, Mladen Kolar, Nathan Srebro, Tong Zhang
HIGHLIGHT: We propose a novel, efficient approach for distributed sparse learning with observations randomly partitioned across machines.
- 377, TITLE: Robust Probabilistic Modeling with Bayesian Data Reweighting
<http://proceedings.mlr.press/v70/wang17g.html>
AUTHORS: Yixin Wang, Alp Kucukelbir, David M. Blei
HIGHLIGHT: We propose a way to systematically detect and mitigate mismatch of a large class of probabilistic models.
- 378, TITLE: Batched High-dimensional Bayesian Optimization via Structural Kernel Learning
<http://proceedings.mlr.press/v70/wang17h.html>
AUTHORS: Zi Wang, Chengtao Li, Stefanie Jegelka, Pushmeet Kohli
HIGHLIGHT: In this paper, we propose to tackle these challenges by (1) assuming a latent additive structure in the function and inferring it properly for more efficient and effective BO, and (2) performing multiple evaluations in parallel to reduce the number of iterations required by the method.
- 379, TITLE: Tensor Decomposition via Simultaneous Power Iteration
<http://proceedings.mlr.press/v70/wang17i.html>
AUTHORS: Po-An Wang, Chi-Jen Lu
HIGHLIGHT: In this paper, we show how to find the eigenvectors simultaneously with the help of a new initialization procedure.
- 380, TITLE: Sequence Modeling via Segmentations
<http://proceedings.mlr.press/v70/wang17j.html>
AUTHORS: Chong Wang, Yining Wang, Po-Sen Huang, Abdelrahman Mohamed, Dengyong Zhou, Li Deng
HIGHLIGHT: In this paper, we present a probabilistic model for sequences via their segmentations.
- 381, TITLE: Variational Policy for Guiding Point Processes
<http://proceedings.mlr.press/v70/wang17k.html>
AUTHORS: Yichen Wang, Grady Williams, Evangelos Theodorou, Le Song
HIGHLIGHT: In this paper, we consider the problem of how to design the optimal control policy for point processes, such that the stochastic system driven by the point process is steered to a target state.
- 382, TITLE: Exploiting Strong Convexity from Data with Primal-Dual First-Order Algorithms
<http://proceedings.mlr.press/v70/wang17l.html>
AUTHORS: Jialei Wang, Lin Xiao
HIGHLIGHT: In this paper, we develop both batch and randomized primal-dual algorithms that can exploit strong convexity from data adaptively and are capable of achieving linear convergence even without regularization.

383, TITLE: Beyond Filters: Compact Feature Map for Portable Deep Model
<http://proceedings.mlr.press/v70/wang17m.html>
AUTHORS: Yunhe Wang, Chang Xu, Chao Xu, Dacheng Tao
HIGHLIGHT: We propose to extract intrinsic representation of the feature maps and preserve the discriminability of the features.

384, TITLE: A Unified Variance Reduction-Based Framework for Nonconvex Low-Rank Matrix Recovery
<http://proceedings.mlr.press/v70/wang17n.html>
AUTHORS: Lingxiao Wang, Xiao Zhang, Quanquan Gu
HIGHLIGHT: We propose a generic framework based on a new stochastic variance-reduced gradient descent algorithm for accelerating nonconvex low-rank matrix recovery.

385, TITLE: Source-Target Similarity Modelings for Multi-Source Transfer Gaussian Process Regression
<http://proceedings.mlr.press/v70/wei17a.html>
AUTHORS: Pengfei Wei, Ramon Sagarna, Yiping Ke, Yew-Soon Ong, Chi-Keong Goh
HIGHLIGHT: In this paper, we study different approaches based on Gaussian process models to solve the multi-source transfer regression problem.

386, TITLE: Latent Intention Dialogue Models
<http://proceedings.mlr.press/v70/wen17a.html>
AUTHORS: Tsung-Hsien Wen, Yishu Miao, Phil Blunsom, Steve Young
HIGHLIGHT: Developing a dialogue agent that is capable of making autonomous decisions and communicating by natural language is one of the long-term goals of machine learning research.

387, TITLE: Unifying Task Specification in Reinforcement Learning
<http://proceedings.mlr.press/v70/white17a.html>
AUTHORS: Martha White
HIGHLIGHT: In this work, we introduce the RL task formalism, that provides a unification through simple constructs including a generalization to transition-based discounting.

388, TITLE: Learned Optimizers that Scale and Generalize
<http://proceedings.mlr.press/v70/wichrowska17a.html>
AUTHORS: Olga Wichrowska, Niru Maheswaranathan, Matthew W. Hoffman, Sergio Gomez Colmenarejo, Misha Denil, Nando Freitas, Jascha Sohl-Dickstein
HIGHLIGHT: We introduce a learned gradient descent optimizer that generalizes well to new tasks, and which has significantly reduced memory and computation overhead.

389, TITLE: Exact Inference for Integer Latent-Variable Models
<http://proceedings.mlr.press/v70/winner17a.html>
AUTHORS: Kevin Winner, Debora Sujono, Dan Sheldon
HIGHLIGHT: In this paper we introduce a new approach for inference with PGFs: instead of manipulating PGFs symbolically, we adapt techniques from the autodiff literature to compute the higher-order derivatives necessary for inference.

390, TITLE: Tensor Belief Propagation
<http://proceedings.mlr.press/v70/wrigley17a.html>
AUTHORS: Andrew Wrigley, Wee Sun Lee, Nan Ye
HIGHLIGHT: We propose a new approximate inference algorithm for graphical models, tensor belief propagation, based on approximating the messages passed in the junction tree algorithm.

391, TITLE: A Unified View of Multi-Label Performance Measures
<http://proceedings.mlr.press/v70/wu17a.html>

- AUTHORS: Xi-Zhu Wu, Zhi-Hua Zhou
HIGHLIGHT: In this paper, we propose a unified margin view to revisit eleven performance measures in multi-label classification.
- 392, TITLE: Dual Supervised Learning
<http://proceedings.mlr.press/v70/xia17a.html>
AUTHORS: Yingce Xia, Tao Qin, Wei Chen, Jiang Bian, Nenghai Yu, Tie-Yan Liu
HIGHLIGHT: In this work, we propose training the models of two dual tasks simultaneously, and explicitly exploiting the probabilistic correlation between them to regularize the training process.
- 393, TITLE: Learning Latent Space Models with Angular Constraints
<http://proceedings.mlr.press/v70/xie17a.html>
AUTHORS: Pengtao Xie, Yuntian Deng, Yi Zhou, Abhimanu Kumar, Yaoliang Yu, James Zou, Eric P. Xing
HIGHLIGHT: To bridge this gap, we propose a new diversity-promoting approach that is both theoretically analyzable and empirically effective.
- 394, TITLE: Uncorrelation and Evenness: a New Diversity-Promoting Regularizer
<http://proceedings.mlr.press/v70/xie17b.html>
AUTHORS: Pengtao Xie, Aarti Singh, Eric P. Xing
HIGHLIGHT: To cope with two challenges in LSMs: (1) how to capture infrequent patterns when pattern frequency is imbalanced and (2) how to reduce model size without sacrificing their expressiveness, several studies have been proposed to “diversify” LSMs, which design regularizers to encourage the components therein to be “diverse”.
- 395, TITLE: Stochastic Convex Optimization: Faster Local Growth Implies Faster Global Convergence
<http://proceedings.mlr.press/v70/xu17a.html>
AUTHORS: Yi Xu, Qihang Lin, Tianbao Yang
HIGHLIGHT: In this paper, a new theory is developed for first-order stochastic convex optimization, showing that the global convergence rate is sufficiently quantified by a local growth rate of the objective function in a neighborhood of the optimal solutions.
- 396, TITLE: Learning Hawkes Processes from Short Doubly-Censored Event Sequences
<http://proceedings.mlr.press/v70/xu17b.html>
AUTHORS: Hongteng Xu, Dixin Luo, Hongyuan Zha
HIGHLIGHT: In this paper, we study this critical problem of quantitative asynchronous event sequence analysis under the framework of Hawkes processes by leveraging the general idea of data synthesis.
- 397, TITLE: Adaptive Consensus ADMM for Distributed Optimization
<http://proceedings.mlr.press/v70/xu17c.html>
AUTHORS: Zheng Xu, Gavin Taylor, Hao Li, M?rio A. T. Figueiredo, Xiaoming Yuan, Tom Goldstein
HIGHLIGHT: We present a $O(1/k)$ convergence rate for adaptive ADMM methods with node-specific parameters, and propose adaptive consensus ADMM (ACADMM), which automatically tunes parameters without user oversight.
- 398, TITLE: High-dimensional Non-Gaussian Single Index Models via Thresholded Score Function Estimation
<http://proceedings.mlr.press/v70/yang17a.html>
AUTHORS: Zhuoran Yang, Krishnakumar Balasubramanian, Han Liu
HIGHLIGHT: Utilizing Stein’s Lemma, we propose estimators based on the score function of the covariate.
- 399, TITLE: Towards K-means-friendly Spaces: Simultaneous Deep Learning and Clustering
<http://proceedings.mlr.press/v70/yang17b.html>
AUTHORS: Bo Yang, Xiao Fu, Nicholas D. Sidiropoulos, Mingyi Hong
HIGHLIGHT: In this work, we assume that this transformation is an unknown and possibly nonlinear function.
- 400, TITLE: On The Projection Operator to A Three-view Cardinality Constrained Set

<http://proceedings.mlr.press/v70/yang17c.html>

AUTHORS: Haichuan Yang, Shupeng Gui, Chuyang Ke, Daniel Stefankovic, Ryohei Fujimaki, Ji Liu
HIGHLIGHT: In this paper, we consider the scenario where the overlapped cardinality constraints satisfy a Three-view Cardinality Structure (TVCS), which reflects the natural restriction in many applications, such as identification of gene regulatory networks and task-worker assignment problem.

401, TITLE: Improved Variational Autoencoders for Text Modeling using Dilated Convolutions

<http://proceedings.mlr.press/v70/yang17d.html>

AUTHORS: Zichao Yang, Zhiting Hu, Ruslan Salakhutdinov, Taylor Berg-Kirkpatrick
HIGHLIGHT: In this paper, we experiment with a new type of decoder for VAE: a dilated CNN.

402, TITLE: Tensor-Train Recurrent Neural Networks for Video Classification

<http://proceedings.mlr.press/v70/yang17e.html>

AUTHORS: Yinchong Yang, Denis Krompass, Volker Tresp
HIGHLIGHT: To address this challenge, we propose a new, more general and efficient approach by factorizing the input-to-hidden weight matrix using Tensor-Train decomposition which is trained simultaneously with the weights themselves.

403, TITLE: A Richer Theory of Convex Constrained Optimization with Reduced Projections and Improved Rates

<http://proceedings.mlr.press/v70/yang17f.html>

AUTHORS: Tianbao Yang, Qihang Lin, Lijun Zhang
HIGHLIGHT: In this paper, we develop projection reduced optimization algorithms for both smooth and non-smooth optimization with improved convergence rates under a certain regularity condition of the constraint function.

404, TITLE: Sparse + Group-Sparse Dirty Models: Statistical Guarantees without Unreasonable Conditions and a Case for Non-Convexity

<http://proceedings.mlr.press/v70/yang17g.html>

AUTHORS: Eunho Yang, Aur?lie C. Lozano
HIGHLIGHT: In this paper, we fill the gap between the practical success and suboptimal analysis of sparse + group-sparse models, by providing the first consistency results that do not require unrealistic assumptions.

405, TITLE: Scalable Bayesian Rule Lists

<http://proceedings.mlr.press/v70/yang17h.html>

AUTHORS: Hongyu Yang, Cynthia Rudin, Margo Seltzer
HIGHLIGHT: We present an algorithm for building probabilistic rule lists that is two orders of magnitude faster than previous work.

406, TITLE: Approximate Newton Methods and Their Local Convergence

<http://proceedings.mlr.press/v70/ye17a.html>

AUTHORS: Haishan Ye, Luo Luo, Zhihua Zhang
HIGHLIGHT: In this paper, we aim to fill these gaps.

407, TITLE: A Simulated Annealing Based Inexact Oracle for Wasserstein Loss Minimization

<http://proceedings.mlr.press/v70/ye17b.html>

AUTHORS: Jianbo Ye, James Z. Wang, Jia Li
HIGHLIGHT: In this paper, we introduce a stochastic approach based on simulated annealing for solving WLMs.

408, TITLE: Latent Feature Lasso

<http://proceedings.mlr.press/v70/ye17a.html>

AUTHORS: Ian En-Hsu Yen, Wei-Cheng Lee, Sung-En Chang, Arun Sai Suggala, Shou-De Lin, Pradeep Ravikumar

HIGHLIGHT: In this paper, we address this outstanding problem of tractable estimation of LFM's via a novel atomic-norm regularization, which gives an algorithm with polynomial run-time and sample complexity without impractical assumptions on the data distribution.

409, TITLE: Combined Group and Exclusive Sparsity for Deep Neural Networks
<http://proceedings.mlr.press/v70/yoon17a.html>

AUTHORS: Jaehong Yoon, Sung Ju Hwang

HIGHLIGHT: To resolve this issue, we propose a sparsity regularization method that exploits both positive and negative correlations among the features to enforce the network to be sparse, and at the same time remove any redundancies among the features to fully utilize the capacity of the network.

410, TITLE: Latent LSTM Allocation: Joint Clustering and Non-Linear Dynamic Modeling of Sequence Data
<http://proceedings.mlr.press/v70/zaheer17a.html>

AUTHORS: Manzil Zaheer, Amr Ahmed, Alexander J. Smola

HIGHLIGHT: In this paper, we introduce Latent LSTM Allocation (LLA) for user modeling combining hierarchical Bayesian models with LSTMs.

411, TITLE: Canopy Fast Sampling with Cover Trees

<http://proceedings.mlr.press/v70/zaheer17b.html>

AUTHORS: Manzil Zaheer, Satwik Kottur, Amr Ahmed, Jos? Moura, Alex Smola

HIGHLIGHT: In this work, we propose Canopy, a sampler based on Cover Trees that is exact, has guaranteed runtime logarithmic in the number of atoms, and is provably polynomial in the inherent dimensionality of the underlying parameter space.

412, TITLE: Continual Learning Through Synaptic Intelligence

<http://proceedings.mlr.press/v70/zenke17a.html>

AUTHORS: Friedemann Zenke, Ben Poole, Surya Ganguli

HIGHLIGHT: In this study, we introduce intelligent synapses that bring some of this biological complexity into artificial neural networks.

413, TITLE: Stochastic Gradient Monomial Gamma Sampler

<http://proceedings.mlr.press/v70/zhang17a.html>

AUTHORS: Yizhe Zhang, Changyou Chen, Zhe Gan, Ricardo Henao, Lawrence Carin

HIGHLIGHT: We propose a generalized framework to improve the sampling efficiency of stochastic gradient MCMC, by leveraging a generalized kinetics that delivers superior stationary mixing, especially in multimodal distributions, and propose several techniques to overcome the practical issues.

414, TITLE: Adversarial Feature Matching for Text Generation

<http://proceedings.mlr.press/v70/zhang17b.html>

AUTHORS: Yizhe Zhang, Zhe Gan, Kai Fan, Zhi Chen, Ricardo Henao, Dinghan Shen, Lawrence Carin

HIGHLIGHT: We propose a framework for generating realistic text via adversarial training.

415, TITLE: Scaling Up Sparse Support Vector Machines by Simultaneous Feature and Sample Reduction

<http://proceedings.mlr.press/v70/zhang17c.html>

AUTHORS: Weizhong Zhang, Bin Hong, Wei Liu, Jieping Ye, Deng Cai, Xiaofei He, Jie Wang

HIGHLIGHT: By noting that sparse SVMs induce sparsities in both feature and sample spaces, we propose a novel approach, which is based on accurate estimations of the primal and dual optima of sparse SVMs, to simultaneously identify the features and samples that are guaranteed to be irrelevant to the outputs.

416, TITLE: Re-visiting Learning on Hypergraphs: Confidence Interval and Subgradient Method

<http://proceedings.mlr.press/v70/zhang17d.html>

AUTHORS: Chenzi Zhang, Shuguang Hu, Zhihao Gavin Tang, T-H. Hubert Chan

HIGHLIGHT: We revisit semi-supervised learning on hypergraphs.

417, TITLE: ZipML: Training Linear Models with End-to-End Low Precision, and a Little Bit of Deep Learning
<http://proceedings.mlr.press/v70/zhang17e.html>
AUTHORS: Hantian Zhang, Jerry Li, Kaan Kara, Dan Alistarh, Ji Liu, Ce Zhang
HIGHLIGHT: We mainly focus on linear models, and the answer is yes for linear models.

418, TITLE: Convexified Convolutional Neural Networks
<http://proceedings.mlr.press/v70/zhang17f.html>
AUTHORS: Yuchen Zhang, Percy Liang, Martin J. Wainwright
HIGHLIGHT: We describe the class of convexified convolutional neural networks (CCNNs), which capture the parameter sharing of convolutional neural networks in a convex manner.

419, TITLE: Projection-free Distributed Online Learning in Networks
<http://proceedings.mlr.press/v70/zhang17g.html>
AUTHORS: Wenpeng Zhang, Peilin Zhao, Wenwu Zhu, Steven C. H. Hoi, Tong Zhang
HIGHLIGHT: In this paper, we fill this gap by proposing the distributed online conditional gradient algorithm, which eschews the expensive projection operation needed in its counterpart algorithms by exploiting much simpler linear optimization steps.

420, TITLE: Multi-Class Optimal Margin Distribution Machine
<http://proceedings.mlr.press/v70/zhang17h.html>
AUTHORS: Teng Zhang, Zhi-Hua Zhou
HIGHLIGHT: In this paper, we propose mcODM (multi-class Optimal margin Distribution Machine), which can solve this problem efficiently.

421, TITLE: Leveraging Node Attributes for Incomplete Relational Data
<http://proceedings.mlr.press/v70/zhao17a.html>
AUTHORS: He Zhao, Lan Du, Wray Buntine
HIGHLIGHT: This paper presents a Bayesian probabilistic approach that incorporates various kinds of node attributes encoded in binary form in relational models with Poisson likelihood.

422, TITLE: Theoretical Properties for Neural Networks with Weight Matrices of Low Displacement Rank
<http://proceedings.mlr.press/v70/zhao17b.html>
AUTHORS: Liang Zhao, Siyu Liao, Yanzhi Wang, Zhe Li, Jian Tang, Bo Yuan
HIGHLIGHT: This paper gives theoretical study on LDR neural networks.

423, TITLE: Learning Hierarchical Features from Deep Generative Models
<http://proceedings.mlr.press/v70/zhao17c.html>
AUTHORS: Shengjia Zhao, Jiaming Song, Stefano Ermon
HIGHLIGHT: In this paper, we prove that hierarchical latent variable models do not take advantage of the hierarchical structure when trained with existing variational methods, and provide some limitations on the kind of features existing models can learn.

424, TITLE: Learning Sleep Stages from Radio Signals: A Conditional Adversarial Architecture
<http://proceedings.mlr.press/v70/zhao17d.html>
AUTHORS: Mingmin Zhao, Shichao Yue, Dina Katabi, Tommi S. Jaakkola, Matt T. Bianchi
HIGHLIGHT: We introduce a new predictive model that combines convolutional and recurrent neural networks to extract sleep-specific subject-invariant features from RF signals and capture the temporal progression of sleep.

425, TITLE: Follow the Moving Leader in Deep Learning
<http://proceedings.mlr.press/v70/zheng17a.html>
AUTHORS: Shuai Zheng, James T. Kwok
HIGHLIGHT: Inspired by the close connection between stochastic optimization and online learning, we propose a variant of the follow the regularized leader (FTRL) algorithm called follow the moving leader (FTML).

- 426, TITLE: Asynchronous Stochastic Gradient Descent with Delay Compensation
<http://proceedings.mlr.press/v70/zheng17b.html>
AUTHORS: Shuxin Zheng, Qi Meng, Taifeng Wang, Wei Chen, Nenghai Yu, Zhi-Ming Ma, Tie-Yan Liu
HIGHLIGHT: We propose a novel technology to compensate this delay, so as to make the optimization behavior of ASGD closer to that of sequential SGD.
- 427, TITLE: Collect at Once, Use Effectively: Making Non-interactive Locally Private Learning Possible
<http://proceedings.mlr.press/v70/zheng17c.html>
AUTHORS: Kai Zheng, Wenlong Mou, Liwei Wang
HIGHLIGHT: In this paper, we extend the frontiers of Non-interactive LDP learning and estimation from several aspects.
- 428, TITLE: Recovery Guarantees for One-hidden-layer Neural Networks
<http://proceedings.mlr.press/v70/zhong17a.html>
AUTHORS: Kai Zhong, Zhao Song, Prateek Jain, Peter L. Bartlett, Inderjit S. Dhillon
HIGHLIGHT: In this paper, we consider regression problems with one-hidden-layer neural networks (1NNs).
- 429, TITLE: Stochastic Adaptive Quasi-Newton Methods for Minimizing Expected Values
<http://proceedings.mlr.press/v70/zhou17a.html>
AUTHORS: Chaoxu Zhou, Wenbo Gao, Donald Goldfarb
HIGHLIGHT: We propose a novel class of stochastic, adaptive methods for minimizing self-concordant functions which can be expressed as an expected value.
- 430, TITLE: Identify the Nash Equilibrium in Static Games with Random Payoffs
<http://proceedings.mlr.press/v70/zhou17b.html>
AUTHORS: Yichi Zhou, Jialian Li, Jun Zhu
HIGHLIGHT: We introduce a multi-armed bandit model to this problem due to its ability to find the best arm efficiently among random arms and propose two algorithms for this problem—LUCB-G based on the confidence bounds and a racing algorithm based on successive action elimination.
- 431, TITLE: When can Multi-Site Datasets be Pooled for Regression? Hypothesis Tests, ℓ_2 -consistency and Neuroscience Applications
<http://proceedings.mlr.press/v70/zhou17c.html>
AUTHORS: Hao Henry Zhou, Yilin Zhang, Vamsi K. Ithapu, Sterling C. Johnson, Grace Wahba, Vikas Singh
HIGHLIGHT: In this paper, we present a hypothesis test to answer this question, both for classical and high dimensional linear regression.
- 432, TITLE: High-Dimensional Variance-Reduced Stochastic Gradient Expectation-Maximization Algorithm
<http://proceedings.mlr.press/v70/zhu17a.html>
AUTHORS: Rongda Zhu, Lingxiao Wang, Chengxiang Zhai, Quanquan Gu
HIGHLIGHT: We propose a generic stochastic expectation-maximization (EM) algorithm for the estimation of high-dimensional latent variable models.
- 433, TITLE: Recurrent Highway Networks
<http://proceedings.mlr.press/v70/zilly17a.html>
AUTHORS: Julian Georg Zilly, Rupesh Kumar Srivastava, Jan Koutník, Jürgen Schmidhuber
HIGHLIGHT: We introduce a novel theoretical analysis of recurrent networks based on Gersgorin’s circle theorem that illuminates several modeling and optimization issues and improves our understanding of the LSTM cell.
- 434, TITLE: Online Learning to Rank in Stochastic Click Models
<http://proceedings.mlr.press/v70/zoghi17a.html>

AUTHORS: Masrour Zoghi, Tomas Tunys, Mohammad Ghavamzadeh, Branislav Kveton, Csaba Szepesvari,
Zheng Wen
HIGHLIGHT: In this work, we propose BatchRank, the first online learning to rank algorithm for a broad class of
click models.