

# **35th International Conference on Machine Learning (ICML 2018)**

Stockholm, Sweden  
10-15 July 2018

Volume 1 of 13

**Editors:**

**Jennifer Dy**

**Andreas Krause**

ISBN: 978-1-5108-6796-3

**Printed from e-media with permission by:**

Curran Associates, Inc.  
57 Morehouse Lane  
Red Hook, NY 12571



**Some format issues inherent in the e-media version may also appear in this print version.**

Copyright© (2018) by the Authors  
All rights reserved.

Printed by Curran Associates, Inc. (2018)

For permission requests, please contact International Machine Learning Society (IMLS)  
at the address below.

International Machine Learning Society (IMLS)  
c/o Priscilla Rasmussen  
209 North Eighth Street  
Stroudsburg PA 18360  
USA

Phone: (570) 476-8006  
Fax: (570) 476-0860

rasmussen@ptd.net

**Additional copies of this publication are available from:**

Curran Associates, Inc.  
57 Morehouse Lane  
Red Hook, NY 12571 USA  
Phone: 845-758-0400  
Fax: 845-758-2633  
Email: curran@proceedings.com  
Web: www.proceedings.com

# TABLE OF CONTENTS

## VOLUME 1

<b>Improved Regret Bounds for Thompson Sampling in Linear Quadratic Control Problems</b> .....	1
<i>Marc Abeille, Alessandro Lazaric</i>	
<b>State Abstractions for Lifelong Reinforcement Learning</b> .....	17
<i>David Abel, Dilip Arumugam, Lucas Lehnert, Michael Littman</i>	
<b>Policy and Value Transfer in Lifelong Reinforcement Learning</b> .....	30
<i>David Abel, Yuu Jinnai, Sophie Yue Guo, George Konidaris, Michael Littman</i>	
<b>INSPECTRE: Privately Estimating the Unseen</b> .....	42
<i>Jayadev Acharya, Gautam Kamath, Ziteng Sun, Huanyu Zhang</i>	
<b>Learning Representations and Generative Models for 3D Point Clouds</b> .....	67
<i>Panos Achlioptas, Olga Diamanti, Ioannis Mitliagkas, Leonidas Guibas</i>	
<b>Discovering Interpretable Representations for Both Deep Generative and Discriminative Models</b> .....	86
<i>Tameem Adel, Zoubin Ghahramani, Adrian Weller</i>	
<b>A Reductions Approach to Fair Classification</b> .....	102
<i>Alekh Agarwal, Alina Beygelzimer, Miroslav Dudik, John Langford, Hanna Wallach</i>	
<b>Accelerated Spectral Ranking</b> .....	120
<i>Arpit Agarwal, Prathamesh Patil, Shivani Agarwal</i>	
<b>MISSION: Ultra Large-Scale Feature Selection using Count-Sketches</b> .....	143
<i>Amirali Aghazadeh, Ryan Spring, Daniel Lejeune, Gautam Dasarathy, Anshumali Shrivastava, Richard G. Baraniuk</i>	
<b>Minimal I-MAP MCMC for Scalable Structure Discovery in Causal DAG Models</b> .....	155
<i>Raj Agrawal, Caroline Uhler, Tamara Broderick</i>	
<b>Proportional Allocation: Simple, Distributed, and Diverse Matching with High Entropy</b> .....	171
<i>Shipra Agrawal, Morteza Zadimoghaddam, Vahab Mirrokni</i>	
<b>Bucket Renormalization for Approximate Inference</b> .....	183
<i>Sungsoo Ahn, Michael Chertkov, Adrian Weller, Jinwoo Shin</i>	
<b>oi-VAE: Output Interpretable VAEs for Nonlinear Group Factor Analysis</b> .....	194
<i>Samuel K. Ainsworth, Nicholas J. Foti, Adrian K. C. Lee, Emily B. Fox</i>	
<b>Limits of Estimating Heterogeneous Treatment Effects: Guidelines for Practical Algorithm Design</b> .....	209
<i>Ahmed Alaa, Mihaela Schaar</i>	
<b>AutoPrognosis: Automated Clinical Prognostic Modeling via Bayesian Optimization with Structured Kernel Learning</b> .....	223
<i>Ahmed Alaa, Mihaela Schaar</i>	
<b>Information Theoretic Guarantees for Empirical Risk Minimization with Applications to Model Selection and Large-Scale Optimization</b> .....	235
<i>Ibrahim Alabdulmohsin</i>	
<b>Fixing a Broken ELBO</b> .....	245
<i>Alexander Alemi, Ben Poole, Ian Fischer, Joshua Dillon, Rif A. Saurous, Kevin Murphy</i>	
<b>Differentially Private Identity and Equivalence Testing of Discrete Distributions</b> .....	266
<i>Maryam Aliakbarpour, Ilias Diakonikolas, Ronitt Rubinfeld</i>	
<b>Katyusha X: Practical Momentum Method for Stochastic Sum-of-Nonconvex Optimization</b> .....	284
<i>Zeyuan Allen-Zhu</i>	
<b>Make the Minority Great Again: First-Order Regret Bound for Contextual Bandits</b> .....	291
<i>Zeyuan Allen-Zhu, Sebastien Bubeck, Yuanzhi Li</i>	
<b>Augmented CycleGAN: Learning Many-to-Many Mappings from Unpaired Data</b> .....	300
<i>Amjad Almahairi, Sai Rajeshwar, Alessandro Sordani, Philip Bachman, Aaron Courville</i>	
<b>Meta-Learning by Adjusting Priors Based on Extended PAC-Bayes Theory</b> .....	310
<i>Ron Amit, Ron Meir</i>	
<b>MAGAN: Aligning Biological Manifolds</b> .....	327
<i>Matthew Amodio, Smita Krishnaswamy</i>	
<b>Subspace Embedding and Linear Regression with Orlicz Norm</b> .....	336
<i>Alexandr Andoni, Chengyu Lin, Ying Sheng, Peilin Zhong, Ruiqi Zhong</i>	
<b>Efficient Gradient-Free Variational Inference using Policy Search</b> .....	359
<i>Oleg Arenz, Mingjun Zhong, Gerhard Neumann</i>	
<b>On the Optimization of Deep Networks: Implicit Acceleration by Overparameterization</b> .....	372
<i>Sanjeev Arora, Nadav Cohen, Elad Hazan</i>	
<b>Stronger Generalization Bounds for Deep Nets via a Compression Approach</b> .....	390
<i>Sanjeev Arora, Rong Ge, Behnam Neyshabur, Yi Zhang</i>	
<b>Lipschitz Continuity in Model-based Reinforcement Learning</b> .....	419
<i>Kavosh Asadi, Dipendra Misra, Michael Littman</i>	
<b>Obfuscated Gradients Give a False Sense of Security: Circumventing Defenses to Adversarial Examples</b> .....	436
<i>Anish Athalye, Nicholas Carlini, David Wagner</i>	
<b>Synthesizing Robust Adversarial Examples</b> .....	449
<i>Anish Athalye, Logan Engstrom, Andrew Ilyas, Kevin Kwok</i>	
<b>Clustering Semi-Random Mixtures of Gaussians</b> .....	469
<i>Pranjal Awasthi, Aravindan Vijayaraghavan</i>	

<b>Contextual Graph Markov Model: A Deep and Generative Approach to Graph Processing</b> .....	495
<i>Davide Bacciu, Federico Errica, Alessio Micheli</i>	
<b>Greed is Still Good: Maximizing Monotone Submodular+Supermodular (BP) Functions</b> .....	505
<i>Wenruo Bai, Jeffrey Bilmes</i>	
<b>Comparing Dynamics: Deep Neural Networks versus Glassy Systems</b> .....	526
<i>Marco Baity-Jesi, Levent Sagun, Mario Geiger, Stefano Spigler, Gerard Ben Arous, Chiara Cammarota, Yann Lecun, Matthieu Wyart, Giulio Biroli</i>	
<b>SMAC: Simultaneous Mapping and Clustering Using Spectral Decompositions</b> .....	536
<i>Chandrajit Bajaj, Tingran Gao, Zihang He, Qixing Huang, Zhenxiao Liang</i>	
<b>A Boo(n) for Evaluating Architecture Performance</b> .....	570
<i>Ondrej Bajgar, Rudolf Kadlec, Jan Kleindienst</i>	
<b>Learning to Branch</b> .....	582
<i>Maria-Florina Balcan, Travis Dick, Tuomas Sandholm, Ellen Vitercik</i>	
<b>The Mechanics of n-Player Differentiable Games</b> .....	619
<i>David Balduzzi, Sebastien Racaniere, James Martens, Jakob Foerster, Karl Tuyls, Thore Graepel</i>	
<b>Spline Filters For End-to-End Deep Learning</b> .....	636
<i>Randall Balestriero, Romain Cosentino, Herve Glotin, Richard Baraniuk</i>	
<b>A Spline Theory of Deep Networks</b> .....	646
<i>Randall Balestriero, Richard Baraniuk</i>	
<b>Approximation Guarantees for Adaptive Sampling</b> .....	661
<i>Eric Balkanski, Yaron Singer</i>	
<b>Improving the Gaussian Mechanism for Differential Privacy: Analytical Calibration and Optimal Denoising</b> .....	678
<i>Borja Balle, Yu-Xiang Wang</i>	
<b>Dissecting Adam: The Sign, Magnitude and Variance of Stochastic Gradients</b> .....	693
<i>Lukas Balles, Philipp Hennig</i>	
<b>Differentially Private Database Release via Kernel Mean Embeddings</b> .....	710
<i>Matej Balog, Ilya Tolstikhin, Bernhard Schölkopf</i>	
<b>Improving Optimization in Models With Continuous Symmetry Breaking</b> .....	727
<i>Robert Bamler, Stephan Mandt</i>	
<b>Improved Training of Generative Adversarial Networks using Representative Features</b> .....	737
<i>Duhyeon Bang, Hyunjung Shim</i>	

## VOLUME 2

<b>Using Inherent Structures to Design Lean 2-layer RBMs</b> .....	747
<i>Abhishek Bansal, Abhinav Anand, Chiranjib Bhattacharyya</i>	
<b>Classification from Pairwise Similarity and Unlabeled Data</b> .....	763
<i>Han Bao, Gang Niu, Masashi Sugiyama</i>	
<b>Bayesian Optimization of Combinatorial Structures</b> .....	782
<i>Ricardo Baptista, Matthias Poloczek</i>	
<b>Geodesic Convolutional Shape Optimization</b> .....	797
<i>Pierre Baque, Edoardo Remelli, Francois Fleuret, Pascal Fua</i>	
<b>Learning to Coordinate with Coordination Graphs in Repeated Single-Stage Multi-Agent Decision Problems</b> .....	810
<i>Eugenio Bargiacchi, Timothy Verstraeten, Diederik Roijers, Ann Nowé, Hado Hasselt</i>	
<b>Testing Sparsity over Known and Unknown Bases</b> .....	819
<i>Siddharth Barman, Arnab Bhattacharyya, Suprovat Ghoshal</i>	
<b>Transfer in Deep Reinforcement Learning Using Successor Features and Generalised Policy Improvement</b> .....	844
<i>Andre Barreto, Diana Borsa, John Quan, Tom Schaul, David Silver, Matteo Hessel, Daniel Mankowitz, Augustin Zidek, Remi Munos</i>	
<b>Gradient Descent with Identity Initialization Efficiently Learns Positive Definite Linear Transformations By Deep Residual Networks</b> .....	854
<i>Peter Bartlett, Dave Helmbold, Phil Long</i>	
<b>Mutual Information Neural Estimation</b> .....	864
<i>Mohamed Ishmael Belghazi, Aristide Baratin, Sai Rajeshwar, Sherjil Ozair, Yoshua Bengio, Devon Hjelm, Aaron Courville</i>	
<b>To Understand Deep Learning We Need to Understand Kernel Learning</b> .....	874
<i>Mikhail Belkin, Siyuan Ma, Soumik Mandal</i>	
<b>Understanding and Simplifying One-Shot Architecture Search</b> .....	883
<i>Gabriel Bender, Pieter-Jan Kindermans, Barret Zoph, Vijay Vasudevan, Quoc Le</i>	
<b>SIGNSGD: Compressed Optimisation for Non-Convex Problems</b> .....	894
<i>Jeremy Bernstein, Yu-Xiang Wang, Kamyar Azizzadenesheli, Animashree Anandkumar</i>	
<b>Distributed Clustering via LSH Based Data Partitioning</b> .....	919
<i>Aditya Bhaskara, Maheshakya Wijewardena</i>	
<b>Autoregressive Convolutional Neural Networks for Asynchronous Time Series</b> .....	933
<i>Mikolaj Binkowski, Gautier Marti, Philippe Donnat</i>	
<b>Adaptive Sampled Softmax with Kernel Based Sampling</b> .....	946
<i>Guy Blanc, Steffen Rendle</i>	
<b>Optimizing the Latent Space of Generative Networks</b> .....	960
<i>Piotr Bojanowski, Armand Joulin, David Lopez-Pas, Arthur Szlam</i>	
<b>NetGAN: Generating Graphs via Random Walks</b> .....	973
<i>Aleksandar Bojchevski, Oleksandr Shchur, Daniel Zügner, Stephan Günnemann</i>	

<b>A Progressive Batching L-BFGS Method for Machine Learning</b> .....	989
<i>Raghu Bollapragada, Dheevatsa Mudigere, Jorge Nocedal, Hao-Jun Michael Shi, Ping Tak Peter Tang</i>	
<b>Prediction Rule Reshaping</b> .....	1014
<i>Matt Bonakdarpour, Sabyasachi Chatterjee, Rina Foygel Barber, John Lafferty</i>	
<b>QuantTree: Histograms for Change Detection in Multivariate Data Streams</b> .....	1023
<i>Giacomo Boracchi, Diego Carrera, Cristiano Cervellera, Danilo Macciò</i>	
<b>Matrix Norms in Data Streams: Faster, Multi-Pass and Row-Order</b> .....	1036
<i>Vladimir Braverman, Stephen Chestnut, Robert Krauthgamer, Yi Li, David Woodruff, Lin Yang</i>	
<b>Predict and Constrain: Modeling Cardinality in Deep Structured Prediction</b> .....	1046
<i>Nataly Brukhim, Amir Globerson</i>	
<b>Quasi-Monte Carlo Variational Inference</b> .....	1055
<i>Alexander Buchholz, Florian Wenzel, Stephan Mandt</i>	
<b>Path-Level Network Transformation for Efficient Architecture Search</b> .....	1069
<i>Han Cai, Jiacheng Yang, Weinan Zhang, Song Han, Yong Yu</i>	
<b>Improved Large-Scale Graph Learning through Ridge Spectral Sparsification</b> .....	1081
<i>Daniele Calandriello, Ioannis Koutis, Alessandro Lazaric, Michal Valko</i>	
<b>Bayesian Coreset Construction via Greedy Iterative Geodesic Ascent</b> .....	1091
<i>Trevor Campbell, Tamara Broderick</i>	
<b>Adversarial Learning with Local Coordinate Coding</b> .....	1104
<i>Jiezhong Cao, Yong Guo, Qingyao Wu, Chunhua Shen, Junzhou Huang, Mingkui Tan</i>	
<b>Fair and Diverse DPP-Based Data Summarization</b> .....	1118
<i>Elisa Celis, Vijay Keswani, Damian Straszak, Amit Deshpande, Tarun Kathuria, Nisheeth Vishnoi</i>	
<b>Conditional Noise-Contrastive Estimation of Unnormalised Models</b> .....	1134
<i>Ciwan Ceylan, Michael U. Gutmann</i>	
<b>Adversarial Time-to-Event Modeling</b> .....	1143
<i>Paidamoyo Chapfuwa, Chenyang Tao, Chunyuan Li, Courtney Page, Benjamin Goldstein, Lawrence Carin, Ricardo Henao</i>	
<b>Stability and Generalization of Learning Algorithms that Converge to Global Optima</b> .....	1157
<i>Zachary Charles, Dimitris Papailiopoulos</i>	
<b>Learning and Memorization</b> .....	1174
<i>Satrajit Chatterjee</i>	
<b>On the Theory of Variance Reduction for Stochastic Gradient Monte Carlo</b> .....	1183
<i>Niladri Chatterji, Nicolas Flammarion, Yian Ma, Peter Bartlett, Michael Jordan</i>	
<b>Hierarchical Clustering with Structural Constraints</b> .....	1213
<i>Vaggos Chatziafratis, Rad Niazadeh, Moses Charikar</i>	
<b>Hierarchical Deep Generative Models for Multi-Rate Multivariate Time Series</b> .....	1227
<i>Zhengping Che, Sanjay Purushotham, Guangyu Li, Bo Jiang, Yan Liu</i>	
<b>GradNorm: Gradient Normalization for Adaptive Loss Balancing in Deep Multitask Networks</b> .....	1240
<i>Zhao Chen, Vijay Badrinarayanan, Chen-Yu Lee, Andrew Rabinovich</i>	
<b>Weakly Submodular Maximization Beyond Cardinality Constraints: Does Randomization Help Greedy?</b> .....	1252
<i>Lin Chen, Moran Feldman, Amin Karbasi</i>	
<b>Projection-Free Online Optimization with Stochastic Gradient: From Convexity to Submodularity</b> .....	1266
<i>Lin Chen, Christopher Harshaw, Hamed Hassani, Amin Karbasi</i>	
<b>Continuous-Time Flows for Efficient Inference and Density Estimation</b> .....	1285
<i>Changyou Chen, Chunyuan Li, Liqun Chen, Wenlin Wang, Yunchen Pu, Lawrence Carin</i>	
<b>Scalable Bilinear <math>\pi</math> Learning Using State and Action Features</b> .....	1305
<i>Yichen Chen, Lihong Li, Mengdi Wang</i>	
<b>Stein Points</b> .....	1320
<i>Wilson Chen, Lester Mackey, Jackson Gorham, Francois-Xavier Briol, Chris Oates</i>	
<b>Learning K-way D-dimensional Discrete Codes for Compact Embedding Representations</b> .....	1351
<i>Ting Chen, Martin Renqiang Min, Yizhou Sun</i>	
<b>PixelSNAIL: An Improved Autoregressive Generative Model</b> .....	1364
<i>Xi Chen, Nikhil Mishra, Mostafa Rohaninejad, Pieter Abbeel</i>	
<b>Dynamical Isometry and a Mean Field Theory of RNNs: Gating Enables Signal Propagation in Recurrent Neural Networks</b> .....	1373
<i>Minmin Chen, Jeffrey Pennington, Samuel Schoenholz</i>	
<b>Learning to Explain: An Information-Theoretic Perspective on Model Interpretation</b> .....	1386
<i>Jianbo Chen, Le Song, Martin Wainwright, Michael Jordan</i>	
<b>Variational Inference and Model Selection with Generalized Evidence Bounds</b> .....	1419
<i>Liqun Chen, Chenyang Tao, Ruiyi Zhang, Ricardo Henao, Lawrence Carin</i>	
<b>DRACO: Byzantine-resilient Distributed Training via Redundant Gradients</b> .....	1436
<i>Lingjiao Chen, Hongyi Wang, Zachary Charles, Dimitris Papailiopoulos</i>	
<b>SADAGRAD: Strongly Adaptive Stochastic Gradient Methods</b> .....	1455
<i>Zaiyi Chen, Yi Xu, Enhong Chen, Tianbao Yang</i>	
<b>Covariate Adjusted Precision Matrix Estimation via Nonconvex Optimization</b> .....	1464
<i>Jinghui Chen, Pan Xu, Lingxiao Wang, Jian Ma, Quanquan Gu</i>	

## VOLUME 3

<b>End-to-End Learning for the Deep Multivariate Probit Model</b> .....	1490
<i>Di Chen, Yexiang Xue, Carla Gomes</i>	
<b>Stochastic Training of Graph Convolutional Networks with Variance Reduction</b> .....	1503
<i>Jianfei Chen, Jun Zhu, Le Song</i>	
<b>Extreme Learning to Rank via Low Rank Assumption</b> .....	1533
<i>Minhao Cheng, Ian Davidson, Cho-Jui Hsieh</i>	
<b>Learning a Mixture of Two Multinomial Logits</b> .....	1545
<i>Flavio Chierichetti, Ravi Kumar, Andrew Tomkins</i>	
<b>Structured Evolution with Compact Architectures for Scalable Policy Optimization</b> .....	1562
<i>Krzysztof Choromanski, Mark Rowland, Vikas Sindhwani, Richard E. Turner, Adrian Weller</i>	
<b>Path Consistency Learning in Tsallis Entropy Regularized MDPs</b> .....	1578
<i>Yinlam Chow, Ofir Nachum, Mohammad Ghavamzadeh</i>	
<b>An Iterative, Sketching-based Framework for Ridge Regression</b> .....	1595
<i>Agniva Chowdhury, Jiasen Yang, Petros Drineas</i>	
<b>Stochastic Wasserstein Barycenters</b> .....	1627
<i>Sebastian Claiici, Edward Chien, Justin Solomon</i>	
<b>Self-Consistent Trajectory Autoencoder: Hierarchical Reinforcement Learning with Trajectory Embeddings</b> .....	1637
<i>John Co-Reyes, Yuxuan Liu, Abhishek Gupta, Benjamin Eysenbach, Pieter Abbeel, Sergey Levine</i>	
<b>On Acceleration with Noise-Corrupted Gradients</b> .....	1648
<i>Michael Cohen, Jelena Diakonikolas, Lorenzo Orecchia</i>	
<b>Online Linear Quadratic Control</b> .....	1667
<i>Alon Cohen, Avinatán Hasidim, Tomer Koren, Nevena Lazic, Yishay Mansour, Kunal Talwar</i>	
<b>GEP-PG: Decoupling Exploration and Exploitation in Deep Reinforcement Learning Algorithms</b> .....	1682
<i>Cédric Colas, Olivier Sigaud, Pierre-Yves Oudayer</i>	
<b>Leveraging Well-Conditioned Bases: Streaming and Distributed Summaries in Minkowski p-Norms</b> .....	1692
<i>Graham Cormode, Charlie Dickens, David Woodruff</i>	
<b>Efficient ModelBased Deep Reinforcement Learning with Variational State Tabulation</b> .....	1708
<i>Dane Corneil, Wulfram Gerstner, Johanni Brea</i>	
<b>Online Learning with Abstention</b> .....	1726
<i>Corinna Cortes, Giulia Desalvo, Claudio Gentile, Mehryar Mohri, Scott Yang</i>	
<b>Constrained Interacting Submodular Groupings</b> .....	1735
<i>Andrew Cotter, Mahdi Milani Fard, Seungil You, Maya Gupta, Jeff Bilmes</i>	
<b>Inference Suboptimality in Variational Autoencoders</b> .....	1749
<i>Chris Cremer, Xuechen Li, David Duvenaud</i>	
<b>Mix &amp; Match Agent Curricula for Reinforcement Learning</b> .....	1761
<i>Wojciech Marian Czarnecki, Siddhant M. Jayakumar, Max Jaderberg, Leonard Hasenclever, Yee Whye Teh, Nicolas Heess, Simon Osindero, Razvan Pascanu</i>	
<b>Implicit Quantile Networks for Distributional Reinforcement Learning</b> .....	1774
<i>Will Dabney, Georg Ostrovski, David Silver, Remi Munos</i>	
<b>Learning Steady-States of Iterative Algorithms over Graphs</b> .....	1788
<i>Hanjun Dai, Zornitsa Kozareva, Bo Dai, Alex Smola, Le Song</i>	
<b>Adversarial Attack on Graph Structured Data</b> .....	1799
<i>Hanjun Dai, Hui Li, Tian Tian, Xin Huang, Lin Wang, Jun Zhu, Le Song</i>	
<b>SBED: Convergent Reinforcement Learning with Nonlinear Function Approximation</b> .....	1809
<i>Bo Dai, Albert Shaw, Lihong Li, Lin Xiao, Niao He, Zhen Liu, Jianshu Chen, Le Song</i>	
<b>Compressing Neural Networks using the Variational Information Bottleneck</b> .....	1819
<i>Bin Dai, Chen Zhu, Baining Guo, David Wipf</i>	
<b>Asynchronous Byzantine Machine Learning (the case of SGD)</b> .....	1829
<i>Georgios Damaskinos, El Mahdi El Mhamdi, Rachid Guerraoui, Rhicheck Patra, Mahsa Taziki</i>	
<b>Escaping Saddles with Stochastic Gradients</b> .....	1859
<i>Hadi Daneshmand, Jonas Kohler, Aurelien Lucchi, Thomas Hofmann</i>	
<b>Minibatch Gibbs Sampling on Large Graphical Models</b> .....	1885
<i>Christopher De Sa, Vincent Chen, Wing Wong</i>	
<b>Stochastic Video Generation with a Learned Prior</b> .....	1906
<i>Emily Denton, Rob Fergus</i>	
<b>Decomposition of Uncertainty in Bayesian Deep Learning for Efficient and Risk-sensitive Learning</b> .....	1920
<i>Stefan Depeweg, Jose-Miguel Hernandez-Lobato, Finale Doshi-Velez, Steffen Udluft</i>	
<b>Accurate Inference for Adaptive Linear Models</b> .....	1935
<i>Yash Deshpande, Lester Mackey, Vasilis Syrgkanis, Matt Taddy</i>	
<b>Variational Network Inference: Strong and Stable with Concrete Support</b> .....	1945
<i>Amir Dezfouli, Edwin V. Bonilla, Richard Nock</i>	
<b>Modeling Sparse Deviations for Compressed Sensing using Generative Models</b> .....	1990
<i>Manik Dhar, Aditya Grover, Stefano Ermon</i>	
<b>Alternating Randomized Block Coordinate Descent</b> .....	2006
<i>Jelena Diakonikolas, Lorenzo Orecchia</i>	
<b>Learning to Act in Decentralized Partially Observable MDPs</b> .....	2019
<i>Jilles Dibangoye, Olivier Buffet</i>	

<b>Noisin: Unbiased Regularization for Recurrent Neural Networks</b> .....	2030
<i>Adji Bousso Dieng, Rajesh Ranganath, Jaan Allosaer, David Blei</i>	
<b>Discovering and Removing Exogenous State Variables and Rewards for Reinforcement Learning</b> .....	2040
<i>Thomas Dietterich, George Trimpontias, Zhitang Chen</i>	
<b>Coordinated Exploration in Concurrent Reinforcement Learning</b> .....	2051
<i>Maria Dimakopoulou, Benjamin Van Roy</i>	
<b>Probabilistic Recurrent State-Space Models</b> .....	2060
<i>Andreas Doerr, Christian Daniel, Martin Schiegg, Duy Nguyen-Tuong, Stefan Schaal, Marc Toussaint, Sebastian Trimpe</i>	
<b>Randomized Block Cubic Newton Method</b> .....	2076
<i>Nikita Doikov, Peter Richtarik</i>	
<b>Low-Rank Riemannian Optimization on Positive Semidefinite Stochastic Matrices with Applications to Graph Clustering</b> .....	2091
<i>Ahmed Douik, Babak Hassibi</i>	
<b>Essentially No Barriers in Neural Network Energy Landscape</b> .....	2101
<i>Felix Draxler, Kambis Veschgini, Manfred Salmhofer, Fred Hamprecht</i>	
<b>Weakly Consistent Optimal Pricing Algorithms in Repeated Posted-Price Auctions with Strategic Buyer</b> .....	2113
<i>Alexey Drutsa</i>	
<b>On the Power of Over-parametrization in Neural Networks with Quadratic Activation</b> .....	2132
<i>Simon S. Du, Jason D. Lee</i>	
<b>Gradient Descent Learns One-hidden-layer CNN: Don't be Afraid of Spurious Local Minima</b> .....	2142
<i>Simon S. Du, Jason D. Lee, Yuandong Tian, Aarti Singh, Barnabas Poczos</i>	
<b>Investigating Human Priors for Playing Video Games</b> .....	2160
<i>Rachit Dubey, Pulkit Agrawal, Deepak Pathak, Tom Griffiths, Alexei Efros</i>	
<b>A Distributed Second-Order Algorithm You Can Trust</b> .....	2169
<i>Celestine Duenner, Aurelien Lucchi, Matilde Gargiani, An Bian, Thomas Hofmann, Martin Jaggi</i>	
<b>Computational Optimal Transport: Complexity by Accelerated Gradient Descent Is Better Than by Sinkhorn's Algorithm</b> .....	2196
<i>Pavel Dvurechensky, Alexander Gasnikov, Alexey Kroshnin</i>	
<b>Entropy-SGD Optimizes the Prior of a PAC-Bayes Bound: Generalization Properties of Entropy-SGD and Data-Dependent Priors</b> .....	2221
<i>Gintare Karolina Dziugaite, Daniel Roy</i>	

#### VOLUME 4

<b>Beyond the One-Step Greedy Approach in Reinforcement Learning</b> .....	2239
<i>Yonathan Efroni, Gal Dalal, Bruno Scherrer, Shie Mannor</i>	
<b>Parallel and Streaming Algorithms for K-Core Decomposition</b> .....	2253
<i>Hossein Esfandiari, Silvio Lattanzi, Vahab Mirrokni</i>	
<b>IMPALA: Scalable Distributed Deep-RL with Importance Weighted Actor-Learner Architectures</b> .....	2263
<i>Lasse Espeholt, Hubert Soyer, Remi Munos, Karen Simonyan, Volodymyr Mnih, Tom Ward, Yotam Doron, Vlad Firoiu, Tim Harley, Iain Dunning, Shane Legg, Koray Kavukcuoglu</i>	
<b>Scalable Gaussian Processes with Grid-Structured Eigenfunctions (GP-GRIEF)</b> .....	2285
<i>Trefor Evans, Prasanth Nair</i>	
<b>The Limits of Maxing, Ranking, and Preference Learning</b> .....	2296
<i>Moein Falahatgar, Ayush Jain, Alon Orlitsky, Venkatesh Pichapati, Vaishakh Ravindrakumar</i>	
<b>BOHB: Robust and Efficient Hyperparameter Optimization at Scale</b> .....	2323
<i>Stefan Falkner, Aaron Klein, Frank Hutter</i>	
<b>More Robust Doubly Robust Off-policy Evaluation</b> .....	2342
<i>Mehrdad Farajtabar, Yinlam Chow, Mohammad Ghavamzadeh</i>	
<b>Efficient and Consistent Adversarial Bipartite Matching</b> .....	2372
<i>Rizal Fathony, Sima Behpour, Xinhua Zhang, Brian Ziebart</i>	
<b>Global Convergence of Policy Gradient Methods for the Linear Quadratic Regulator</b> .....	2385
<i>Maryam Fazel, Rong Ge, Sham Kakade, Mehran Mesbahi</i>	
<b>CRVI: Convex Relaxation for Variational Inference</b> .....	2414
<i>Ghazal Fazelnia, John Paisley</i>	
<b>Fourier Policy Gradients</b> .....	2423
<i>Matthew Fellows, Kamil Ciosek, Shimon Whiteson</i>	
<b>Nonparametric Variable Importance Using an Augmented Neural Network with Multi-task Learning</b> .....	2436
<i>Jean Feng, Brian D. Williamson, Marco Carone, Noah Simon</i>	
<b>Closed-form Marginal Likelihood in Gamma-Poisson Matrix Factorization</b> .....	2449
<i>Louis Filstroff, Alberto Lumbreras, Cédric Fèveotte</i>	
<b>Automatic Goal Generation for Reinforcement Learning Agents</b> .....	2458
<i>Carlos Florensa, David Held, Xinyang Geng, Pieter Abbeel</i>	
<b>DiCE: The Infinitely Differentiable Monte Carlo Estimator</b> .....	2472
<i>Jakob Foerster, Gregory Farquhar, Maruan Al-Shedivat, Tim Rocktäschel, Eric Xing, Shimon Whiteson</i>	
<b>Practical Contextual Bandits with Regression Oracles</b> .....	2482
<i>Dylan Foster, Alekh Agarwal, Miroslav Dudik, Haipeng Luo, Robert Schapire</i>	
<b>Generative Temporal Models with Spatial Memory for Partially Observed Environments</b> .....	2518
<i>Marco Fraccaro, Danilo Rezende, Yori Zwols, Alexander Pritzel, S. M. Ali Eslami, Fabio Viola</i>	

<b>ADMM and Accelerated ADMM as Continuous Dynamical Systems</b> .....	2528
<i>Guilherme Franca, Daniel Robinson, Rene Vidal</i>	
<b>Bilevel Programming for Hyperparameter Optimization and Meta-Learning</b> .....	2537
<i>Luca Franceschi, Paolo Frasconi, Saverio Salzo, Riccardo Grazi, Massimiliano Pontil</i>	
<b>Efficient Bias-Span-Constrained Exploration-Exploitation in Reinforcement Learning</b> .....	2549
<i>Ronan Fruit, Matteo Pirotta, Alessandro Lazaric, Ronald Ortner</i>	
<b>Addressing Function Approximation Error in Actor-Critic Methods</b> .....	2587
<i>Scott Fujimoto, Herke Hoof, David Meger</i>	
<b>Clipped Action Policy Gradient</b> .....	2602
<i>Yasuhiro Fujita, Shin-Ichi Maeda</i>	
<b>Born-Again Neural Networks</b> .....	2615
<i>Tommaso Furlanello, Zachary Lipton, Michael Tschannen, Laurent Itti, Anima Anandkumar</i>	
<b>Local Private Hypothesis Testing: Chi-Square Tests</b> .....	2625
<i>Marco Gaboardi, Ryan Rogers</i>	
<b>Inductive Two-layer Modeling with Parametric Bregman Transfer</b> .....	2646
<i>Vignesh Ganapathiraman, Zhan Shi, Xinhua Zhang, Yaoliang Yu</i>	
<b>Hyperbolic Entailment Cones for Learning Hierarchical Embeddings</b> .....	2661
<i>Octavian-Eugen Ganeva, Gary Becigneul, Thomas Hofmann</i>	
<b>Parameterized Algorithms for the Matrix Completion Problem</b> .....	2674
<i>Robert Ganian, Iyad Kanj, Sebastian Ordyniak, Stefan Szeider</i>	
<b>Synthesizing Programs for Images using Reinforced Adversarial Learning</b> .....	2684
<i>Yaroslav Ganin, Tejas Kulkarni, Igor Babuschkin, S. M. Ali Eslami, Oriol Vinyals</i>	
<b>Spotlight: Optimizing Device Placement for Training Deep Neural Networks</b> .....	2696
<i>Yuanxiang Gao, Li Chen, Baochun Li</i>	
<b>Parallel Bayesian Network Structure Learning</b> .....	2705
<i>Tian Gao, Dennis Wei</i>	
<b>Structured Output Learning with Abstention: Application to Accurate Opinion Prediction</b> .....	2716
<i>Alexandre Garcia, Chloé Clavel, Slim Essid, Florence D'Alche-Buc</i>	
<b>Conditional Neural Processes</b> .....	2738
<i>Marta Garnelo, Dan Rosenbaum, Christopher Maddison, Tiago Ramalho, David Saxton, Murray Shanahan, Yee Whye Teh, Danilo Rezende, S. M. Ali Eslami</i>	
<b>Temporal Poisson Square Root Graphical Models</b> .....	2748
<i>Sinong Geng, Zhaobin Kuang, Peggy Peissig, David Page</i>	
<b>The Generalization Error of Dictionary Learning with Moreau Envelopes</b> .....	2764
<i>Alexandros Georgogiannis</i>	
<b>Budgeted Experiment Design for Causal Structure Learning</b> .....	2788
<i>Ahmed Ghassami, Saber Salehkaleybar, Negar Kiyavash, Elias Bareinboim</i>	
<b>Linear Spectral Estimators and an Application to Phase Retrieval</b> .....	2802
<i>Ramina Ghods, Andrew Lan, Tom Goldstein, Christoph Studer</i>	
<b>Structured Variational Learning of Bayesian Neural Networks with Horseshoe Priors</b> .....	2819
<i>Soumya Ghosh, Jiayu Yao, Finale Doshi-Velez</i>	
<b>Learning Maximum-A-Posteriori Perturbation Models for Structured Prediction in Polynomial Time</b> .....	2833
<i>Asish Ghoshal, Jean Honorio</i>	
<b>Robust and Scalable Models of Microbiome Dynamics</b> .....	2844
<i>Travis Gibson, Georg Gerber</i>	
<b>Non-Linear Motor Control by Local Learning in Spiking Neural Networks</b> .....	2855
<i>Aditya Gilra, Wulfram Gerstner</i>	
<b>Learning One Convolutional Layer with Overlapping Patches</b> .....	2865
<i>Surbhi Goel, Adam Klivans, Raghu Meka</i>	
<b>Visualizing and Understanding Atari Agents</b> .....	2877
<i>Samuel Greydanus, Anurag Koul, Jonathan Dodge, Alan Fern</i>	
<b>Learning Policy Representations in Multiagent Systems</b> .....	2887
<i>Aditya Grover, Maruan Al-Shedivat, Jayesh Gupta, Yuri Burda, Harrison Edwards</i>	
<b>Faster Derivative-Free Stochastic Algorithm for Shared Memory Machines</b> .....	2898
<i>Bin Gu, Zhouyuan Huo, Cheng Deng, Heng Huang</i>	
<b>Learning to Search with MCTSnets</b> .....	2920
<i>Arthur Guez, Theophane Weber, Ioannis Antonoglou, Karen Simonyan, Oriol Vinyals, Daan Wierstra, Remi Munos, David Silver</i>	
<b>Characterizing Implicit Bias in Terms of Optimization Geometry</b> .....	2932
<i>Suriya Gunasekar, Jason Lee, Daniel Soudry, Nathan Srebro</i>	
<b>Shampoo: Preconditioned Stochastic Tensor Optimization</b> .....	2956
<i>Vineet Gupta, Tomer Koren, Yoram Singer</i>	
<b>Latent Space Policies for Hierarchical Reinforcement Learning</b> .....	2965
<i>Tuomas Haarnoja, Kristian Hartikainen, Pieter Abbeel, Sergey Levine</i>	

## VOLUME 5

<b>Soft Actor-Critic: Off-Policy Maximum Entropy Deep Reinforcement Learning with a Stochastic Actor</b> .....	2976
<i>Tuomas Haarnoja, Aurick Zhou, Pieter Abbeel, Sergey Levine</i>	



<b>Comparison-Based Random Forests</b> .....	2990
<i>Siavash Haghighi, Damien Garreau, Ulrike Luxburg</i>	
<b>K-Beam Minimax: Efficient Optimization for Deep Adversarial Learning</b> .....	3000
<i>Jihun Hamm, Yung-Kyun Noh</i>	
<b>Candidates vs. Noises Estimation for Large Multi-Class Classification Problem</b> .....	3013
<i>Lei Han, Yiheng Huang, Tong Zhang</i>	
<b>Stein Variational Gradient Descent Without Gradient</b> .....	3030
<i>Jun Han, Qiang Liu</i>	
<b>Rectify Heterogeneous Models with Semantic Mapping</b> .....	3040
<i>Han-Jia Ye, De-Chuan Zhan, Yuan Jiang, Zhi-Hua Zhou</i>	
<b>Deep Models of Interactions Across Sets</b> .....	3050
<i>Jason Hartford, Devon Graham, Kevin Leyton-Brown, Siamak Ravanbakhsh</i>	
<b>Learning Memory Access Patterns</b> .....	3062
<i>Milad Hashemi, Kevin Swersky, Jamie Smith, Grant Ayers, Heiner Litz, Jichuan Chang, Christos Kozyrakis, Parthasarathy Ranganathan</i>	
<b>Fairness Without Demographics in Repeated Loss Minimization</b> .....	3077
<i>Tatsunori Hashimoto, Megha Srivastava, Hongseok Namkoong, Percy Liang</i>	
<b>Multicalibration: Calibration for the (Computationally-Identifiable) Masses</b> .....	3087
<i>Ursula Hebert-Johnson, Michael Kim, Omer Reingold, Guy Rothblum</i>	
<b>Recurrent Predictive State Policy Networks</b> .....	3104
<i>Ahmed Hefny, Zita Marinho, Wen Sun, Siddhartha Srinivasa, Geoffrey Gordon</i>	
<b>Learning Unknown ODE Models with Gaussian Processes</b> .....	3120
<i>Markus Heinonen, Cagatay Yildiz, Henrik Mannerström, Jukka Intosalmi, Harri Lähdesmäki</i>	
<b>Orthogonal Recurrent Neural Networks with Scaled Cayley Transform</b> .....	3133
<i>Kyle Helfrich, Devin Willmott, Qiang Ye</i>	
<b>Fast Bellman Updates for Robust MDPs</b> .....	3144
<i>Chin Pang Ho, Marek Petrik, Wolfram Wiesemann</i>	
<b>CyCADA: Cycle-Consistent Adversarial Domain Adaptation</b> .....	3162
<i>Judy Hoffman, Eric Tzeng, Taesung Park, Jun-Yan Zhu, Phillip Isola, Kate Saenko, Alexei Efros, Trevor Darrell</i>	
<b>Sound Abstraction and Decomposition of Probabilistic Programs</b> .....	3175
<i>Steven Holtzen, Guy Broeck, Todd Millstein</i>	
<b>Gradient Primal-Dual Algorithm Converges to Second-Order Stationary Solution for Nonconvex Distributed Optimization Over Networks</b> .....	3189
<i>Mingyi Hong, Meisam Razaviyayn, Jason Lee</i>	
<b>Variational Bayesian Dropout: Pitfalls and Fixes</b> .....	3199
<i>Jiri Hron, Alexander G. G. Matthews, Zoubin Ghahramani</i>	
<b>Does Distributionally Robust Supervised Learning Give Robust Classifiers?</b> .....	3220
<i>Weihua Hu, Gang Niu, Issei Sato, Masashi Sugiyama</i>	
<b>Dissipativity Theory for Accelerating Stochastic Variance Reduction: A Unified Analysis of SVRG and Katyusha Using Semidefinite Programs</b> .....	3250
<i>Bin Hu, Stephen Wright, Laurent Lessard</i>	
<b>Near Optimal Frequent Directions for Sketching Dense and Sparse Matrices</b> .....	3260
<i>Zengfeng Huang</i>	
<b>Learning Deep ResNet Blocks Sequentially using Boosting Theory</b> .....	3272
<i>Furong Huang, Jordan Ash, John Langford, Robert Schapire</i>	
<b>Learning Hidden Markov Models from Pairwise Co-occurrences with Application to Topic Modeling</b> .....	3291
<i>Kejun Huang, Xiao Fu, Nicholas Sidiropoulos</i>	
<b>Neural Autoregressive Flows</b> .....	3309
<i>Chin-Wei Huang, David Krueger, Alexandre Lacoste, Aaron Courville</i>	
<b>Topological Mixture Estimation</b> .....	3325
<i>Steve Huntsman</i>	
<b>Decoupled Parallel Backpropagation with Convergence Guarantee</b> .....	3335
<i>Zhouyuan Huo, Bin Gu, Qian Yang, Heng Huang</i>	
<b>Using Reward Machines for High-Level Task Specification and Decomposition in Reinforcement Learning</b> .....	3347
<i>Rodrigo Toro Icarte, Toryn Klassen, Richard Valenzano, Sheila McIlraith</i>	
<b>Deep Variational Reinforcement Learning for POMDPs</b> .....	3359
<i>Maximilian Igl, Luisa Zintgraf, Tuan Anh Le, Frank Wood, Shimon Whiteson</i>	
<b>Attention-based Deep Multiple Instance Learning</b> .....	3376
<i>Maximilian Ilse, Jakub Tomczak, Max Welling</i>	
<b>Black-box Adversarial Attacks with Limited Queries and Information</b> .....	3392
<i>Andrew Ilyas, Logan Engstrom, Anish Athalye, Jessy Lin</i>	
<b>Analysis of Minimax Error Rate for Crowdsourcing and Its Application to Worker Clustering Model</b> .....	3402
<i>Hideaki Imamura, Issei Sato, Masashi Sugiyama</i>	
<b>Improving Regression Performance with Distributional Losses</b> .....	3414
<i>Ehsan Imani, Martha White</i>	
<b>Deep Density Destructors</b> .....	3426
<i>David Inouye, Pradeep Ravikumar</i>	
<b>Unbiased Objective Estimation in Predictive Optimization</b> .....	3438
<i>Shinji Ito, Akihiro Yabe, Ryohei Fujimaki</i>	

<b>Anonymous Walk Embeddings</b> .....	3448
<i>Sergey Ivanov, Evgeny Burnaev</i>	
<b>Learning Binary Latent Variable Models: A Tensor Eigenpair Approach</b> .....	3458
<i>Ariel Jaffe, Roi Weiss, Shai Carmi, Yuval Kluger, Boaz Nadler</i>	
<b>Firing Bandits: Optimizing Crowdfunding</b> .....	3473
<i>Lalit Jain, Kevin Jamieson</i>	
<b>Differentially Private Matrix Completion Revisited</b> .....	3486
<i>Prateek Jain, Om Dipakbhai Thakkar, Abhradeep Thakurta</i>	
<b>Video Prediction with Appearance and Motion Conditions</b> .....	3496
<i>Yunseok Jang, Gunhee Kim, Yale Song</i>	
<b>Pathwise Derivatives Beyond the Reparameterization Trick</b> .....	3511
<i>Martin Jankowiak, Fritz Obermeyer</i>	
<b>Detecting Non-causal Artifacts in Multivariate Linear Regression Models</b> .....	3529
<i>Dominik Janzing, Bernhard Schölkopf</i>	
<b>A Unified Framework for Structured Low-rank Matrix Learning</b> .....	3540
<i>Pratik Javanpuria, Bamdev Mishra</i>	
<b>Efficient End-to-End Learning for Quantizable Representations</b> .....	3550
<i>Yeonwoo Jeong, Hyun Oh Song</i>	
<b>Exploring Hidden Dimensions in Parallelizing Convolutional Neural Networks</b> .....	3560
<i>Zhihao Jia, Sina Lin, Charles R. Qi, Alex Aiken</i>	
<b>Feedback-Based Tree Search for Reinforcement Learning</b> .....	3572
<i>Daniel Jiang, Emmanuel Ekwedike, Han Liu</i>	
<b>Quickshift++: Provably Good Initializations for Sample-Based Mean Shift</b> .....	3591
<i>Heinrich Jiang, Jennifer Jang, Samory Kpotufe</i>	
<b>MentorNet: Learning Data-Driven Curriculum for Very Deep Neural Networks on Corrupted Labels</b> .....	3601
<i>Lu Jiang, Zhengyuan Zhou, Thomas Leung, Li-Jia Li, Li Fei-Fei</i>	
<b>The Weighted Kendall and High-order Kernels for Permutations</b> .....	3621
<i>Yunlong Jiao, Jean-Philippe Vert</i>	
<b>Junction Tree Variational Autoencoder for Molecular Graph Generation</b> .....	3632
<i>Wengong Jin, Regina Barzilay, Tommi Jaakkola</i>	
<b>Network Global Testing by Counting Graphlets</b> .....	3649
<i>Jiashun Jin, Zheng Tracy Ke, Shengming Luo</i>	
<b>Regret Minimization for Partially Observable Deep Reinforcement Learning</b> .....	3669
<i>Peter Jin, Kurt Keutzer, Sergey Levine</i>	
<b>WSNet: Compact and Efficient Networks Through Weight Sampling</b> .....	3683
<i>Xiaojie Jin, Yingzhen Yang, Ning Xu, Jianchao Yang, Nebojsa Jojic, Jiashi Feng, Shuicheng Yan</i>	
<b>Large-Scale Cox Process Inference using Variational Fourier Features</b> .....	3697
<i>ST John, James Hensman</i>	
<b>Composite Functional Gradient Learning of Generative Adversarial Models</b> .....	3706
<i>Rie Johnson, Tong Zhang</i>	

## VOLUME 6

<b>Kronecker Recurrent Units</b> .....	3725
<i>Cijo Jose, Moustapha Cisse, Francois Fleuret</i>	
<b>Fast Decoding in Sequence Models Using Discrete Latent Variables</b> .....	3743
<i>Lukasz Kaiser, Samy Bengio, Aurko Roy, Ashish Vaswani, Niki Parmar, Jakob Uszkoreit, Noam Shazeer</i>	
<b>Kernel Recursive ABC: Point Estimation with Intractable Likelihood</b> .....	3753
<i>Takafumi Kajihara, Motonobu Kanagawa, Keisuke Yamazaki, Kenji Fukumizu</i>	
<b>Efficient Neural Audio Synthesis</b> .....	3775
<i>Nal Kalchbrenner, Erich Elsen, Karen Simonyan, Seb Noury, Norman Casagrande, Edward Lockhart, Florian Stimberg, Aaron Oord, Sander Dieleman, Koray Kavukcuoglu</i>	
<b>Learning Diffusion using Hyperparameters</b> .....	3785
<i>Dimitris Kalimeris, Yaron Singer, Karthik Subbian, Udi Weinsberg</i>	
<b>Signal and Noise Statistics Oblivious Orthogonal Matching Pursuit</b> .....	3800
<i>Sreejith Kallummil, Sheetal Kalyani</i>	
<b>Residual Unfairness in Fair Machine Learning from Prejudiced Data</b> .....	3821
<i>Nathan Kallus, Angela Zhou</i>	
<b>Learn from Your Neighbor: Learning Multi-modal Mappings from Sparse Annotations</b> .....	3835
<i>Ashwin Kalyan, Stefan Lee, Anitha Kannan, Dhruv Batra</i>	
<b>Semi-Supervised Learning via Compact Latent Space Clustering</b> .....	3845
<i>Konstantinos Kamnitsas, Daniel Castro, Loic Le Folgoc, Ian Walker, Ryutaro Tanno, Daniel Rueckert, Ben Glocker, Antonio Criminisi, Aditya Nori</i>	
<b>Policy Optimization with Demonstrations</b> .....	3855
<i>Bingyi Kang, Zequn Jie, Jiashi Feng</i>	
<b>Improving Sign Random Projections With Additional Information</b> .....	3870
<i>Keegan Kang, Wong Wei Pin</i>	
<b>Let's be Honest: An Optimal No-Regret Framework for Zero-Sum Games</b> .....	3879
<i>Ehsan Asadi Kangarshahi, Ya-Ping Hsieh, Mehmet Fatih Sahin, Volkan Cevher</i>	

<b>Continual Reinforcement Learning with Complex Synapses</b> .....	3893
<i>Christos Kaplanis, Murray Shanahan, Claudia Clopath</i>	
<b>LaVAN: Localized and Visible Adversarial Noise</b> .....	3903
<i>Danny Karmon, Daniel Zoran, Yoav Goldberg</i>	
<b>Riemannian Stochastic Recursive Gradient Algorithm with Retraction and Vector Transport and Its Convergence Analysis</b> .....	3912
<i>Hiroyuki Kasai, Hiroyuki Sato, Bamdev Mishra</i>	
<b>Not All Samples Are Created Equal: Deep Learning with Importance Sampling</b> .....	3936
<i>Angelos Katharopoulos, Francois Fleuret</i>	
<b>Feasible Arm Identification</b> .....	3950
<i>Julian Katz-Samuels, Clay Scott</i>	
<b>Scalable Deletion-Robust Submodular Maximization: Data Summarization with Privacy and Fairness Constraints</b> .....	3980
<i>Ehsan Kazemi, Morteza Zadimoghaddam, Amin Karbasi</i>	
<b>Focused Hierarchical RNNs for Conditional Sequence Processing</b> .....	3998
<i>Nan Rosemary Ke, Konrad Zolna, Alessandro Sordani, Zhouhan Lin, Adam Trischler, Yoshua Bengio, Joelle Pineau, Laurent Charlin, Christopher Pal</i>	
<b>Preventing Fairness Gerrymandering: Auditing and Learning for Subgroup Fairness</b> .....	4008
<i>Michael Kearns, Seth Neel, Aaron Roth, Zhiwei Steven Wu</i>	
<b>Improved Nearest Neighbor Search Using Auxiliary Information and Priority Functions</b> .....	4017
<i>Omid Keivani, Kaushik Sinha</i>	
<b>ContextNet: Deep learning for Star Galaxy Classification</b> .....	4030
<i>Noble Kennamer, David Kirkby, Alexander Ihler, Francisco Javier Sanchez-Lopez</i>	
<b>Frank-Wolfe with Subsampling Oracle</b> .....	4039
<i>Thomas Kerdreux, Fabian Pedregosa, Alexandre D'Aspremont</i>	
<b>Convergence Guarantees for a Class of Non-convex and Non-smooth Optimization Problems</b> .....	4062
<i>Koulik Khamaru, Martin Wainwright</i>	
<b>Fast and Scalable Bayesian Deep Learning by Weight-Perturbation in Adam</b> .....	4088
<i>Mohammad Emtiyaz Khan, Didrik Nielsen, Voot Tangkaratt, Wu Lin, Yarin Gal, Akash Srivastava</i>	
<b>Geometry Score: A Method For Comparing Generative Adversarial Networks</b> .....	4114
<i>Valentin Khrulkov, Ivan Oseledets</i>	
<b>Blind Justice: Fairness with Encrypted Sensitive Attributes</b> .....	4123
<i>Niki Kilbertus, Adria Gascon, Matt Kusner, Michael Veale, Krishna P. Gummadi, Adrian Weller</i>	
<b>Markov Modulated Gaussian Cox Processes for Semi-Stationary Intensity Modeling of Events Data</b> .....	4138
<i>Minyoung Kim</i>	
<b>Disentangling by Factorising</b> .....	4153
<i>Hyunjik Kim, Andriy Mnih</i>	
<b>Self-Bounded Prediction Suffix Tree via Approximate String Matching</b> .....	4172
<i>Dongwoo Kim, Christian Walder</i>	
<b>Interpretability Beyond Feature Attribution: Quantitative Testing with Concept Activation Vectors (TCAV)</b> .....	4186
<i>Been Kim, Martin Wattenberg, Justin Gilmer, Carrie Cai, James Wexler, Fernanda Viegas, Rory Sayres</i>	
<b>Semi-Amortized Variational Autoencoders</b> .....	4196
<i>Yoon Kim, Sam Wiseman, Andrew Miller, David Sontag, Alexander Rush</i>	
<b>Neural Relational Inference for Interacting Systems</b> .....	4209
<i>Thomas Kipf, Ethan Fetaya, Kuan-Chieh Wang, Max Welling, Richard Zemel</i>	
<b>An Alternative View: When Does SGD Escape Local Minima?</b> .....	4226
<i>Robert Kleinberg, Yuanzhi Li, Yang Yuan</i>	
<b>Crowdsourcing with Arbitrary Adversaries</b> .....	4238
<i>Matthaeus Kleindessner, Pranjal Awasthi</i>	
<b>Spatio-temporal Bayesian On-line Changepoint Detection with Model Selection</b> .....	4260
<i>Jeremias Knoblauch, Theodoros Damoulas</i>	
<b>Fast Gradient-Based Methods with Exponential Rate: A Hybrid Control Framework</b> .....	4271
<i>Arman Sharifi Kolarijani, Peyman Mohajerin Esfahani, Tamas Keviczky</i>	
<b>Nonconvex Optimization for Regression with Fairness Constraints</b> .....	4280
<i>Junpei Komiyama, Akiko Takeda, Junya Honda, Hajime Shimao</i>	
<b>On the Generalization of Equivariance and Convolution in Neural Networks to the Action of Compact Groups</b> .....	4295
<i>Risi Kondor, Shubhendu Trivedi</i>	
<b>Compiling Combinatorial Prediction Games</b> .....	4310
<i>Frederic Koriche</i>	
<b>Dynamic Evaluation of Neural Sequence Models</b> .....	4320
<i>Ben Krause, Emmanuel Kahembwe, Iain Murray, Steve Renals</i>	
<b>Semiparametric Contextual Bandits</b> .....	4330
<i>Akshay Krishnamurthy, Zhiwei Steven Wu, Vasilis Syrgkanis</i>	
<b>Fast Maximization of Non-Submodular, Monotonic Functions on the Integer Lattice</b> .....	4350
<i>Alan Kuhnle, J. David Smith, Victoria G. Crawford, My T. Thai</i>	
<b>Accurate Uncertainties for Deep Learning Using Calibrated Regression</b> .....	4369
<i>Volodymyr Kuleshov, Nathan Fenner, Stefano Ermon</i>	
<b>Trainable Calibration Measures For Neural Networks From Kernel Mean Embeddings</b> .....	4378
<i>Aviral Kumar, Sunita Sarawagi, Ujjwal Jain</i>	
<b>Data-Dependent Stability of Stochastic Gradient Descent</b> .....	4390
<i>Ilya Kuzborskij, Christoph Lampert</i>	

<b>Explicit Inductive Bias for Transfer Learning with Convolutional Networks</b> .....	4408
<i>Xuhong Li, Yves Grandvalet, Franck Davoine</i>	
<b>Understanding the Loss Surface of Neural Networks for Binary Classification</b> .....	4420
<i>Shiyu Liang, Ruoyu Sun, Yixuan Li, Rayadurgam Srikant</i>	
<b>Mixed Batches and Symmetric Discriminators for GAN Training</b> .....	4430
<i>Thomas Lucas, Corentin Tallec, Yann Ollivier, Jakob Verbeek</i>	
<b>Binary Partitions with Approximate Minimum Impurity</b> .....	4448
<i>Eduardo S. Laber, Marco Molinaro, Felipe A. Mello Pereira</i>	

## VOLUME 7

<b>Canonical Tensor Decomposition for Knowledge Base Completion</b> .....	4475
<i>Timothee Lacroix, Nicolas Usunier, Guillaume Obozinski</i>	
<b>Generalization without Systematicity: On the Compositional Skills of Sequence-to-Sequence Recurrent Networks</b> .....	4487
<i>Brenden Lake, Marco Baroni</i>	
<b>An Estimation and Analysis Framework for the Rasch Model</b> .....	4500
<i>Andrew Lan, Mung Chiang, Christoph Studer</i>	
<b>Partial Optimality and Fast Lower Bounds for Weighted Correlation Clustering</b> .....	4509
<i>Jan-Hendrik Lange, Andreas Karrenbauer, Bjoern Andres</i>	
<b>Deep Linear Networks with Arbitrary Loss: All Local Minima Are Global</b> .....	4522
<i>Thomas Laurent, James Brecht</i>	
<b>The Multilinear Structure of ReLU Networks</b> .....	4528
<i>Thomas Laurent, James Brecht</i>	
<b>Hierarchical Imitation and Reinforcement Learning</b> .....	4560
<i>Hoang Le, Nan Jiang, Alekh Agarwal, Miroslav Dudik, Yisong Yue, Hal Daumé</i>	
<b>Gradient-Based Meta-Learning with Learned Layerwise Metric and Subspace</b> .....	4574
<i>Yoonho Lee, Seungjin Choi</i>	
<b>Deep Reinforcement Learning in Continuous Action Spaces: a Case Study in the Game of Simulated Curling</b> .....	4587
<i>Kywoon Lee, Sol-A Kim, Jaesik Choi, Seong-Whan Lee</i>	
<b>Gated Path Planning Networks</b> .....	4597
<i>Lisa Lee, Emilio Parisotto, Devendra Singh Chaplot, Eric Xing, Ruslan Salakhutdinov</i>	
<b>Deep Asymmetric Multi-task Feature Learning</b> .....	4609
<i>Hae Beom Lee, Eunho Yang, Sung Ju Hwang</i>	
<b>Noise2Noise: Learning Image Restoration without Clean Data</b> .....	4620
<i>Jaakko Lehtinen, Jacob Munkberg, Jon Hasselgren, Samuli Laine, Tero Karras, Miika Aittala, Timo Aila</i>	
<b>Out-of-sample Extension of Graph Adjacency Spectral Embedding</b> .....	4632
<i>Keith Levin, Farbod Roosta-Khorasani, Michael W. Mahoney, Carey E. Priebe</i>	
<b>An Optimal Control Approach to Deep Learning and Applications to Discrete-Weight Neural Networks</b> .....	4642
<i>Qianxiao Li, Shuji Hao</i>	
<b>Towards Binary-Valued Gates for Robust LSTM Training</b> .....	4662
<i>Zhuohan Li, Di He, Fei Tian, Wei Chen, Tao Qin, Liwei Wang, Tie-Yan Liu</i>	
<b>On the Limitations of First-Order Approximation in GAN Dynamics</b> .....	4672
<i>Jerry Li, Aleksander Madry, John Peebles, Ludwig Schmidt</i>	
<b>Submodular Hypergraphs: p-Laplacians, Cheeger Inequalities and Spectral Clustering</b> .....	4690
<i>Pan Li, Olgica Milenkovic</i>	
<b>The Well-Tempered Lasso</b> .....	4720
<i>Yuanzhi Li, Yoram Singer</i>	
<b>Estimation of Markov Chain via Rank-constrained Likelihood</b> .....	4729
<i>Xudong Li, Mengdi Wang, Anru Zhang</i>	
<b>Asynchronous Decentralized Parallel Stochastic Gradient Descent</b> .....	4745
<i>Xiangru Lian, Wei Zhang, Ce Zhang, Ji Liu</i>	
<b>RLlib: Abstractions for Distributed Reinforcement Learning</b> .....	4768
<i>Eric Liang, Richard Liaw, Robert Nishihara, Philipp Moritz, Roy Fox, Ken Goldberg, Joseph E. Gonzalez, Michael I. Jordan, Ion Stoica</i>	
<b>On the Spectrum of Random Features Maps of High Dimensional Data</b> .....	4781
<i>Zhenyu Liao, Romain Couillet</i>	
<b>The Dynamics of Learning: A Random Matrix Approach</b> .....	4793
<i>Zhenyu Liao, Romain Couillet</i>	
<b>Revisiting and Improving Recurrent Back-Propagation</b> .....	4807
<i>Renjie Liao, Yuwen Xiong, Ethan Fetaya, Lisa Zhang, Kijung Yoon, Xaq Pitkow, Raquel Urtasun, Richard Zemel</i>	
<b>Optimal Distributed Learning with Multi-pass Stochastic Gradient Methods</b> .....	4821
<i>Junhong Lin, Volkan Cevher</i>	
<b>Optimal Rates of Sketched-regularized Algorithms for Least-Squares Regression over Hilbert Spaces</b> .....	4848
<i>Junhong Lin, Volkan Cevher</i>	
<b>Level-Set Methods for Finite-Sum Constrained Convex Optimization</b> .....	4868
<i>Qihang Lin, Runchao Ma, Tianbao Yang</i>	
<b>Detecting and Correcting for Label Shift with Black Box Predictors</b> .....	4887
<i>Zachary Lipton, Yu-Xiang Wang, Alexander Smola</i>	

<b>Generalized Robust Bayesian Committee Machine for Large-scale Gaussian Process Regression</b> .....	4898
<i>Haitao Liu, Jianfei Cai, Yi Wang, Yew Soon Ong</i>	
<b>Towards Black-box Iterative Machine Teaching</b> .....	4911
<i>Weiyang Liu, Bo Dai, Xingguo Li, Zhen Liu, James Rehg, Le Song</i>	
<b>Delayed Impact of Fair Machine Learning</b> .....	4929
<i>Lydia Liu, Sarah Dean, Esther Rolf, Max Simchowitz, Moritz Hardt</i>	
<b>A Two-Step Computation of the Exact GAN Wasserstein Distance</b> .....	4959
<i>Huidong Liu, Xianfeng Gu, Dimitris Samaras</i>	
<b>Open Category Detection with PAC Guarantees</b> .....	4985
<i>Si Liu, Risheek Garrepalli, Thomas Dietterich, Alan Fern, Dan Hendrycks</i>	
<b>Fast Variance Reduction Method with Stochastic Batch Size</b> .....	5017
<i>Xuanqing Liu, Cho-Jui Hsieh</i>	
<b>Fast Stochastic AUC Maximization with <math>O(1/n)</math>-Convergence Rate</b> .....	5034
<i>Mingrui Liu, Xiaoxuan Zhang, Zaiyi Chen, Xiaoyu Wang, Tianbao Yang</i>	
<b>On Matching Pursuit and Coordinate Descent</b> .....	5048
<i>Francesco Locatello, Anant Raj, Sai Praneeth Karimireddy, Gunnar Raetsch, Bernhard Schölkopf, Sebastian Stich, Martin Jaggi</i>	
<b>PDE-Net: Learning PDEs from Data</b> .....	5067
<i>Zichao Long, Yiping Lu, Xianzhong Ma, Bin Dong</i>	
<b>Error Estimation for Randomized Least-Squares Algorithms via the Bootstrap</b> .....	5079
<i>Miles Lopes, Shusen Wang, Michael Mahoney</i>	
<b>Constraining the Dynamics of Deep Probabilistic Models</b> .....	5089
<i>Marco Lorenzi, Maurizio Filippone</i>	
<b>Spectrally Approximating Large Graphs with Smaller Graphs</b> .....	5102
<i>Andreas Loukas, Pierre Vandergheynst</i>	
<b>The Edge Density Barrier: Computational-Statistical Tradeoffs in Combinatorial Inference</b> .....	5119
<i>Hao Lu, Yuan Cao, Junwei Lu, Han Liu, Zhaoran Wang</i>	
<b>Accelerating Greedy Coordinate Descent Methods</b> .....	5149
<i>Haihao Lu, Robert Freund, Vahab Mirrokni</i>	
<b>Structured Variationally Auto-encoded Optimization</b> .....	5169
<i>Xiaoyu Lu, Javier Gonzalez, Zhenwen Dai, Neil D. Lawrence</i>	
<b>Beyond Finite Layer Neural Networks: Bridging Deep Architectures and Numerical Differential Equations</b> .....	5181
<i>Yiping Lu, Aoxiao Zhong, Quanzheng Li, Bin Dong</i>	
<b>End-to-end Active Object Tracking via Reinforcement Learning</b> .....	5191
<i>Wenhan Luo, Peng Sun, Fangwei Zhong, Wei Liu, Tong Zhang, Yizhou Wang</i>	
<b>Competitive Caching with Machine Learned Advice</b> .....	5201
<i>Thodoris Lykouris, Sergei Vassilytiskii</i>	

## VOLUME 8

<b>Batch Bayesian Optimization via Multi-objective Acquisition Ensemble for Automated Analog Circuit Design</b> .....	5214
<i>Wenlong Lyu, Fan Yang, Changhao Yan, Dian Zhou, Xuan Zeng</i>	
<b>Celer: A Fast Solver for the Lasso with Dual Extrapolation</b> .....	5226
<i>Mathurin Massias, Joseph Salmon, Alexandre Gramfort</i>	
<b>The Power of Interpolation: Understanding the Effectiveness of SGD in Modern Over-parametrized Learning</b> .....	5238
<i>Siyuan Ma, Raef Bassily, Mikhail Belkin</i>	
<b>Gradient Descent for Sparse Rank-One Matrix Completion for Crowd-Sourced Aggregation of Sparsely Interacting Workers</b> .....	5248
<i>Yao Ma, Alexander Olshevsky, Csaba Szepesvari, Venkatesh Saligrama</i>	
<b>Implicit Regularization in Nonconvex Statistical Estimation: Gradient Descent Converges Linearly for Phase Retrieval and Matrix Completion</b> .....	5264
<i>Cong Ma, Kaizheng Wang, Yuejie Chi, Yuxin Chen</i>	
<b>Dimensionality-Driven Learning with Noisy Labels</b> .....	5332
<i>Xingjun Ma, Yisen Wang, Michael E. Houle, Shuo Zhou, Sarah M. Erfani, Shu-Tao Xia, Sudanthi Wijewickrema, James Bailey</i>	
<b>Approximate Message Passing for Amplitude Based Optimization</b> .....	5342
<i>Junjie Ma, Ji Xu, Arian Maleki</i>	
<b>Learning Adversarially Fair and Transferable Representations</b> .....	5423
<i>David Madras, Elliot Creager, Toniann Pitassi, Richard Zemel</i>	
<b>An Efficient, Generalized Bellman Update For Cooperative Inverse Reinforcement Learning</b> .....	5435
<i>Dhruv Malik, Malayandi Palaniappan, Jaime Fisac, Dylan Hadfield-Menell, Stuart Russell, Anca Dragan</i>	
<b>Iterative Amortized Inference</b> .....	5444
<i>Joseph Marino, Yisong Yue, Stephan Mandt</i>	
<b>Streaming Principal Component Analysis in Noisy Settings</b> .....	5463
<i>Teodor Vanislavov Marinov, Poorya Mianjy, Raman Arora</i>	
<b>Fast Approximate Spectral Clustering for Dynamic Networks</b> .....	5484
<i>Lionel Martin, Andreas Loukas, Pierre Vandergheynst</i>	
<b>Bayesian Model Selection for Change Point Detection and Clustering</b> .....	5497
<i>Othmane Mazhar, Cristian Rojas, Carlo Fischione, Mohammad Reza Hesamzadeh</i>	
<b>Optimization, Fast and Slow: Optimally Switching between Local and Bayesian Optimization</b> .....	5521
<i>Mark McLeod, Stephen Roberts, Michael A. Osborne</i>	

<b>Bounds on the Approximation Power of Feedforward Neural Networks</b> .....	5531
<i>Mohammad Mehrabi, Aslan Tchamkerten, Mansoor Yousefi</i>	
<b>Differentiable Dynamic Programming for Structured Prediction and Attention</b> .....	5540
<i>Arthur Mensch, Mathieu Blondel</i>	
<b>Ranking Distributions based on Noisy Sorting</b> .....	5563
<i>Adil El Mesaoudi-Paul, Eyke Hüllermeier, Robert Busa-Fekete</i>	
<b>Which Training Methods for GANs Do Actually Converge?</b> .....	5589
<i>Lars Mescheder, Andreas Geiger, Sebastian Nowozin</i>	
<b>Configurable Markov Decision Processes</b> .....	5627
<i>Alberto Maria Metelli, Mirco Mutti, Marcello Restelli</i>	
<b>prDeep: Robust Phase Retrieval with a Flexible Deep Network</b> .....	5654
<i>Christopher Metzler, Phillip Schniter, Ashok Veeraraghavan, Richard G. Baraniuk</i>	
<b>Pseudo-task Augmentation: From Deep Multitask Learning to Intratask Sharing-and Back</b> .....	5664
<i>Elliot Meyerson, Risto Miikkulainen</i>	
<b>The Hidden Vulnerability of Distributed Learning in Byzantium</b> .....	5674
<i>El Mahdi El Mhamdi, Rachid Guerraoui, Sébastien Rouault</i>	
<b>Stochastic PCA with <math>\ell_2</math> and <math>\ell_1</math> Regularization</b> .....	5687
<i>Poorya Mianjy, Raman Arora</i>	
<b>On the Implicit Bias of Dropout</b> .....	5701
<i>Poorya Mianjy, Raman Arora, Rene Vidal</i>	
<b>One-Shot Segmentation in Clutter</b> .....	5718
<i>Claudio Michaelis, Matthias Bethge, Alexander Ecker</i>	
<b>Differentiable Plasticity: Training Plastic Neural Networks with Backpropagation</b> .....	5728
<i>Thomas Miconi, Kenneth Stanley, Jeff Clune</i>	
<b>Training Neural Machines with Trace-Based Supervision</b> .....	5740
<i>Mathew Mirman, Dimitar Dimitrov, Pavle Djordjevic, Timon Gehr, Martin Vechev</i>	
<b>Differentiable Abstract Interpretation for Provably Robust Neural Networks</b> .....	5761
<i>Mathew Mirman, Timon Gehr, Martin Vechev</i>	
<b>A Delay-tolerant Proximal-Gradient Algorithm for Distributed Learning</b> .....	5774
<i>Konstantin Mishchenko, Franck Iutzeler, Jérôme Malick, Massih-Reza Amini</i>	
<b>Data Summarization at Scale: A Two-Stage Submodular Approach</b> .....	5789
<i>Marko Mitrovic, Ehsan Kazemi, Morteza Zadimoghaddam, Amin Karbasi</i>	
<b>The Hierarchical Adaptive Forgetting Variational Filter</b> .....	5806
<i>Vincent Moens</i>	
<b>Decentralized Submodular Maximization: Bridging Discrete and Continuous Settings</b> .....	5822
<i>Aryan Mokhtari, Hamed Hassani, Amin Karbasi</i>	
<b>DICOD: Distributed Convolutional Coordinate Descent for Convolutional Sparse Coding</b> .....	5846
<i>Thomas Moreau, Laurent Oudre, Nicolas Vayatis</i>	
<b>WHInter: A Working Set Algorithm for High-dimensional Sparse Second Order Interaction Models</b> .....	5861
<i>Marine Le Morvan, Jean-Philippe Vert</i>	
<b>Dropout Training, Data-dependent Regularization, and Generalization Bounds</b> .....	5879
<i>Wenlong Mou, Yuchen Zhou, Jun Gao, Liwei Wang</i>	
<b>Kernelized Synaptic Weight Matrices</b> .....	5888
<i>Lorenz Müller, Julien Martel, Giacomo Indiveri</i>	
<b>Rapid Adaptation with Conditionally Shifted Neurons</b> .....	5898
<i>Tsendsuren Munkhdalai, Xingdi Yuan, Soroush Mehri, Adam Trischler</i>	
<b>On the Relationship Between Data Efficiency and Error for Uncertainty Sampling</b> .....	5910
<i>Stephen Mussmann, Percy Liang</i>	
<b>Fitting New Speakers Based on a Short Untranscribed Sample</b> .....	5932
<i>Eliya Nachmani, Adam Polyak, Yaniv Taigman, Lior Wolf</i>	
<b>Smoothed Action Value Functions for Learning Gaussian Policies</b> .....	5941
<i>Ofir Nachum, Mohammad Norouzi, George Tucker, Dale Schuurmans</i>	

## VOLUME 9

<b>Nearly Optimal Robust Subspace Tracking</b> .....	5954
<i>Praneeth Narayanamurthy, Namrata Vaswani</i>	
<b>Stochastic Proximal Algorithms for AUC Maximization</b> .....	5963
<i>Michael Natole, Yiming Ying, Siwei Lyu</i>	
<b>Mitigating Bias in Adaptive Data Gathering via Differential Privacy</b> .....	5973
<i>Seth Neel, Aaron Roth</i>	
<b>Optimization Landscape and Expressivity of Deep CNNs</b> .....	5983
<i>Quynh Nguyen, Matthias Hein</i>	
<b>Neural Networks Should Be Wide Enough to Learn Disconnected Decision Regions</b> .....	6001
<i>Quynh Nguyen, Mahesh Chandra Mukkamala, Matthias Hein</i>	
<b>SGD and Hogwild! Convergence Without the Bounded Gradients Assumption</b> .....	6012
<i>Lam Nguyen, Phuong Ha Nguyen, Marten Dijk, Peter Richtarik, Katya Scheinberg, Martin Takac</i>	
<b>Active Testing: An Efficient and Robust Framework for Estimating Accuracy</b> .....	6036
<i>Phuc Nguyen, Deva Ramanan, Charless Fowlkes</i>	

<b>On Learning Sparsely Used Dictionaries from Incomplete Samples</b> .....	6050
<i>Thanh Nguyen, Akshay Soni, Chinmay Hegde</i>	
<b>Learning Continuous Hierarchies in the Lorentz Model of Hyperbolic Geometry</b> .....	6071
<i>Maximillian Nickel, Douwe Kiela</i>	
<b>State Space Gaussian Processes with Non-Gaussian Likelihood</b> .....	6081
<i>Hannes Nickisch, Arno Solin, Alexander Grigorevskiy</i>	
<b>SparseMAP: Differentiable Sparse Structured Inference</b> .....	6091
<i>Vlad Niculae, Andre Martins, Mathieu Blondel, Claire Cardie</i>	
<b>A Theoretical Explanation for Perplexing Behaviors of Backpropagation-based Visualizations</b> .....	6105
<i>Weili Nie, Yang Zhang, Ankit Patel</i>	
<b>Functional Gradient Boosting based on Residual Network Perception</b> .....	6126
<i>Atsushi Nitanda, Taiji Suzuki</i>	
<b>Beyond 1/2-Approximation for Submodular Maximization on Massive Data Streams</b> .....	6144
<i>Ashkan Norouzi-Fard, Jakub Tarnawski, Slobodan Mitrovic, Amir Zandieh, Aidasadat Mousavifar, Ola Svensson</i>	
<b>The Uncertainty Bellman Equation and Exploration</b> .....	6154
<i>Brendan O'Donoghue, Ian Osband, Remi Munos, Volodymyr Mnih</i>	
<b>Is Generator Conditioning Causally Related to GAN Performance?</b> .....	6174
<i>Augustus Odena, Jacob Buckman, Catherine Olsson, Tom Brown, Christopher Olah, Colin Raffel, Ian Goodfellow</i>	
<b>Learning in Reproducing Kernel Krein Spaces</b> .....	6188
<i>Dino Oglic, Thomas Gaertner</i>	
<b>BOCK : Bayesian Optimization with Cylindrical Kernels</b> .....	6201
<i>Changyong Oh, Efstratios Gavves, Max Welling</i>	
<b>Self-Imitation Learning</b> .....	6214
<i>Junhyuk Oh, Yijie Guo, Satinder Singh, Honglak Lee</i>	
<b>A Probabilistic Framework for Multi-view Feature Learning with Many-to-many Associations Via Neural Networks</b> .....	6227
<i>Akifumi Okuno, Tetsuya Hada, Hidetoshi Shimodaira</i>	
<b>Transformation Autoregressive Networks</b> .....	6243
<i>Junier Oliva, Avinava Dubey, Manzil Zaheer, Barnabas Poczos, Ruslan Salakhutdinov, Eric Xing, Jeff Schneider</i>	
<b>Design of Experiments for Model Discrimination Hybridising Analytical and Data-Driven Approaches</b> .....	6259
<i>Simon Olofsson, Marc Deisenroth, Ruth Misener</i>	
<b>Parallel WaveNet: Fast High-Fidelity Speech Synthesis</b> .....	6270
<i>Aaron Oord, Yazhe Li, Igor Babuschkin, Karen Simonyan, Oriol Vinyals, Koray Kavukcuoglu, George Driessche, Edward Lockhart, Luis Cobo, Florian Stimberg, Norman Casagrande, Dominik Grewe, Seb Noury, Sander Dieleman, Erich Elsen, Nal Kalchbrenner, Heiga Zen, Alex Graves, Helen King, Tom Walters, Dan Belov, Demis Hassabis</i>	
<b>Learning Localized Spatio-Temporal Models From Streaming Data</b> .....	6279
<i>Muhammad Osama, Dave Zachariah, Thomas Schön</i>	
<b>Autoregressive Quantile Networks for Generative Modeling</b> .....	6293
<i>Georg Ostrovski, Will Dabney, Remi Munos</i>	
<b>Efficient First-Order Algorithms for Adaptive Signal Denoising</b> .....	6309
<i>Dmitrii Ostrovskii, Zaid Harchaoui</i>	
<b>Analyzing Uncertainty in Neural Machine Translation</b> .....	6329
<i>Myle Ott, Michael Auli, David Grangier, Marc'Aurelio Ranzato</i>	
<b>Learning Compact Neural Networks with Regularization</b> .....	6341
<i>Samet Oymak</i>	
<b>Tree Edit Distance Learning via Adaptive Symbol Embeddings</b> .....	6351
<i>Benjamin Paaßen, Claudio Gallicchio, Alessio Micheli, Barbara Hammer</i>	
<b>Reinforcement Learning with Function-Valued Action Spaces for Partial Differential Equation Control</b> .....	6361
<i>Yangchen Pan, Amir-Massoud Farahmand, Martha White, Saleh Nabi, Piyush Grover, Daniel Nikovski</i>	
<b>Learning to Speed Up Structured Output Prediction</b> .....	6376
<i>Xingyuan Pan, Vivek Srikumar</i>	
<b>Theoretical Analysis of Image-to-Image Translation with Adversarial Learning</b> .....	6389
<i>Xudong Pan, Mi Zhang, Daizong Ding</i>	
<b>Max-Mahalanobis Linear Discriminant Analysis Networks</b> .....	6409
<i>Tianyu Pang, Chao Du, Jun Zhu</i>	
<b>Stochastic Variance-Reduced Policy Gradient</b> .....	6422
<i>Matteo Papini, Damiano Binaghi, Giuseppe Canonaco, Matteo Piroitta, Marcello Restelli</i>	
<b>Learning Independent Causal Mechanisms</b> .....	6432
<i>Giambattista Parascandolo, Niki Kilbertus, Mateo Rojas-Carulla, Bernhard Schölkopf</i>	
<b>Time Limits in Reinforcement Learning</b> .....	6443
<i>Fabio Pardo, Arash Tavakoli, Vitaly Levdivk, Petar Kormushev</i>	
<b>Image Transformer</b> .....	6453
<i>Niki Parmar, Ashish Vaswani, Jakob Uszkoreit, Lukasz Kaiser, Noam Shazeer, Alexander Ku, Dustin Tran</i>	
<b>PIPPS: Flexible Model-Based Policy Search Robust to the Curse of Chaos</b> .....	6463
<i>Paavo Parmas, Carl Edward Rasmussen, Jan Peters, Kenji Doya</i>	
<b>High-Quality Prediction Intervals for Deep Learning: A Distribution-Free, Ensembled Approach</b> .....	6473
<i>Tim Pearce, Alexandra Brintrup, Mohamed Zaki, Andy Neely</i>	
<b>Adaptive Three Operator Splitting</b> .....	6486
<i>Fabian Pedregosa, Gauthier Gidel</i>	

<b>Efficient Neural Architecture Search via Parameter Sharing</b> .....	6522
<i>Hieu Pham, Melody Guan, Barret Zoph, Quoc Le, Jeff Dean</i>	
<b>Bandits with Delayed, Aggregated Anonymous Feedback</b> .....	6533
<i>Ciara Pike-Burke, Shipra Agrawal, Csaba Szepesvari, Steffen Grunewalder</i>	
<b>Constant-Time Predictive Distributions for Gaussian Processes</b> .....	6575
<i>Geoff Pleiss, Jacob Gardner, Kilian Weinberger, Andrew Gordon Wilson</i>	
<b>Local Convergence Properties of SAGA/Prox-SVRG and Acceleration</b> .....	6585
<i>Clarice Poon, Jingwei Liang, Carola-Bibiane Schoenlieb</i>	
<b>Equivalence of Multicategory SVM and Simplex Cone SVM: Fast Computations and Statistical Theory</b> .....	6604
<i>Guillaume Pouliot</i>	
<b>Learning Dynamics of Linear Denoising Autoencoders</b> .....	6612
<i>Arnu Pretorius, Steve Kroon, Herman Kamper</i>	
<b>JointGAN: Multi-Domain Joint Distribution Learning with Generative Adversarial Nets</b> .....	6626
<i>Yunchen Pu, Shuyang Dai, Zhe Gan, Weiyao Wang, Guoyin Wang, Yizhe Zhang, Ricardo Henao, Lawrence Carin</i>	
<b>Selecting Representative Examples for Program Synthesis</b> .....	6644
<i>Yewen Pu, Zachery Miranda, Armando Solar-Lezama, Leslie Kaelbling</i>	
<b>Generalized Earley Parser: Bridging Symbolic Grammars and Sequence Data for Future Prediction</b> .....	6656
<i>Siyuan Qi, Baoxiang Jia, Song-Chun Zhu</i>	
<b>Do Outliers Ruin Collaboration?</b> .....	6668
<i>Mingda Qiao</i>	
<b>Gradually Updated Neural Networks for Large-Scale Image Recognition</b> .....	6676
<i>Siyuan Qiao, Zhishuai Zhang, Wei Shen, Bo Wang, Alan Yuille</i>	
<b>DCFNet: Deep Neural Network with Decomposed Convolutional Filters</b> .....	6687
<i>Qiang Qiu, Xiuyuan Cheng, Robert Calderbank, Guillermo Sapiro</i>	

## VOLUME 10

<b>Non-convex Conditional Gradient Sliding</b> .....	6701
<i>Chao Qu, Yan Li, Huan Xu</i>	
<b>Machine Theory of Mind</b> .....	6723
<i>Neil Rabinowitz, Frank Perbet, Francis Song, Chiyuan Zhang, S. M. Ali Eslami, Matthew Botvinick</i>	
<b>Fast Parametric Learning with Activation Memorization</b> .....	6739
<i>Jack Rae, Chris Dyer, Peter Dayan, Timothy Lillicrap</i>	
<b>Can Deep Reinforcement Learning Solve Erdos-Selfridge-Spencer Games?</b> .....	6756
<i>Maithra Raghu, Alex Irpan, Jacob Andreas, Robert Kleinberg, Quoc Le, Jon Kleinberg</i>	
<b>Cut-Pursuit Algorithm for Regularizing Nonsmooth Functionals with Graph Total Variation</b> .....	6769
<i>Hugo Raguet, Loic Landrieu</i>	
<b>Modeling Others using Oneself in Multi-Agent Reinforcement Learning</b> .....	6779
<i>Roberta Raileanu, Emily Denton, Arthur Szlam, Rob Fergus</i>	
<b>On Nesting Monte Carlo Estimators</b> .....	6789
<i>Tom Rainforth, Robert Cornish, Hongseok Yang, Andrew Warrington</i>	
<b>Tighter Variational Bounds are Not Necessarily Better</b> .....	6818
<i>Tom Rainforth, Adam Kosioerek, Tuan Anh Le, Chris Maddison, Maximilian Igl, Frank Wood, Yee Whye Teh</i>	
<b>SAFFRON: an Adaptive Algorithm for Online Control of the False Discovery Rate</b> .....	6833
<i>Aaditya Ramdas, Tijana Zrnic, Martin Wainwright, Michael Jordan</i>	
<b>QMIX: Monotonic Value Function Factorisation for Deep Multi-Agent Reinforcement Learning</b> .....	6846
<i>Tabish Rashid, Mikayel Samvelyan, Christian Schroeder Witt, Gregory Farquhar, Jakob Foerster, Shimon Whiteson</i>	
<b>Gradient Coding from Cyclic MDS Codes and Expander Graphs</b> .....	6860
<i>Netanel Raviv, Rashish Tandon, Alex Dimakis, Itzhak Tamo</i>	
<b>Learning Implicit Generative Models with the Method of Learned Moments</b> .....	6869
<i>Suman Ravuri, Shakir Mohamed, Mihaela Rosca, Oriol Vinyals</i>	
<b>Weightless: Lossy Weight Encoding for Deep Neural Network Compression</b> .....	6886
<i>Brandon Reagan, Udit Gupta, Bob Adolf, Michael Mitzenmacher, Alexander Rush, Gu-Yeon Wei, David Brooks</i>	
<b>Learning to Reweight Examples for Robust Deep Learning</b> .....	6900
<i>Mengye Ren, Wenyuan Zeng, Bin Yang, Raquel Urtasun</i>	
<b>Learning by Playing Solving Sparse Reward Tasks from Scratch</b> .....	6910
<i>Martin Riedmiller, Roland Hafner, Thomas Lampe, Michael Neunert, Jonas Degraeve, Tom Wiele, Vlad Mnih, Nicolas Heess, Jost Tobias Springenberg</i>	
<b>Been There, Done That: Meta-Learning with Episodic Recall</b> .....	6929
<i>Samuel Ritter, Jane Wang, Zeb Kurth-Nelson, Siddhant Jayakumar, Charles Blundell, Razvan Pascanu, Matthew Botvinick</i>	
<b>A Hierarchical Latent Vector Model for Learning Long-Term Structure in Music</b> .....	6939
<i>Adam Roberts, Jesse Engel, Colin Raffel, Curtis Hawthorne, Douglas Eck</i>	
<b>Learning to Optimize Combinatorial Functions</b> .....	6955
<i>Nir Rosenfeld, Eric Balkanski, Amir Globerson, Yaron Singer</i>	
<b>Fast Information-theoretic Bayesian Optimisation</b> .....	6967
<i>Binxin Ru, Mark McLeod, Diego Granziol, Michael A. Osborne</i>	
<b>Deep One-Class Classification</b> .....	6981
<i>Lukas Ruff, Nico Goernitz, Lucas Deecke, Shoaib Ahmed Siddiqui, Robert Vandermeulen, Alexander Binder, Emmanuel Müller, Marius Kloft</i>	



<b>Augment and Reduce: Stochastic Inference for Large Categorical Distributions</b> .....	6997
<i>Francisco J. R. Ruiz, Michalis K. Titsias, Adji B. Dieng, David M. Blei</i>	
<b>Probabilistic Boolean Tensor Decomposition</b> .....	7007
<i>Tammo Rukat, Chris Holmes, Christopher Yau</i>	
<b>Black-Box Variational Inference for Stochastic Differential Equations</b> .....	7021
<i>Thomas Ryder, Andrew Golightly, A. Stephen McGough, Dennis Prangle</i>	
<b>Spurious Local Minima are Common in Two-Layer ReLU Neural Networks</b> .....	7031
<i>Itay Safran, Ohad Shamir</i>	
<b>Learning Equations for Extrapolation and Control</b> .....	7053
<i>Subham Sahoo, Christoph Lampert, Georg Martius</i>	
<b>Tempered Adversarial Networks</b> .....	7062
<i>Mehdi S. M. Sajjadi, Giambattista Parascandolo, Arash Mehrjou, Bernhard Schölkopf</i>	
<b>Representation Tradeoffs for Hyperbolic Embeddings</b> .....	7071
<i>Frederic Sala, Chris De Sa, Albert Gu, Christopher Re</i>	
<b>Graph Networks as Learnable Physics Engines for Inference and Control</b> .....	7097
<i>Alvaro Sanchez-Gonzalez, Nicolas Heess, Jost Tobias Springenberg, Josh Merel, Martin Riedmiller, Raia Hadsell, Peter Battaglia</i>	
<b>Measuring Abstract Reasoning in Neural Networks</b> .....	7118
<i>Adam Santoro, Felix Hill, David Barrett, Ari Morcos, Timothy Lillicrap</i>	
<b>A Classification-Based Study of Covariate Shift in GAN Distributions</b> .....	7135
<i>Shibani Santurkar, Ludwig Schmidt, Aleksander Madry</i>	
<b>TAPAS: Tricks to Accelerate (encrypted) Prediction As a Service</b> .....	7145
<i>Amartya Sanyal, Matt Kusner, Adria Gascon, Varun Kanade</i>	
<b>Tight Regret Bounds for Bayesian Optimization in One Dimension</b> .....	7155
<i>Jonathan Scarlett</i>	
<b>Learning with Abandonment</b> .....	7172
<i>Sven Schmit, Ramesh Johari</i>	
<b>Not to Cry Wolf: Distantly Supervised Multitask Learning in Critical Care</b> .....	7185
<i>Patrick Schwab, Emanuela Keller, Carl Muroi, David J. Mack, Christian Strässle, Walter Karlen</i>	
<b>Progress &amp; Compress: A Scalable Framework for Continual Learning</b> .....	7199
<i>Jonathan Schwarz, Wojciech Czarnecki, Jelena Luketina, Agnieszka Grabska-Barwinska, Yee Whye Teh, Razvan Pascanu, Raia Hadsell</i>	
<b>Multi-Fidelity Black-Box Optimization with Hierarchical Partitions</b> .....	7212
<i>Rajat Sen, Kirthevasan Kandasamy, Sanjay Shakkottai</i>	
<b>Overcoming Catastrophic Forgetting with Hard Attention to the Task</b> .....	7225
<i>Joan Serra, Didac Suris, Marius Miron, Alexandros Karatzoglou</i>	
<b>Bounding and Counting Linear Regions of Deep Neural Networks</b> .....	7243
<i>Thiago Serra, Christian Tjandraatmadja, Srikumar Ramalingam</i>	
<b>First Order Generative Adversarial Networks</b> .....	7262
<i>Calvin Seward, Thomas Unterthiner, Urs Bergmann, Nikolay Jetchev, Sepp Hochreiter</i>	
<b>Finding Influential Training Samples for Gradient Boosted Decision Trees</b> .....	7287
<i>Boris Sharchilev, Yury Ustinovskiy, Pavel Serdyukov, Maarten Rijke</i>	
<b>Solving Partial Assignment Problems using Random Clique Complexes</b> .....	7297
<i>Charu Sharma, Deepak Nathani, Manohar Kaul</i>	
<b>Adafactor: Adaptive Learning Rates with Sublinear Memory Cost</b> .....	7322
<i>Noam Shazeer, Mitchell Stern</i>	
<b>Locally Private Hypothesis Testing</b> .....	7331
<i>Or Sheffet</i>	
<b>Learning in Integer Latent Variable Models with Nested Automatic Differentiation</b> .....	7350
<i>Daniel Sheldon, Kevin Winner, Debora Sujono</i>	
<b>Towards More Efficient Stochastic Decentralized Learning: Faster Convergence and Sparse Communication</b> .....	7361
<i>Zebang Shen, Aryan Mokhtari, Tengfei Zhou, Peilin Zhao, Hui Qian</i>	
<b>An Algorithmic Framework of Variable Metric Over-Relaxed Hybrid Proximal Extra-Gradient Method</b> .....	7376
<i>Li Shen, Peng Sun, Yitong Wang, Wei Liu, Tong Zhang</i>	
<b>A Spectral Approach to Gradient Estimation for Implicit Distributions</b> .....	7404
<i>Jiaxin Shi, Shengyang Sun, Jun Zhu</i>	
<b>TACO: Learning Task Decomposition via Temporal Alignment for Control</b> .....	7419
<i>Kyriacos Shiarlis, Markus Wulfmeier, Sasha Salter, Shimon Whiteson, Ingmar Posner</i>	
<b>CRAFTML, an Efficient Clustering-based Random Forest for Extreme Multi-label Learning</b> .....	7432
<i>Wissam Siblini, Frank Meyer, Pascale Kuntz</i>	

## VOLUME 11

<b>Asynchronous Stochastic Quasi-Newton MCMC for Non-Convex Optimization</b> .....	7442
<i>Umut Simsekli, Cagatay Yildiz, Than Huy Nguyen, A. Taylan Cemgil, Gael Richard</i>	
<b>K-means Clustering Using Random Matrix Sparsification</b> .....	7460
<i>Kaushik Sinha</i>	
<b>Towards End-to-End Prosody Transfer for Expressive Speech Synthesis with Tacotron</b> .....	7471
<i>RJ Skerry-Ryan, Eric Battenberg, Ying Xiao, Yuxuan Wang, Daisy Stanton, Joel Shor, Ron Weiss, Rob Clark, Rif A. Saurous</i>	

<b>An Inference-Based Policy Gradient Method for Learning Options</b> .....	7481
<i>Matthew Smith, Herke Hoof, Joelle Pineau</i>	
<b>Accelerating Natural Gradient with Higher-Order Invariance</b> .....	7491
<i>Yang Song, Jiaming Song, Stefano Ermon</i>	
<b>Knowledge Transfer with Jacobian Matching</b> .....	7515
<i>Suraj Srinivas, Francois Fleuret</i>	
<b>Universal Planning Networks</b> .....	7526
<i>Aravind Srinivas, Allan Jabri, Pieter Abbeel, Sergey Levine, Chelsea Finn</i>	
<b>Structured Control Nets for Deep Reinforcement Learning</b> .....	7557
<i>Mario Strouji, Jian Zhang, Ruslan Salakhutdinov</i>	
<b>Approximation Algorithms for Cascading Prediction Models</b> .....	7569
<i>Matthew Streeter</i>	
<b>Learning Low-Dimensional Temporal Representations</b> .....	7578
<i>Bing Su, Ying Wu</i>	
<b>Exploiting the Potential of Standard Convolutional Autoencoders for Image Restoration by Evolutionary Search</b> .....	7592
<i>Masanori Suganuma, Mete Ozay, Takayuki Okatani</i>	
<b>Stagewise Safe Bayesian Optimization with Gaussian Processes</b> .....	7602
<i>Yanan Sui, Vincent Zhuang, Joel Burdick, Yisong Yue</i>	
<b>Neural Program Synthesis from Diverse Demonstration Videos</b> .....	7614
<i>Shao-Hua Sun, Hyeonwoo Noh, Sriram Somasundaram, Joseph Lim</i>	
<b>Scalable Approximate Bayesian Inference for Particle Tracking Data</b> .....	7627
<i>Ruoxi Sun, Liam Paninski</i>	
<b>Graphical Nonconvex Optimization via an Adaptive Convex Relaxation</b> .....	7638
<i>Qiang Sun, Kean Ming Tan, Han Liu, Tong Zhang</i>	
<b>Convolutional Imputation of Matrix Networks</b> .....	7663
<i>Qingyun Sun, Mengyuan Yan, David Donoho, Stephen Boyd</i>	
<b>Differentiable Compositional Kernel Learning for Gaussian Processes</b> .....	7676
<i>Shengyang Sun, Guodong Zhang, Chaoqi Wang, Wenyuan Zeng, Jiaman Li, Roger Grosse</i>	
<b>Learning the Reward Function for a Misspecified Model</b> .....	7697
<i>Erik Talvitie</i>	
<b>D<sub>2</sub> : Decentralized Training over Decentralized Data</b> .....	7711
<i>Hanlin Tang, Xiangru Lian, Ming Yan, Ce Zhang, Ji Liu</i>	
<b>Neural Inverse Rendering for General Reflectance Photometric Stereo</b> .....	7731
<i>Tatsunori Taniati, Takanori Maehara</i>	
<b>Black Box FDR</b> .....	7761
<i>Wesley Tansey, Yixin Wang, David Blei, Raul Rabadan</i>	
<b>Best Arm Identification in Linear Bandits with Linear Dimension Dependency</b> .....	7773
<i>Chao Tao, Saúl Blanco, Yuan Zhou</i>	
<b>X<sup>2</sup> Generative Adversarial Network</b> .....	7787
<i>Chenyang Tao, Liqun Chen, Ricardo Henao, Jianfeng Feng, Lawrence Carin</i>	
<b>Lyapunov Functions for First-Order Methods: Tight Automated Convergence Guarantees</b> .....	7810
<i>Adrien Taylor, Bryan Van Scoy, Laurent Lessard</i>	
<b>Bayesian Uncertainty Estimation for Batch Normalized Deep Networks</b> .....	7824
<i>Mattias Teye, Hossein Azizpour, Kevin Smith</i>	
<b>Decoupling Gradient-Like Learning Rules from Representations</b> .....	7853
<i>Philip Thomas, Christoph Dann, Emma Brunskill</i>	
<b>CoVeR: Learning Covariate-Specific Vector Representations with Tensor Decompositions</b> .....	7869
<i>Kevin Tian, Teng Zhang, James Zou</i>	
<b>Importance Weighted Transfer of Samples in Reinforcement Learning</b> .....	7881
<i>Andrea Tirinzoni, Andrea Sessa, Matteo Pirota, Marcello Restelli</i>	
<b>Adversarial Regression with Multiple Learners</b> .....	7899
<i>Liang Tong, Sixie Yu, Scott Alfeld, Yevgeny Vorobeychik</i>	
<b>Convergent TREE BACKUP and RETRACE with Function Approximation</b> .....	7914
<i>Ahmed Touati, Pierre-Luc Bacon, Doina Precup, Pascal Vincent</i>	
<b>Learning Longer-term Dependencies in RNNs with Auxiliary Losses</b> .....	7930
<i>Trieu Trinh, Andrew Dai, Thang Luong, Quoc Le</i>	
<b>Theoretical Analysis of Sparse Subspace Clustering with Missing Entries</b> .....	7940
<i>Manolis Tsakiris, Rene Vidal</i>	
<b>StrassenNets: Deep Learning with a Multiplication Budget</b> .....	7950
<i>Michael Tschannen, Aran Khanna, Anima Anandkumar</i>	
<b>Invariance of Weight Distributions in Rectified MLPs</b> .....	7962
<i>Russell Tsuchida, Farbod Roosta-Khorasani, Marcus Gallagher</i>	
<b>Least-Squares Temporal Difference Learning for the Linear Quadratic Regulator</b> .....	7975
<i>Stephen Tu, Benjamin Recht</i>	
<b>The Mirage of Action-Dependent Baselines in Reinforcement Learning</b> .....	7985
<i>George Tucker, Surya Bhupatiraju, Shixiang Gu, Richard Turner, Zoubin Ghahramani, Sergey Levine</i>	
<b>Adversarial Risk and the Dangers of Evaluating Against Weak Attacks</b> .....	7995
<i>Jonathan Uesato, Brendan O'Donoghue, Pushmeet Kohli, Aaron Oord</i>	
<b>DVAE++: Discrete Variational Autoencoders with Overlapping Transformations</b> .....	8008
<i>Arash Vahdat, William Macready, Zhengbing Bian, Amir Khoshman, Evgeny Andriyash</i>	

<b>Programmatically Interpretable Reinforcement Learning</b> .....	8024
<i>Abhinav Verma, Vijayaraghavan Murali, Rishabh Singh, Pushmeet Kohli, Swarat Chaudhuri</i>	
<b>A Probabilistic Theory of Supervised Similarity Learning for Pointwise ROC Curve Optimization</b> .....	8037
<i>Robin Vogel, Aurélien Bellet, Stéphan Cléménçon</i>	
<b>Transfer Learning via Learning to Transfer</b> .....	8059
<i>Ying Wei, Yu Zhang, Junzhou Huang, Qiang Yang</i>	
<b>Semi-Supervised Learning on Data Streams via Temporal Label Propagation</b> .....	8078
<i>Tal Wagner, Sudipto Guha, Shiva Prasad Kasiviswanathan, Nina Mishra</i>	
<b>Neural Dynamic Programming for Musical Self Similarity</b> .....	8088
<i>Christian Walder, Dongwoo Kim</i>	
<b>Thompson Sampling for Combinatorial Semi-Bandits</b> .....	8097
<i>Siwei Wang, Wei Chen</i>	
<b>PredRNN++: Towards A Resolution of the Deep-in-Time Dilemma in Spatiotemporal Predictive Learning</b> .....	8122
<i>Yunbo Wang, Zhifeng Gao, Mingsheng Long, Jianmin Wang, Philip S Yu</i>	
<b>Analyzing the Robustness of Nearest Neighbors to Adversarial Examples</b> .....	8132
<i>Yizhen Wang, Somesh Jha, Kamalika Chaudhuri</i>	
<b>Competitive Multi-agent Inverse Reinforcement Learning with Sub-optimal Demonstrations</b> .....	8148
<i>Xingyu Wang, Diego Klabjan</i>	

## VOLUME 12

<b>Coded Sparse Matrix Multiplication</b> .....	8176
<i>Simong Wang, Jiashang Liu, Ness Shroff</i>	
<b>A Fast and Scalable Joint Estimator for Integrating Additional Knowledge in Learning Multiple Related Sparse Gaussian Graphical Models</b> .....	8194
<i>Beilun Wang, Arshdeep Sekhon, Yanjun Qi</i>	
<b>Provable Variable Selection for Streaming Features</b> .....	8220
<i>Jing Wang, Jie Shen, Ping Li</i>	
<b>Style Tokens: Unsupervised Style Modeling, Control and Transfer in End-to-End Speech Synthesis</b> .....	8229
<i>Yuxuan Wang, Daisy Stanton, Yu Zhang, Rj-Skerry Ryan, Eric Battenberg, Joel Shor, Ying Xiao, Ye Jia, Fei Ren, Rif A. Saurous</i>	
<b>Adversarial Distillation of Bayesian Neural Network Posteriors</b> .....	8239
<i>Kuan-Chieh Wang, Paul Vicol, James Lucas, Li Gu, Roger Grosse, Richard Zemel</i>	
<b>Minimax Concave Penalized Multi-Armed Bandit Model with High-Dimensional Covariates</b> .....	8249
<i>Xue Wang, Mike Mingcheng Wei, Tao Yao</i>	
<b>Online Convolutional Sparse Coding with Sample-Dependent Dictionary</b> .....	8267
<i>Yaqing Wang, Quanming Yao, James Tin-Yau Kwok, Lionel M. Ni</i>	
<b>Stein Variational Message Passing for Continuous Graphical Models</b> .....	8278
<i>Dilin Wang, Zhe Zeng, Qiang Liu</i>	
<b>Approximate Leave-One-Out for Fast Parameter Tuning in High Dimensions</b> .....	8290
<i>Shuaiwen Wang, Wenda Zhou, Haihao Lu, Arian Maleki, Vahab Mirrokni</i>	
<b>Hierarchical Multi-Label Classification Networks</b> .....	8321
<i>Jonatas Wehrmann, Ricardo Cerri, Rodrigo Barros</i>	
<b>Curriculum Learning by Transfer Learning: Theory and Experiments with Deep Networks</b> .....	8331
<i>Daphna Weinshall, Gad Cohen, Dan Amir</i>	
<b>Extracting Automata from Recurrent Neural Networks Using Queries and Counterexamples</b> .....	8340
<i>Gail Weiss, Yoav Goldberg, Eran Yahav</i>	
<b>LEAPSANDBOUNDS: A Method for Approximately Optimal Algorithm Configuration</b> .....	8358
<i>Gellert Weisz, Andras Gyorgy, Csaba Szepesvari</i>	
<b>Deep Predictive Coding Network for Object Recognition</b> .....	8369
<i>Haiguang Wen, Kuan Han, Junxing Shi, Yizhen Zhang, Eugenio Culurciello, Zhongming Liu</i>	
<b>Towards Fast Computation of Certified Robustness for ReLU Networks</b> .....	8379
<i>Tsui-Wei Weng, Huan Zhang, Hongge Chen, Zhao Song, Cho-Jui Hsieh, Luca Daniel, Duane Boning, Inderjit Dhillon</i>	
<b>Provable Defenses Against Adversarial Examples via the Convex Outer Adversarial Polytope</b> .....	8405
<i>Eric Wong, Zico Kolter</i>	
<b>Local Density Estimation in High Dimensions</b> .....	8424
<i>Xian Wu, Moses Charikar, Vishnu Natchu</i>	
<b>Adaptive Exploration-Exploitation Tradeoff for Opportunistic Bandits</b> .....	8440
<i>Huasen Wu, Xueying Guo, Xin Liu</i>	
<b>SQL-Rank: A Listwise Approach to Collaborative Ranking</b> .....	8458
<i>Liwei Wu, Cho-Jui Hsieh, James Sharpnack</i>	
<b>Error Compensated Quantized SGD and its Applications to Large-scale Distributed Optimization</b> .....	8472
<i>Jiaxiang Wu, Weidong Huang, Junzhou Huang, Tong Zhang</i>	
<b>Reinforcing Adversarial Robustness using Model Confidence Induced by Adversarial Training</b> .....	8484
<i>Xi Wu, Uyeong Jang, Jiefeng Chen, Lingjiao Chen, Somesh Jha</i>	
<b>Discrete-Continuous Mixtures in Probabilistic Programming: Generalized Semantics and Inference Algorithms</b> .....	8494
<i>Yi Wu, Siddharth Srivastava, Nicholas Hay, Simon Du, Stuart Russell</i>	
<b>Variance Regularized Counterfactual Risk Minimization via Variational Divergence Minimization</b> .....	8508
<i>Hang Wu, May Wang</i>	

<b>Deep k-Means: Re-Training and Parameter Sharing with Harder Cluster Assignments for Compressing Deep Convolutions</b> .....	8523
<i>Junru Wu, Yue Wang, Zhenyu Wu, Zhangyang Wang, Ashok Veeraraghavan, Yingyan Lin</i>	
<b>Bayesian Quadrature for Multiple Related Integrals</b> .....	8533
<i>Xiaoyue Xi, Francois-Xavier Briol, Mark Girolami</i>	
<b>Model-Level Dual Learning</b> .....	8552
<i>Yingce Xia, Xu Tan, Fei Tian, Tao Qin, Nenghai Yu, Tie-Yan Liu</i>	
<b>Dynamical Isometry and a Mean Field Theory of CNNs: How to Train 10,000-Layer Vanilla Convolutional Neural Networks</b> .....	8565
<i>Lechao Xiao, Yasaman Bahri, Jascha Sohl-Dickstein, Samuel Schoenholz, Jeffrey Pennington</i>	
<b>Orthogonality-Promoting Distance Metric Learning: Convex Relaxation and Theoretical Analysis</b> .....	8581
<i>Pengtao Xie, Wei Wu, Yichen Zhu, Eric P. Xing</i>	
<b>Nonoverlap-Promoting Variable Selection</b> .....	8612
<i>Pengtao Xie, Hongbao Zhang, Yichen Zhu, Eric P. Xing</i>	
<b>Learning Semantic Representations for Unsupervised Domain Adaptation</b> .....	8626
<i>Shaoan Xie, Zibin Zheng, Liang Chen, Chuan Chen</i>	
<b>Rates of Convergence of Spectral Methods for Graphon Estimation</b> .....	8636
<i>Jiaming Xu</i>	
<b>Learning Registered Point Processes from Idiosyncratic Observations</b> .....	8646
<i>Hongteng Xu, Lawrence Carin, Hongyuan Zha</i>	
<b>Representation Learning on Graphs with Jumping Knowledge Networks</b> .....	8662
<i>Keyulu Xu, Chengtao Li, Yonglong Tian, Tomohiro Sonobe, Ken-Ichi Kawarabayashi, Stefanie Jegelka</i>	
<b>Learning to Explore via Meta-Policy Gradient</b> .....	8676
<i>Tianbing Xu, Qiang Liu, Liang Zhao, Jian Peng</i>	
<b>Nonparametric Regression with Comparisons: Escaping the Curse of Dimensionality with Ordinal Information</b> .....	8686
<i>Yichong Xu, Hariank Muthakana, Sivaraman Balakrishnan, Aarti Singh, Artur Dubrawski</i>	
<b>Optimal Tuning for Divide-and-conquer Kernel Ridge Regression with Massive Data</b> .....	8707
<i>Ganggang Xu, Zuofeng Shang, Guang Cheng</i>	
<b>Continuous and Discrete-time Accelerated Stochastic Mirror Descent for Strongly Convex Functions</b> .....	8719
<i>Pan Xu, Tianhao Wang, Quanquan Gu</i>	
<b>A Semantic Loss Function for Deep Learning with Symbolic Knowledge</b> .....	8738
<i>Jingyi Xu, Zilu Zhang, Tal Friedman, Yitao Liang, Guy Broeck</i>	
<b>Causal Bandits with Propagating Inference</b> .....	8752
<i>Akihiro Yabe, Daisuke Hatano, Hanna Sumita, Shinji Ito, Naonori Kakimura, Takuro Fukunaga, Ken-Ichi Kawarabayashi</i>	
<b>Active Learning with Logged Data</b> .....	8761
<i>Songbai Yan, Kamalika Chaudhuri, Tara Javidi</i>	
<b>Binary Classification with Karmic, Threshold-Quasi-Concave Metrics</b> .....	8782
<i>Bowei Yan, Oluwasanmi Koyejo, Kai Zhong, Pradeep Ravikumar</i>	
<b>Characterizing and Learning Equivalence Classes of Causal DAGs under Interventions</b> .....	8805
<i>Karren Yang, Abigail Katoff, Caroline Uhler</i>	
<b>Dependent Relational Gamma Process Models for Longitudinal Networks</b> .....	8823
<i>Sikun Yang, Heinz Koeppl</i>	
<b>Goodness-of-fit Testing for Discrete Distributions via Stein Discrepancy</b> .....	8840
<i>Jiasen Yang, Qiang Liu, Vinayak Rao, Jennifer Neville</i>	
<b>Mean Field Multi-Agent Reinforcement Learning</b> .....	8853
<i>Yaodong Yang, Rui Luo, Minne Li, Ming Zhou, Weinan Zhang, Jun Wang</i>	
<b>Yes, but Did It Work?: Evaluating Variational Inference</b> .....	8869
<i>Yuling Yao, Aki Vehtari, Daniel Simpson, Andrew Gelman</i>	
<b>Hierarchical Text Generation and Planning for Strategic Dialogue</b> .....	8887
<i>Denis Yarats, Mike Lewis</i>	
<b>Massively Parallel Algorithms and Hardness for Single-Linkage Clustering under <math>l_p</math> Distances</b> .....	8896
<i>Grigory Yaroslavtsev, Adithya Vadapalli</i>	
<b>Communication-Computation Efficient Gradient Coding</b> .....	8912
<i>Min Ye, Emmanuel Abbe</i>	

### VOLUME 13

<b>Variable Selection via Penalized Neural Network: a Drop-Out-One Loss Approach</b> .....	8922
<i>Mao Ye, Yan Sun</i>	
<b>Loss Decomposition for Fast Learning in Large Output Spaces</b> .....	8937
<i>Ian En-Hsu Yen, Satyen Kale, Felix Yu, Daniel Holtmann-Rice, Sanjiv Kumar, Pradeep Ravikumar</i>	
<b>Byzantine-Robust Distributed Learning: Towards Optimal Statistical Rates</b> .....	8947
<i>Dong Yin, Yudong Chen, Kannan Ramchandran, Peter Bartlett</i>	
<b>Semi-Implicit Variational Inference</b> .....	8976
<i>Mingzhang Yin, Mingyuan Zhou</i>	
<b>Disentangled Sequential Autoencoder</b> .....	8992
<i>Li Yingzhen, Stephan Mandt</i>	
<b>Probably Approximately Metric-Fair Learning</b> .....	9002
<i>Gal Yona, Guy Rothblum</i>	

<b>GAIN: Missing Data Imputation using Generative Adversarial Nets</b> .....	9042
<i>Jinsung Yoon, James Jordon, Mihaela Schaar</i>	
<b>RadialGAN: Leveraging Multiple Datasets to Improve Target-Specific Predictive Models Using Generative Adversarial Networks</b> .....	9060
<i>Jinsung Yoon, James Jordon, Mihaela Schaar</i>	
<b>GraphRNN: Generating Realistic Graphs with Deep Auto-regressive Models</b> .....	9072
<i>Jiaxuan You, Rex Ying, Xiang Ren, William Hamilton, Jure Leskovec</i>	
<b>An Efficient Semismooth Newton Based Algorithm for Convex Clustering</b> .....	9085
<i>Yancheng Yuan, Defeng Sun, Kim-Chuan Toh</i>	
<b>A Conditional Gradient Framework for Composite Convex Minimization with Applications to Semidefinite Programming</b> .....	9096
<i>Alp Yurtsever, Olivier Fercoq, Francesco Locatello, Volkan Cevher</i>	
<b>Orthogonal Machine Learning: Power and Limitations</b> .....	9112
<i>Ilias Zadik, Lester Mackey, Vasilis Syrgkanis</i>	
<b>Problem Dependent Reinforcement Learning Bounds Which Can Identify Bandit Structure in MDPs</b> .....	9125
<i>Andrea Zanette, Emma Brunskill</i>	
<b>Policy Optimization as Wasserstein Gradient Flows</b> .....	9134
<i>Ruiyi Zhang, Changyou Chen, Chunyuan Li, Lawrence Carin</i>	
<b>Fast and Sample Efficient Inductive Matrix Completion via Multi-Phase Procrustes Flow</b> .....	9149
<i>Xiao Zhang, Simon Du, Quanquan Gu</i>	
<b>Large-Scale Sparse Inverse Covariance Estimation via Thresholding and Max-Det Matrix Completion</b> .....	9178
<i>Richard Zhang, Salar Fattahi, Somayeh Sojoudi</i>	
<b>High Performance Zero-Memory Overhead Direct Convolutions</b> .....	9201
<i>Jiyuan Zhang, Franz Franchetti, Tze Meng Low</i>	
<b>Safe Element Screening for Submodular Function Minimization</b> .....	9211
<i>Weizhong Zhang, Bin Hong, Lin Ma, Wei Liu, Tong Zhang</i>	
<b>Improving the Privacy and Accuracy of ADMM-Based Distributed Algorithms</b> .....	9221
<i>Xueru Zhang, Mohammad Mahdi Khalili, Mingyan Liu</i>	
<b>Stabilizing Gradients for Deep Neural Networks via Efficient SVD Parameterization</b> .....	9239
<i>Jiong Zhang, Qi Lei, Inderjit Dhillon</i>	
<b>Learning Long Term Dependencies via Fourier Recurrent Units</b> .....	9256
<i>Jiong Zhang, Yibo Lin, Zhao Song, Inderjit Dhillon</i>	
<b>Tropical Geometry of Deep Neural Networks</b> .....	9265
<i>Liwen Zhang, Gregory Naitzat, Lek-Heng Lim</i>	
<b>Deep Bayesian Nonparametric Tracking</b> .....	9283
<i>Aonan Zhang, John Paisley</i>	
<b>Composable Planning with Attributes</b> .....	9292
<i>Amy Zhang, Sainbayar Sukhbaatar, Adam Lerer, Arthur Szlam, Rob Fergus</i>	
<b>Noisy Natural Gradient as Variational Inference</b> .....	9308
<i>Guodong Zhang, Shengyang Sun, David Duvenaud, Roger Grosse</i>	
<b>A Primal-Dual Analysis of Global Optimality in Nonconvex Low-Rank Matrix Recovery</b> .....	9322
<i>Xiao Zhang, Lingxiao Wang, Yaodong Yu, Quanquan Gu</i>	
<b>Fully Decentralized Multi-Agent Reinforcement Learning with Networked Agents</b> .....	9340
<i>Kaiqing Zhang, Zhuoran Yang, Han Liu, Tong Zhang, Tamer Basar</i>	
<b>Dynamic Regret of Strongly Adaptive Methods</b> .....	9372
<i>Lijun Zhang, Tianbao Yang, Rong Jin, Zhi-Hua Zhou</i>	
<b>Inter and Intra Topic Structure Learning with Word Embeddings</b> .....	9388
<i>He Zhao, Lan Du, Wray Buntine, Mingyuan Zhou</i>	
<b>Adversarially Regularized Autoencoders</b> .....	9405
<i>Junbo Zhao, Yoon Kim, Kelly Zhang, Alexander Rush, Yann Lecun</i>	
<b>MSplit LBI: Realizing Feature Selection and Dense Estimation Simultaneously in Few-shot and Zero-shot Learning</b> .....	9421
<i>Bo Zhao, Xinwei Sun, Yanwei Fu, Yuan Yao, Yizhou Wang</i>	
<b>Composite Marginal Likelihood Methods for Random Utility Models</b> .....	9433
<i>Zhibing Zhao, Lirong Xia</i>	
<b>Lightweight Stochastic Optimization for Minimizing Finite Sums with Infinite Data</b> .....	9455
<i>Shuai Zheng, James Tin-Yau Kwok</i>	
<b>A Robust Approach to Sequential Information Theoretic Planning</b> .....	9476
<i>Sue Zheng, Jason Pacheco, John Fisher</i>	
<b>Revealing Common Statistical Behaviors in Heterogeneous Populations</b> .....	9490
<i>Andrey Zhitnikov, Rotem Mulayoff, Tomer Michaeli</i>	
<b>Understanding Generalization and Optimization Performance of Deep CNNs</b> .....	9506
<i>Pan Zhou, Jiashi Feng</i>	
<b>Distributed Asynchronous Optimization with Unbounded Delays: How Slow Can You Go?</b> .....	9544
<i>Zhengyuan Zhou, Panayotis Mertikopoulos, Nicholas Bambos, Peter Glynn, Yinyu Ye, Li-Jia Li, Li Fei-Fei</i>	
<b>A Simple Stochastic Variance Reduced Algorithm with Fast Convergence Rates</b> .....	9568
<i>Kaiwen Zhou, Fanhua Shang, James Cheng</i>	
<b>Stochastic Variance-Reduced Cubic Regularized Newton Method</b> .....	9597
<i>Dongruo Zhou, Pan Xu, Quanquan Gu</i>	

<b>Racing Thompson: an Efficient Algorithm for Thompson Sampling with Non-conjugate Priors</b> .....	9607
<i>Yichi Zhou, Jun Zhu, Jingwei Zhuo</i>	
<b>Distributed Nonparametric Regression under Communication Constraints</b> .....	9616
<i>Yuancheng Zhu, John Lafferty</i>	
<b>Message Passing Stein Variational Gradient Descent</b> .....	9629
<i>Jingwei Zhuo, Chang Liu, Jiaxin Shi, Jun Zhu, Ning Chen, Bo Zhang</i>	
<b>Stochastic Variance-Reduced Hamilton Monte Carlo Methods</b> .....	9647
<i>Difan Zou, Pan Xu, Quanquan Gu</i>	
<b>Hierarchical Long-term Video Prediction without Supervision</b> .....	9657
<i>Nevan Wichers, Ruben Villegas, Dumitru Erhan, Honglak Lee</i>	
<b>Author Index</b>	