

34th International Conference on Machine Learning (ICML 2017)

Sydney, Australia
6 - 11 August 2017

Volume 1 of 8

ISBN: 978-1-5108-5514-4

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© (2017) 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

Uncovering Causality from Multivariate Hawkes Integrated Cumulants	1
<i>Massil Achab, Emmanuel Bacry, Stéphane Gaïffas, Iacopo Mastromatteo, Jean-François Muzy</i>	
A Unified Maximum Likelihood Approach for Estimating Symmetric Properties of Discrete Distributions	17
<i>Jayadev Acharya, Hirakendu Das, Alon Orlitsky, Ananda Theertha Suresh</i>	
Constrained Policy Optimization	30
<i>Joshua Achiam, David Held, Aviv Tamar, Pieter Abbeel</i>	
The Price of Differential Privacy for Online Learning	48
<i>Naman Agarwal, Karan Singh</i>	
Local Bayesian Optimization of Motor Skills	59
<i>Riad Akrouf, Dmitry Sorokin, Jan Peters, Gerhard Neumann</i>	
Connected Subgraph Detection with Mirror Descent on SDPs	69
<i>Cem Aksoylar, Lorenzo Orecchia, Venkatesh Saligrama</i>	
Learning from Clinical Judgments: Semi-Markov-Modulated Marked Hawkes Processes for Risk Prognosis	80
<i>Ahmed M. Alaa, Scott Hu, Mihaela Schaar</i>	
A Semismooth Newton Method for Fast, Generic Convex Programming	95
<i>Ahmed M. Alaa, Eric Wong, J. Zico Kolter</i>	
Learning Continuous Semantic Representations of Symbolic Expressions	118
<i>Miltiadis Allamanis, Pankajan Chanthirasegaran, Pushmeet Kohli, Charles Sutton</i>	
Natasha: Faster Non-Convex Stochastic Optimization via Strongly Non-Convex Parameter	132
<i>Zeyuan Allen-Zhu</i>	
Doubly Accelerated Methods for Faster CCA and Generalized Eigendecomposition	141
<i>Zeyuan Allen-Zhu, Yuanzhi Li</i>	
Faster Principal Component Regression and Stable Matrix Chebyshev Approximation	150
<i>Zeyuan Allen-Zhu, Yuanzhi Li</i>	
Follow the Compressed Leader: Faster Online Learning of Eigenvectors and Faster MMWU	159
<i>Zeyuan Allen-Zhu, Yuanzhi Li</i>	
Near-Optimal Design of Experiments via Regret Minimization	169
<i>Zeyuan Allen-Zhu, Yuanzhi Li, Aarti Singh, Yining Wang</i>	
OptNet: Differentiable Optimization as a Layer in Neural Networks	179
<i>Brandon Amos, J. Zico Kolter</i>	
Input Convex Neural Networks	192
<i>Brandon Amos, Lei Xu, J. Zico Kolter</i>	
An Efficient, Sparsity-Preserving, Online Algorithm for Low-Rank Approximation	207
<i>David Anderson, Ming Gu</i>	
Modular Multitask Reinforcement Learning with Policy Sketches	229
<i>Jacob Andreas, Dan Klein, Sergey Levine</i>	
Averaged-DQN: Variance Reduction and Stabilization for Deep Reinforcement Learning	240
<i>Oron Anschel, Nir Baram, Nahum Shimkin</i>	
A Simple Multi-Class Boosting Framework with Theoretical Guarantees and Empirical Proficiency	254
<i>Ron Appel, Pietro Perona</i>	
Deep Voice: Real-time Neural Text-to-Speech	264
<i>Sercan Ö. Arik, Mike Chrzanowski, Adam Coates, Gregory Diamos, Andrew Gibiansky, Yongguo Kang, Xian Li, John Miller, Andrew Ng, Jonathan Raiman, Shubho Sengupta, Mohammad Shoeybi</i>	
Oracle Complexity of Second-Order Methods for Finite-Sum Problems	274
<i>Yossi Arjevani, Ohad Shamir</i>	
Wasserstein Generative Adversarial Networks	298
<i>Martin Arjovsky, Soumith Chintala, Léon Bottou</i>	
Generalization and Equilibrium in Generative Adversarial Nets (GANs)	322
<i>Sanjeev Arora, Rong Ge, Yingyu Liang, Tengyu Ma, Yi Zhang</i>	
A Closer Look at Memorization in Deep Networks	350
<i>Devansh Arpit, Stanislaw Jastrzebski, Nicolas Ballas, David Krueger, Emmanuel Bengio, Maxinder S. Kanwal, Tegan Maharaj, Asja Fischer, Aaron Courville, Yoshua Bengio, Simon Lacoste-Julien</i>	
An Alternative Softmax Operator for Reinforcement Learning	360
<i>Kavosh Asadi, Michael L. Littman</i>	
Random Fourier Features for Kernel Ridge Regression: Approximation Bounds and Statistical Guarantees	370
<i>Haim Avron, Michael Kapralov, Cameron Musco, Christopher Musco, Ameya Velingker, Amir Zandieh</i>	
Minimax Regret Bounds for Reinforcement Learning	405
<i>Mohammad Gheshlaghi Azar, Ian Osband, Rémi Munos</i>	
Learning the Structure of Generative Models without Labeled Data	434
<i>Stephen H. Bach, Bryan He, Alexander Ratner, Christopher Ré</i>	
Uniform Deviation Bounds for k-Means Clustering	450
<i>Olivier Bachem, Mario Lucic, S. Hamed Hassani, Andreas Krause</i>	
Distributed and Provably Good Seedings for k-Means in Constant Rounds	468
<i>Olivier Bachem, Mario Lucic, Andreas Krause</i>	

Learning Algorithms for Active Learning	477
<i>Philip Bachman, Alessandro Sordoni, Adam Trischler</i>	
Improving Viterbi is Hard: Better Runtimes Imply Faster Clique Algorithms	487
<i>Arturs Backurs, Christos Tzamos</i>	
Differentially Private Clustering in High-Dimensional Euclidean Spaces	501
<i>Maria-Florina Balcan, Travis Dick, Yingyu Liang, Wenlong Mou, Hongyang Zhang</i>	
Strongly-Typed Agents are Guaranteed to Interact Safely	523
<i>David Balduzzi</i>	
The Shattered Gradients Problem: If Resnets Are the Answer, Then What is the Question?	536
<i>David Balduzzi, Marcus Frean, Lennox Leary, J. P. Lewis, Kurt Wan-Duo Ma, Brian McWilliams</i>	
Neural Taylor Approximations: Convergence and Exploration in Rectifier Networks	550
<i>David Balduzzi, Brian McWilliams, Tony Butler-Yeoman</i>	
Spectral Learning from a Single Trajectory under Finite-State Policies	562
<i>Borja Balle, Odalric-Ambrym Maillard</i>	
Lost Relatives of the Gumbel Trick	588
<i>Matej Balog, Nilesh Tripuraneni, Zoubin Ghahramani, Adrian Weller</i>	
Dynamic Word Embeddings	607
<i>Robert Bamler, Stephan Mandt</i>	
End-to-End Differentiable Adversarial Imitation Learning	622
<i>Nir Baram, Oron Anshel, Itai Caspi, Shie Mannor</i>	
Emulating the Expert: Inverse Optimization through Online Learning	632
<i>Andreas Bärman, Sebastian Pokutta, Oskar Schneider</i>	
Unimodal Probability Distributions for Deep Ordinal Classification	647
<i>Christopher Beckham, Christopher Pal</i>	
Globally Induced Forest: A Prepruning Compression Scheme	656
<i>Jean-Michel Begon, Arnaud Joly, Pierre Geurts</i>	
End-to-End Learning for Structured Prediction Energy Networks	670
<i>David Belanger, Bishan Yang, Andrew McCallum</i>	
Learning to Discover Sparse Graphical Models	682
<i>Eugene Belilovsky, Kyle Kastner, Gael Varoquaux, Matthew B. Blaschko</i>	
A Distributional Perspective on Reinforcement Learning	693
<i>Marc G. Bellemare, Will Dabney, Rémi Munos</i>	
Neural Optimizer Search with Reinforcement Learning	712
<i>Irwan Bello, Barret Zoph, Vijay Vasudevan, Quoc V. Le</i>	
Learning Texture Manifolds with the Periodic Spatial GAN	722
<i>Urs Bergmann, Nikolay Jetchev, Roland Vollgraf</i>	
Differentially Private Learning of Undirected Graphical Models Using Collective Graphical Models	731
<i>Garrett Bernstein, Ryan McKenna, Tao Sun, Daniel Sheldon, Michael Hay, Jerome Miklau</i>	
Efficient Online Bandit Multiclass Learning with $O(\sqrt{T})$ Regret	742
<i>Alina Beygelzimer, Francesco Orabona, Chicheng Zhang</i>	
Guarantees for Greedy Maximization of Non-submodular Functions with Applications	756
<i>Andrew An Bian, Joachim M. Buhmann, Andreas Krause, Sebastian Tschiatschek</i>	
Robust Submodular Maximization: A Non-Uniform Partitioning Approach	783
<i>Ilija Bogunovic, Slobodan Mitrovic, Jonathan Scarlett, Volkan Cevher</i>	

VOLUME 2

Unsupervised Learning by Predicting Noise	802
<i>Piotr Bojanowski, Armand Joulin</i>	
Adaptive Neural Networks for Efficient Inference	812
<i>Tolga Bolukbasi, Joseph Wang, Ofer Dekel, Venkatesh Saligrama</i>	
Compressed Sensing using Generative Models	822
<i>Ashish Bora, Ajil Jalal, Eric Price, Alexandros G. Dimakis</i>	
Programming with a Differentiable Forth Interpreter	842
<i>Matko Bošnjak, Tim Rocktäschel, Jason Naradowsky, Sebastian Riedel</i>	
Practical Gauss-Newton Optimisation for Deep Learning	860
<i>Aleksandar Botev, Hippolyt Ritter, David Barber</i>	
Lazifying Conditional Gradient Algorithms	881
<i>Gábor Braun, Sebastian Pokutta, Daniel Zink</i>	
Clustering High Dimensional Dynamic Data Streams	927
<i>Vladimir Braverman, Gereon Frahling, Harry Lang, Christian Sohler, Lin F. Yang</i>	
On the Sampling Problem for Kernel Quadrature	949
<i>François-Xavier Briol, Chris J. Oates, Jon Cockayne, Wilson Ye Chen, Mark Girolami</i>	
Reduced Space and Faster Convergence in Imperfect-Information Games via Pruning	969
<i>Noam Brown, Tuomas Sandholm</i>	
Globally Optimal Gradient Descent for a ConvNet with Gaussian Inputs	980
<i>Alon Brutzkus, Amir Globerson</i>	
Deep Tensor Convolution on Multicores	1007
<i>David Budden, Alexander Matveev, Shibani Santurkar, Shraman Ray Chaudhuri, Nir Shavit</i>	

Multi-objective Bandits: Optimizing the Generalized Gini Index	1018
<i>Róbert Busa-Fekete, Balázs Szörényi, Paul Weng, Shie Mannor</i>	
Priv'IT: Private and Sample Efficient Identity Testing	1039
<i>Bryan Cai, Constantinos Daskalakis, Gautam Kamath</i>	
Second-Order Kernel Online Convex Optimization with Adaptive Sketching	1052
<i>Daniele Calandriello, Alessandro Lazaric, Michal Valko</i>	
“Convex Until Proven Guilty”: Dimension-Free Acceleration of Gradient Descent on Non-Convex Functions	1069
<i>Yair Carmon, John C. Duchi, Oliver Hinder, Aaron Sidford</i>	
Sliced Wasserstein Kernel for Persistence Diagrams	1092
<i>Mathieu Carrière, Marco Cuturi, Steve Oudot</i>	
Multiple Clustering Views from Multiple Uncertain Experts	1102
<i>Yale Chang, Junxiang Chen, Michael H. Cho, Peter J. Castaldi, Edwin K. Silverman, Jennifer G. Dy</i>	
Uncertainty Assessment and False Discovery Rate Control in High-Dimensional Granger Causal Inference	1120
<i>Aditya Chaudhry, Pan Xu, Quanquan Gu</i>	
Active Heteroscedastic Regression	1156
<i>Kamalika Chaudhuri, Prateek Jain, Nagarajan Natarajan</i>	
Combining Model-Based and Model-Free Updates for Trajectory-Centric Reinforcement Learning	1173
<i>Yevgen Chebotar, Karol Hausman, Marvin Zhang, Gaurav Sukhatme, Stefan Schaal, Sergey Levine</i>	
Robust Structured Estimation with Single-Index Models	1186
<i>Sheng Chen, Arindam Banerjee</i>	
Adaptive Multiple-Arm Identification	1202
<i>Jiecao Chen, Xi Chen, Qin Zhang, Yuan Zhou</i>	
Dueling Bandits with Weak Regret	1211
<i>Bangrui Chen, Peter I. Frazier</i>	
Strong NP-Hardness for Sparse Optimization with Concave Penalty Functions	1230
<i>Yichen Chen, Dongdong Ge, Mengdi Wang, Zizhuo Wang, Yinyu Ye, Hao Yin</i>	
Learning to Learn without Gradient Descent by Gradient Descent	1252
<i>Yutian Chen, Matthew W. Hoffman, Sergio Gómez Colmenarejo, Misha Denil, Timothy P. Lillicrap, Matt Botvinick, Nando Freitas</i>	
Identification and Model Testing in Linear Structural Equation Models using Auxiliary Variables	1261
<i>Bryant Chen, Daniel Kumor, Elias Bareinboim</i>	
Toward Efficient and Accurate Covariance Matrix Estimation on Compressed Data	1271
<i>Xixian Chen, Michael R. Lyu, Irwin King</i>	
Online Partial Least Square Optimization: Dropping Convexity for Better Efficiency and Scalability	1305
<i>Zhehui Chen, Lin F. Yang, Chris Junchi Li, Tuo Zhao</i>	
Learning to Aggregate Ordinal Labels by Maximizing Separating Width	1321
<i>Guangyong Chen, Shengyu Zhang, Di Lin, Hui Huang, Pheng Ann Heng</i>	
Nearly Optimal Robust Matrix Completion	1331
<i>Yeshwanth Cherapanamjeri, Kartik Gupta, Prateek Jain</i>	
Algorithms for ℓ_p Low-Rank Approximation	1358
<i>Flavio Chierichetti, Sreenivas Gollapudi, Ravi Kumar, Silvio Lattanzi, Rina Panigrahy, David P. Woodruff</i>	
MEC: Memory-efficient Convolution for Deep Neural Network	1367
<i>Minsik Cho, Daniel Brand</i>	
On Relaxing Determinism in Arithmetic Circuits	1377
<i>Arthur Choi, Adnan Darwiche</i>	
Improving Stochastic Policy Gradients in Continuous Control with Deep Reinforcement Learning using the Beta Distribution	1386
<i>Po-Wei Chou, Daniel Maturana, Sebastian Scherer</i>	
On Kernelized Multi-armed Bandits	1397
<i>Sayak Ray Chowdhury, Aditya Gopalan</i>	
Parseval Networks: Improving Robustness to Adversarial Examples	1423
<i>Moustapha Cisse, Piotr Bojanowski, Edouard Grave, Yann Dauphin, Nicolas Usunier</i>	
Deep Latent Dirichlet Allocation with Topic-Layer-Adaptive Stochastic Gradient Riemannian MCMC	1433
<i>Yulai Cong, Bo Chen, Hongwei Liu, Mingyuan Zhou</i>	
AdaNet: Adaptive Structural Learning of Artificial Neural Networks	1452
<i>Corinna Cortes, Xavier Gonzalvo, Vitaly Kuznetsov, Mehryar Mohri, Scott Yang</i>	
Random Feature Expansions for Deep Gaussian Processes	1467
<i>Kurt Cutajar, Edwin V. Bonilla, Pietro Michiardi, Maurizio Filippone</i>	
Soft-DTW: a Differentiable Loss Function for Time-Series	1483
<i>Marco Cuturi, Mathieu Blondel</i>	
Understanding Synthetic Gradients and Decoupled Neural Interfaces	1506
<i>Wojciech Marian Czarnecki, Grzegorz Swirszcz, Max Jaderberg, Simon Osindero, Oriol Vinyals, Koray Kavukcuoglu</i>	
Stochastic Generative Hashing	1522
<i>Bo Dai, Ruiqi Guo, Sanjiv Kumar, Niao He, Le Song</i>	
Logarithmic Time One-Against-Some	1539
<i>Hal Daumé III, Nikos Karampatziakis, John Langford, Paul Mineiro</i>	
Language Modeling with Gated Convolutional Networks	1551
<i>Yann N. Dauphin, Angela Fan, Michael Auli, David Grangier</i>	
An Infinite Hidden Markov Model With Similarity-Biased Transitions	1560
<i>Colin Reimer Dawson, Chaofan Huang, Clayton T. Morrison</i>	

Distributed Batch Gaussian Process Optimization	1577
<i>Erik A. Daxberger, Bryan Kian Hsiang Low</i>	

VOLUME 3

Consistency Analysis for Binary Classification Revisited	1597
<i>Krzysztof Dembczynski, Wojciech Kotlowski, Oluwasanmi Koyejo, Nagarajan Natarajan</i>	
iSurvive: An Interpretable, Event-time Prediction Model for mHealth	1614
<i>Walter H. Dempsey, Alexander Moreno, Christy K. Scott, Michael L. Dennis, David H. Gustafson, Susan A. Murphy, James M. Rehg</i>	
Image-to-Markup Generation with Coarse-to-Fine Attention	1631
<i>Yuntian Deng, Anssi Kanervisto, Jeffrey Ling, Alexander M. Rush</i>	
RobustFill: Neural Program Learning under Noisy I/O	1641
<i>Jacob Devlin, Jonathan Uesato, Surya Bhupatiraju, Rishabh Singh, Abdel-Rahman Mohamed, Pushmeet Kohli</i>	
Being Robust (in High Dimensions) Can Be Practical	1659
<i>Ilias Diakonikolas, Gautam Kamath, Daniel M. Kane, Jerry Li, Ankur Moitra, Alistair Stewart</i>	
Probabilistic Path Hamiltonian Monte Carlo	1690
<i>Vu Dinh, Arman Bilge, Cheng Zhang, Frederick A. Matsen IV</i>	
Sharp Minima Can Generalize For Deep Nets	1705
<i>Laurent Dinh, Razvan Pascanu, Samy Bengio, Yoshua Bengio</i>	
A Divergence Bound for Hybrids of MCMC and Variational Inference and an Application to Langevin Dynamics and SGVI	1718
<i>Justin Domke</i>	
Dance Dance Convolution	1734
<i>Chris Donahue, Zachary C. Lipton, Julian McAuley</i>	
Stochastic Variance Reduction Methods for Policy Evaluation	1744
<i>Simon S. Du, Jianshu Chen, Lihong Li, Lin Xiao, Dengyong Zhou</i>	
Rule-Enhanced Penalized Regression by Column Generation using Rectangular Maximum Agreement	1762
<i>Jonathan Eckstein, Noam Goldberg, Ai Kagawa</i>	
Neural Audio Synthesis of Musical Notes with WaveNet Autoencoders	1771
<i>Jesse Engel, Cinjon Resnick, Adam Roberts, Sander Dieleman, Mohammad Norouzi, Douglas Eck, Karen Simonyan</i>	
Statistical Inference for Incomplete Ranking Data: The Case of Rank-Dependent Coarsening	1781
<i>Mohsen Ahmadi Fahandar, Eyke Hüllermeier, Inés Couso</i>	
Maximum Selection and Ranking under Noisy Comparisons	1808
<i>Moein Falahatgar, Alon Orlitsky, Venkatesh Pichapati, Ananda Theertha Suresh</i>	
Fake News Mitigation via Point Process Based Intervention	1823
<i>Mehrdad Farajtabar, Jiachen Yang, Xiaojing Ye, Huan Xu, Rakshit Trivedi, Elias Khalil, Shuang Li, Le Song, Hongyuan Zha</i>	
Regret Minimization in Behaviorally-Constrained Zero-Sum Games	1837
<i>Gabriele Farina, Christian Kroer, Tuomas Sandholm</i>	
Coresets for Vector Summarization with Applications to Network Graphs	1847
<i>Dan Feldman, Sedat Ozer, Daniela Rus</i>	
Model-Agnostic Meta-Learning for Fast Adaptation of Deep Networks	1856
<i>Chelsea Finn, Pieter Abbeel, Sergey Levine</i>	
Input Switched Affine Networks: An RNN Architecture Designed for Interpretability	1869
<i>Jakob N. Foerster, Justin Gilmer, Jascha Sohl-Dickstein, Jan Chorowski, David Sussillo</i>	
Stabilising Experience Replay for Deep Multi-Agent Reinforcement Learning	1879
<i>Jakob Foerster, Nantas Nardelli, Gregory Farquhar, Triantafyllos Afouras, Philip H. S. Torr, Pushmeet Kohli, Shimon Whiteson</i>	
Counterfactual Data-Fusion for Online Reinforcement Learners	1889
<i>Andrew Forney, Judea Pearl, Elias Bareinboim</i>	
Forward and Reverse Gradient-Based Hyperparameter Optimization	1903
<i>Luca Franceschi, Michele Donini, Paolo Frasconi, Massimiliano Pontil</i>	
Learning to Detect Sepsis with a Multitask Gaussian Process RNN Classifier	1914
<i>Joseph Futoma, Sanjay Hariharan, Katherine Heller</i>	
Deep Bayesian Active Learning with Image Data	1923
<i>Yarin Gal, Riashat Islam, Zoubin Ghahramani</i>	
Local-to-Global Bayesian Network Structure Learning	1933
<i>Tian Gao, Kshitij Fadnis, Murray Campbell</i>	
Communication-efficient Algorithms for Distributed Stochastic Principal Component Analysis	1943
<i>Dan Garber, Ohad Shamir, Nathan Srebro</i>	
Differentiable Programs with Neural Libraries	1965
<i>Alexander L. Gaunt, Marc Brockschmidt, Nate Kushman, Daniel Tarlow</i>	
Zonotope Hit-and-run for Efficient Sampling from Projection DPPs	1978
<i>Guillaume Gautier, Rémi Bardenet, Michal Valko</i>	
No Spurious Local Minima in Nonconvex Low Rank Problems: A Unified Geometric Analysis	1990
<i>Rong Ge, Chi Jin, Yi Zheng</i>	
Convolutional Sequence to Sequence Learning	2029
<i>Jonas Gehring, Michael Auli, David Grangier, Denis Yarats, Yann N. Dauphin</i>	
On Context-Dependent Clustering of Bandits	2043
<i>Claudio Gentile, Shuai Li, Purushottam Kar, Alexandros Karatzoglou, Giovanni Zappella, Evans Etrud</i>	

Neural Message Passing for Quantum Chemistry	2053
<i>Justin Gilmer, Samuel S. Schoenholz, Patrick F. Riley, Oriol Vinyals, George E. Dahl</i>	
Convex Phase Retrieval Without Lifting Via PhaseMax	2071
<i>Tom Goldstein, Christoph Studer</i>	
Preferential Bayesian Optimization	2080
<i>Javier González, Zhenwen Dai, Andreas Damianou, Neil D. Lawrence</i>	
Measuring Sample Quality with Kernels	2090
<i>Jackson Gorham, Lester Mackey</i>	
Efficient Softmax Approximation for GPUs	2111
<i>Grave, Armand Joulin, Moustapha Cissé, David Grangier, Hervé Jégou</i>	
Automated Curriculum Learning for Neural Networks	2120
<i>Alex Graves, Marc G. Bellemare, Jacob Menick, Rémi Munos, Koray Kavukcuoglu</i>	
On Calibration of Modern Neural Networks	2130
<i>Chuan Guo, Geoff Pleiss, Yu Sun, Kilian Q. Weinberger</i>	
ProtoNN: Compressed and Accurate kNN for Resource-scarce Devices	2144
<i>Chirag Gupta, Arun Sai Suggala, Ankit Goyal, Harsha Vardhan Simhadri, Bhargavi Paranjape, Ashish Kumar, Saurabh Goyal, Raghavendra Udupa, Manik Varma, Prateek Jain</i>	
Deep Value Networks Learn to Evaluate and Iteratively Refine Structured Outputs	2160
<i>Michael Gygli, Mohammad Norouzi, Anelia Angelova</i>	
Reinforcement Learning with Deep Energy-Based Policies	2171
<i>Tuomas Haarnoja, Haoran Tang, Pieter Abbeel, Sergey Levine</i>	
DeepBach: a Steerable Model for Bach Chorales Generation	2187
<i>Gaëtan Hadjeres, François Pachet, Frank Nielsen</i>	
Consistent On-Line Off-Policy Evaluation	2197
<i>Assaf Hallak, Shie Mannor</i>	
Faster Greedy MAP Inference for Determinantal Point Processes	2215
<i>Insu Han, Prabhajan Kambadur, Kyoungsoo Park, Jinwoo Shin</i>	
Data-Efficient Policy Evaluation Through Behavior Policy Search	2229
<i>Josiah P. Hanna, Philip S. Thomas, Peter Stone, Scott Niekum</i>	
Joint Dimensionality Reduction and Metric Learning: A Geometric Take	2244
<i>Mehrtash Harandi, Mathieu Salzmann, Richard Hartley</i>	
Deep IV: A Flexible Approach for Counterfactual Prediction	2257
<i>Jason Hartford, Greg Lewis, Kevin Leyton-Brown, Matt Taddy</i>	
Robust Guarantees of Stochastic Greedy Algorithms	2269
<i>Avinatan Hassidim, Yaron Singer</i>	
Efficient Regret Minimization in Non-Convex Games	2278
<i>Elad Hazan, Karan Singh, Cyril Zhang</i>	
Kernelized Support Tensor Machines	2289
<i>Lifang He, Chun-Ta Lu, Guixiang Ma, Shen Wang, Linlin Shen, Philip S. Yu, Ann B. Ragin</i>	
The Sample Complexity of Online One-Class Collaborative Filtering	2299
<i>Reinhard Heckel, Kannan Ramchandran</i>	
Warped Convolutions: Efficient Invariance to Spatial Transformations	2316
<i>João F. Henriques, Andrea Vedaldi</i>	
Parallel and Distributed Thompson Sampling for Large-scale Accelerated Exploration of Chemical Space	2325
<i>José Miguel Hernández-Lobato, James Requeima, Edward O. Pyzer-Knapp, Alán Aspuru-Guzik</i>	
DARLA: Improving Zero-Shot Transfer in Reinforcement Learning	2335
<i>Irina Higgins, Arka Pal, Andrei Rusu, Loic Matthey, Christopher Burgess, Alexander Pritzel, Matthew Botvinick, Charles Blundell, Alexander Lerchner</i>	
SPLICE: Fully Tractable Hierarchical Extension of ICA with Pooling	2351
<i>Jun-Ichiro Hirayama, Aapo Hyvärinen, Motoaki Kawanabe</i>	
Multilevel Clustering via Wasserstein Means	2363
<i>Nhat Ho, Xuanlong Nguyen, Mikhail Yurochkin, Hung Hai Bui, Viet Huynh, Dinh Phung</i>	
Learning Deep Latent Gaussian Models with Markov Chain Monte Carlo	2379
<i>Matthew D. Hoffman</i>	

VOLUME 4

Minimizing Trust Leaks for Robust Sybil Detection	2391
<i>János Höner, Shinichi Nakajima, Alexander Bauer, Klaus-Robert Müller, Nico Görnitz</i>	
Prox-PDA: The Proximal Primal-Dual Algorithm for Fast Distributed Nonconvex Optimization and Learning Over Networks	2402
<i>Mingyi Hong, Davood Hajimezhad, Ming-Min Zhao</i>	
Analysis and Optimization of Graph Decompositions by Lifted Multicuts	2434
<i>Andrea Hornáková, Jan-Hendrik Lange, Bjoern Andres</i>	
Dissipativity Theory for Nesterov's Accelerated Method	2456
<i>Bin Hu, Laurent Lessard</i>	
Learning Discrete Representations via Information Maximizing Self-Augmented Training	2467
<i>Weihua Hu, Takeru Miyato, Seiya Tokui, Eiichi Matsumoto, Masashi Sugiyama</i>	
State-Frequency Memory Recurrent Neural Networks	2482
<i>Hao Hu, Guo-Jun Qi</i>	

Deep Generative Models for Relational Data with Side Information	2492
<i>Changwei Hu, Piyush Rai, Lawrence Carin</i>	
Toward Controlled Generation of Text	2503
<i>Zhiting Hu, Zichao Yang, Xiaodan Liang, Ruslan Salakhutdinov, Eric P. Xing</i>	
Tensor Decomposition with Smoothness	2514
<i>Masaaki Imaizumi, Kohei Hayashi</i>	
Variational Inference for Sparse and Undirected Models	2528
<i>John Inghram, Debora Marks</i>	
Fairness in Reinforcement Learning	2542
<i>Shahin Jabbari, Matthew Joseph, Michael Kearns, Jamie Morgenstern, Aaron Roth</i>	
Decoupled Neural Interfaces using Synthetic Gradients	2558
<i>Max Jaderberg, Wojciech Marian Czarnecki, Simon Osindero, Oriol Vinyals, Alex Graves, David Silver, Koray Kavukcuoglu</i>	
Scalable Generative Models for Multi-label Learning with Missing Labels	2578
<i>Vikas Jain, Nirbhay Modhe, Piyush Rai</i>	
Sequence Tutor: Conservative Fine-Tuning of Sequence Generation Models with KL-control	2587
<i>Natasha Jaques, Shixiang Gu, Dmitry Bahdanau, José Miguel Hernández-Lobato, Richard E. Turner, Douglas Eck</i>	
Bayesian Optimization with Tree-structured Dependencies	2597
<i>Rodolphe Jenatton, Cedric Archambeau, Javier González, Matthias Seeger</i>	
Simultaneous Learning of Trees and Representations for Extreme Classification and Density Estimation	2613
<i>Yacine Jernite, Anna Choromanska, David Sontag</i>	
From Patches to Images: A Nonparametric Generative Model	2634
<i>Geng Ji, Michael C. Hughes, Erik B. Sudderth</i>	
Density Level Set Estimation on Manifolds with DBSCAN	2643
<i>Heinrich Jiang</i>	
Uniform Convergence Rates for Kernel Density Estimation	2656
<i>Heinrich Jiang</i>	
Contextual Decision Processes with Low Bellman Rank are PAC-Learnable	2671
<i>Nan Jiang, Akshay Krishnamurthy, Alekh Agarwal, John Langford, Robert E. Schapire</i>	
Efficient Nonmyopic Active Search	2708
<i>Shali Jiang, Gustavo Malkomes, Geoff Converse, Alyssa Shofner, Benjamin Moseley, Roman Garnett</i>	
How to Escape Saddle Points Efficiently	2727
<i>Chi Jin, Rong Ge, Praneeth Netrapalli, Sham M. Kakade, Michael I. Jordan</i>	
Tunable Efficient Unitary Neural Networks (EUNN) and Their Application to RNNs	2753
<i>Li Jing, Yichen Shen, Tena Dubcek, John Peurifoy, Scott Skirlo, Yann Lecun, Max Tegmark, Marin Soljacic</i>	
An Adaptive Test of Independence with Analytic Kernel Embeddings	2762
<i>Wittawat Jitkrittum, Zoltán Szabó, Arthur Gretton</i>	
StingyCD: Safely Avoiding Wasteful Updates in Coordinate Descent	2785
<i>Tyler B. Johnson, Carlos Guestrin</i>	
Differentially Private Chi-squared Test by Unit Circle Mechanism	2805
<i>Kazuya Kakizaki, Kazuto Fukuchi, Jun Sakuma</i>	
Video Pixel Networks	2821
<i>Nal Kalchbrenner, Aaron Oord, Karen Simonyan, Ivo Danihelka, Oriol Vinyals, Alex Graves, Koray Kavukcuoglu</i>	
Adaptive Feature Selection: Computationally Efficient Online Sparse Linear Regression under RIP	2830
<i>Satyen Kale, Zohar Karnin, Tengyuan Liang, Dávid Pál</i>	
Recursive Partitioning for Personalization using Observational Data	2849
<i>Nathan Kallus</i>	
Multi-fidelity Bayesian Optimisation with Continuous Approximations	2861
<i>Kirthevasan Kandasamy, Gautam Dasarathy, Jeff Schneider, Barnabás Póczos</i>	
Schema Networks: Zero-shot Transfer with a Generative Causal Model of Intuitive Physics	2879
<i>Ken Kansky, Tom Silver, David A. Mély, Mohamed Eldawy, Miguel Lázaro-Gredilla, Xinghua Lou, Nimrod Dorfman, Szymon Sidor, Scott Phoenix, Dileep George</i>	
Learning in POMDPs with Monte Carlo Tree Search	2890
<i>Sammie Katt, Frans A. Oliehoek, Christopher Amato</i>	
Meritocratic Fairness for Cross-Population Selection	2905
<i>Michael Kearns, Aaron Roth, Zhiwei Steven Wu</i>	
On Approximation Guarantees for Greedy Low Rank Optimization	2914
<i>Rajiv Khanna, Ethan R. Elenberg, Alexandros G. Dimakis, Joydeep Ghosh, Sahand Negahban</i>	
Graph-based Isometry Invariant Representation Learning	2928
<i>Renata Khasanova, Pascal Frossard</i>	
Learning to Discover Cross-Domain Relations with Generative Adversarial Networks	2941
<i>Taeksoo Kim, Moonsu Cha, Hyunsoo Kim, Jung Kwon Lee, Jiwon Kim</i>	
SplitNet: Learning to Semantically Split Deep Networks for Parameter Reduction and Model Parallelization	2950
<i>Juyong Kim, Yookoon Park, Gunhee Kim, Sung Ju Hwang</i>	
Cost-Optimal Learning of Causal Graphs	2963
<i>Murat Kocaoglu, Alex Dimakis, Sriram Vishwanath</i>	
Understanding Black-box Predictions via Influence Functions	2976
<i>Pang Wei Koh, Percy Liang</i>	
Sub-sampled Cubic Regularization for Non-convex Optimization	2988
<i>Jonas Moritz Kohler, Aurelien Lucchi</i>	

PixelCNN Models with Auxiliary Variables for Natural Image Modeling	3012
<i>Alexander Kolesnikov, Christoph H. Lampert</i>	
Active Learning for Cost-Sensitive Classification	3022
<i>Akshay Krishnamurthy, Alekh Agarwal, Tzu-Kuo Huang, Hal Daumé III, John Langford</i>	
Evaluating Bayesian Models with Posterior Dispersion Indices	3052
<i>Alp Kucukelbir, Yixin Wang, David M. Blei</i>	
Resource-efficient Machine Learning in 2 KB RAM for the Internet of Things	3062
<i>Ashish Kumar, Saurabh Goyal, Manik Varma</i>	
Grammar Variational Autoencoder	3072
<i>Matt J. Kusner, Brooks Paige, José Miguel Hernández-Lobato</i>	
Co-clustering Through Optimal Transport	3085
<i>Charlotte Laclau, Ievgen Redko, Basarab Matei, Younès Bennani, Vincent Brault</i>	
Conditional Accelerated Lazy Stochastic Gradient Descent	3098
<i>Guanghui Lan, Sebastian Pokutta, Yi Zhou, Daniel Zink</i>	
Consistent k-Clustering	3120
<i>Silvio Lattanzi, Sergei Vassilvitskii</i>	
Deep Spectral Clustering Learning	3130
<i>Marc T. Law, Raquel Urtasun, Richard S. Zemel</i>	
Coordinated Multi-Agent Imitation Learning	3140
<i>Hoang M. Le, Yisong Yue, Peter Carr, Patrick Lucey</i>	
Bayesian Inference on Random Simple Graphs with Power Law Degree Distributions	3153
<i>Juho Lee, Creighton Heaukulani, Zoubin Ghahramani, Lancelot F. James, Seungjin Choi</i>	
Confident Multiple Choice Learning	3169
<i>Kimin Lee, Changho Hwang, Kyoungsoo Park, Jinwoo Shin</i>	
Deriving Neural Architectures from Sequence and Graph Kernels	3181
<i>Tao Lei, Wengong Jin, Regina Barzilay, Tommi Jaakkola</i>	

VOLUME 5

Doubly Greedy Primal-Dual Coordinate Descent for Sparse Empirical Risk Minimization	3191
<i>Qi Lei, Ian En-Hsu Yen, Chao-Yuan Wu, Inderjit S. Dhillon, Pradeep Ravikumar</i>	
Learning to Align the Source Code to the Compiled Object Code	3207
<i>Dor Levy, Lior Wolf</i>	
Dropout Inference in Bayesian Neural Networks with Alpha-divergences	3219
<i>Yingzhen Li, Yarin Gal</i>	
Provable Alternating Gradient Descent for Non-negative Matrix Factorization with Strong Correlations	3233
<i>Yuanzhi Li, Yingyu Liang</i>	
Provably Optimal Algorithms for Generalized Linear Contextual Bandits	3266
<i>Lihong Li, Yu Lu, Dengyong Zhou</i>	
Fast k-Nearest Neighbour Search via Prioritized DCI	3282
<i>Ke Li, Jitendra Malik</i>	
Forest-type Regression with General Losses and Robust Forest	3296
<i>Alexander Hanbo Li, Andrew Martin</i>	
Stochastic Modified Equations and Adaptive Stochastic Gradient Algorithms	3306
<i>Qianxiao Li, Cheng Tai, Weinan E</i>	
Convergence Analysis of Proximal Gradient with Momentum for Nonconvex Optimization	3341
<i>Qunwei Li, Yi Zhou, Yingbin Liang, Pramod K. Varshney</i>	
Exact MAP Inference by Avoiding Fractional Vertices	3358
<i>Erik M. Lindgren, Alexandros G. Dimakis, Adam Klivans</i>	
Leveraging Union of Subspace Structure to Improve Constrained Clustering	3368
<i>John Lipor, Laura Balzano</i>	
Zero-Inflated Exponential Family Embeddings	3381
<i>Li-Ping Liu, David M. Blei</i>	
Iterative Machine Teaching	3390
<i>Weiyang Liu, Bo Dai, Ahmad Humayun, Charlene Tay, Chen Yu, Linda B. Smith, James M. Rehg, Le Song</i>	
Algorithmic Stability and Hypothesis Complexity	3413
<i>Tongliang Liu, Gábor Lugosi, Gergely Neu, Dacheng Tao</i>	
Analogical Inference for Multi-relational Embeddings	3422
<i>Hanxiao Liu, Yuexin Wu, Yiming Yang</i>	
Dual Iterative Hard Thresholding: From Non-convex Sparse Minimization to Non-smooth Concave Maximization	3433
<i>Bo Liu, Xiao-Tong Yuan, Lezi Wang, Qingshan Liu, Dimitris N. Metaxas</i>	
Gram-CTC: Automatic Unit Selection and Target Decomposition for Sequence Labelling	3450
<i>Hairong Liu, Zhenyao Zhu, Xiangang Li, Sanjeev Satheesh</i>	
Learning Infinite Layer Networks Without the Kernel Trick	3460
<i>Roi Livni, Daniel Carmon, Amir Globerson</i>	
Deep Transfer Learning with Joint Adaptation Networks	3470
<i>Mingsheng Long, Han Zhu, Jianmin Wang, Michael I. Jordan</i>	
Multiplicative Normalizing Flows for Variational Bayesian Neural Networks	3480
<i>Christos Louizos, Max Welling</i>	

How Close Are the Eigenvectors of the Sample and Actual Covariance Matrices?	3490
<i>Andreas Loukas</i>	
Learning Deep Architectures via Generalized Whitened Neural Networks	3500
<i>Ping Luo</i>	
Learning Gradient Descent: Better Generalization and Longer Horizons	3509
<i>Kaifeng Lv, Shunhua Jiang, Jian Li</i>	
Spherical Structured Feature Maps for Kernel Approximation	3518
<i>Yueming Lyu</i>	
Stochastic Gradient MCMC Methods for Hidden Markov Models	3529
<i>Yi-An Ma, Nicholas J. Foti, Emily B. Fox</i>	
Self-Paced Co-training	3545
<i>Fan Ma, Deyu Meng, Qi Xie, Zina Li, Xuanyi Dong</i>	
Interactive Learning from Policy-Dependent Human Feedback	3557
<i>James Macglashan, Mark K. Ho, Robert Loftin, Bei Peng, Guan Wang, David L. Roberts, Matthew E. Taylor, Michael L. Littman</i>	
A Laplacian Framework for Option Discovery in Reinforcement Learning	3567
<i>Marlos C. Machado, Marc G. Bellemare, Michael Bowling</i>	
Frame-based Data Factorizations	3583
<i>Sebastian Mair, Ahcène Boubekki, Ulf Brefeld</i>	
Global Optimization of Lipschitz Functions	3592
<i>Cédric Malherbe, Nicolas Vayatis</i>	
On Mixed Memberships and Symmetric Nonnegative Matrix Factorizations	3602
<i>Xueyu Mao, Purnamrita Sarkar, Deepayan Chakrabarti</i>	
Bayesian Models of Data Streams with Hierarchical Power Priors	3628
<i>Andrés Masegosa, Thomas D. Nielsen, Helge Langseth, Dario Ramos-López, Antonio Salmerón, Anders L. Madsen</i>	
Just Sort It! A Simple and Effective Approach to Active Preference Learning	3640
<i>Lucas Maystre, Matthias Grossglauser</i>	
ChoiceRank: Identifying Preferences from Node Traffic in Networks	3657
<i>Lucas Maystre, Matthias Grossglauser</i>	
Deciding How to Decide: Dynamic Routing in Artificial Neural Networks	3666
<i>Mason McGill, Pietro Perona</i>	
Risk Bounds for Transferring Representations With and Without Fine-Tuning	3676
<i>Daniel McNamara, Maria-Florina Balcan</i>	
Nonnegative Matrix Factorization for Time Series Recovery From a Few Temporal Aggregates	3685
<i>Jiali Mei, Yohann De Castro, Yannig Goude, Georges Hébrail</i>	
Adversarial Variational Bayes: Unifying Variational Autoencoders and Generative Adversarial Networks	3694
<i>Lars Mescheder, Sebastian Nowozin, Andreas Geiger</i>	
Efficient Orthogonal Parametrisation of Recurrent Neural Networks Using Householder Reflections	3708
<i>Zakaria Mhammedi, Andrew Hellicar, Ashfaqur Rahman, James Bailey</i>	
Discovering Discrete Latent Topics with Neural Variational Inference	3721
<i>Yishu Miao, Edward Grefenstette, Phil Blunsom</i>	
Variational Boosting: Iteratively Refining Posterior Approximations	3732
<i>Andrew C. Miller, Nicholas J. Foti, Ryan P. Adams</i>	
Device Placement Optimization with Reinforcement Learning	3748
<i>Azalia Mirhoseini, Hieu Pham, Quoc V. Le, Benoit Steiner, Rasmus Larsen, Yuefeng Zhou, Naveen Kumar, Mohammad Norouzi, Samy Bengio, Jeff Dean</i>	
Tight Bounds for Approximate Carathéodory and Beyond	3758
<i>Vahab Mirrokni, Renato Paes Leme, Adrian Vladu, Sam Chiu-Wai Wong</i>	
Deletion-Robust Submodular Maximization: Data Summarization with “the Right to be Forgotten”	3780
<i>Baharan Mirzsoleiman, Amin Karbasi, Andreas Krause</i>	
Prediction and Control with Temporal Segment Models	3791
<i>Nikhil Mishra, Pieter Abbeel, Igor Mordatch</i>	
Improving Gibbs Sampler Scan Quality with DoGS	3803
<i>Ioannis Mitliagkas, Lester Mackey</i>	
Differentially Private Submodular Maximization: Data Summarization in Disguise	3818
<i>Marko Mitrovic, Mark Bun, Andreas Krause, Amin Karbasi</i>	
Active Learning for Top-K Rank Aggregation from Noisy Comparisons	3842
<i>Soheil Mohajer, Changho Suh, Adel Elmahdy</i>	
Variational Dropout Sparsifies Deep Neural Networks	3854
<i>Dmitry Molchanov, Arsenii Ashukha, Dmitry Vetrov</i>	
Regularising Non-linear Models Using Feature Side-information	3864
<i>Amina Mollaysa, Pablo Strasser, Alexandros Kalousis</i>	
Coupling Distributed and Symbolic Execution for Natural Language Queries	3876
<i>Lili Mou, Zhengdong Lu, Hang Li, Zhi Jin</i>	
McGAN: Mean and Covariance Feature Matching GAN	3885
<i>Youssef Mroueh, Tom Sercu, Vaibhava Goel</i>	
Sequence to Better Sequence: Continuous Revision of Combinatorial Structures	3900
<i>Jonas Mueller, David Gifford, Tommi Jaakkola</i>	
Variants of RMSProp and Adagrad with Logarithmic Regret Bounds	3917
<i>Mahesh Chandra Mulkamala, Matthias Hein</i>	

Meta Networks	3933
<i>Tsendsuren Munkhdalai, Hong Yu</i>	
Understanding the Representation and Computation of Multilayer Perceptrons: A Case Study in Speech Recognition	3944
<i>Tasha Nagamine, Nima Mesgarani</i>	
Adaptive Sampling Probabilities for Non-Smooth Optimization	3954
<i>Hongseok Namkoong, Aman Sinha, Steve Yadowsky, John C. Duchi</i>	
Delta Networks for Optimized Recurrent Network Computation	3971
<i>Daniel Neil, Jun Haeng Lee, Tobi Delbruck, Shih-Chii Liu</i>	

VOLUME 6

Post-Inference Prior Swapping	3981
<i>Willie Neiswanger, Eric Xing</i>	
The Loss Surface of Deep and Wide Neural Networks	3995
<i>Quynh Nguyen, Matthias Hein</i>	
SARAH: A Novel Method for Machine Learning Problems Using Stochastic Recursive Gradient	4009
<i>Lam M. Nguyen, Jie Liu, Katya Scheinberg, Martin Takác</i>	
Composing Tree Graphical Models with Persistent Homology Features for Clustering Mixed-Type Data	4023
<i>Xiuyan Ni, Novi Quadrianto, Yusu Wang, Chao Chen</i>	
Multichannel End-to-end Speech Recognition	4033
<i>Tsubasa Ochiai, Shinji Watanabe, Takaaki Hori, John R. Hershey</i>	
Conditional Image Synthesis with Auxiliary Classifier GANs	4043
<i>Augustus Odena, Christopher Olah, Jonathon Shlens</i>	
Nyström Method with Kernel K-means++ Samples as Landmarks	4056
<i>Dino Oglic, Thomas Gärtner</i>	
Zero-Shot Task Generalization with Multi-Task Deep Reinforcement Learning	4082
<i>Junhyuk Oh, Satinder Singh, Honglak Lee, Pushmeet Kohli</i>	
The Statistical Recurrent Unit	4098
<i>Junier B. Oliva, Barnabás Póczos, Jeff Schneider</i>	
Deep Decentralized Multi-task Multi-Agent Reinforcement Learning under Partial Observability	4108
<i>Shayegan Omidshafiei, Jason Pazis, Christopher Amato, Jonathan P. How, John Vian</i>	
Algebraic Variety Models for High-Rank Matrix Completion	4123
<i>Greg Ongie, Rebecca Willett, Robert D. Nowak, Laura Balzano</i>	
Why is Posterior Sampling Better than Optimism for Reinforcement Learning?	4133
<i>Ian Osband, Benjamin Van Roy</i>	
Bidirectional Learning for Time-series Models with Hidden Units	4149
<i>Takayuki Osogami, Hiroshi Kajino, Taro Sekiyama</i>	
Count-Based Exploration with Neural Density Models	4161
<i>Georg Ostrovski, Marc G. Bellemare, Aaron Oord, Rémi Munos</i>	
Dictionary Learning Based on Sparse Distribution Tomography	4176
<i>Pedram Pad, Farnood Salehi, Elisa Celis, Patrick Thiran, Michael Unser</i>	
Stochastic Bouncy Particle Sampler	4186
<i>Ari Pakman, Dar Gilboa, David Carlson, Liam Paninski</i>	
A Birth-Death Process for Feature Allocation	4209
<i>Konstantina Palla, David Knowles, Zoubin Ghahramani</i>	
Prediction under Uncertainty in Sparse Spectrum Gaussian Processes with Applications to Filtering and Control	4238
<i>Yunpeng Pan, Xinyan Yan, Evangelos A. Theodorou, Byron Boots</i>	
Clustering by Sum of Norms: Stochastic Incremental Algorithm, Convergence and Cluster Recovery	4247
<i>Ashkan Panahi, Devdatt Dubhashi, Fredrik D. Johansson, Chiranjib Bhattacharyya</i>	
Curiosity-driven Exploration by Self-supervised Prediction	4261
<i>Deepak Pathak, Pulkit Agrawal, Alexei A. Efros, Trevor Darrell</i>	
Asynchronous Distributed Variational Gaussian Process for Regression	4271
<i>Hao Peng, Shandian Zhe, Xiao Zhang, Yuan Qi</i>	
Geometry of Neural Network Loss Surfaces via Random Matrix Theory	4281
<i>Jeffrey Pennington, Yasaman Bahri</i>	
Multi-task Learning with Labeled and Unlabeled Tasks	4292
<i>Anastasia Pentina, Christoph H. Lampert</i>	
Robust Adversarial Reinforcement Learning	4310
<i>Lerrel Pinto, James Davidson, Rahul Sukthankar, Abhinav Gupta</i>	
Neural Episodic Control	4320
<i>Alexander Pritzel, Benigno Uria, Sriram Srinivasan, Adrià Puigdomènech Badia, Oriol Vinyals, Demis Hassabis, Daan Wierstra, Charles Blundell</i>	
Online and Linear-Time Attention by Enforcing Monotonic Alignments	4332
<i>Colin Raffel, Minh-Thang Luong, Peter J. Liu, Ron J. Weiss, Douglas Eck</i>	
On the Expressive Power of Deep Neural Networks	4351
<i>Maithra Raghu, Ben Poole, Jon Kleinberg, Surya Ganguli, Jascha Sohl-Dickstein</i>	
Estimating the Unseen from Multiple Populations	4375
<i>Aditi Raghunathan, Gregory Valiant, James Zou</i>	

Coherence Pursuit: Fast, Simple, and Robust Subspace Recovery	4388
<i>Mostafa Rahmani, George Atia</i>	
Innovation Pursuit: A New Approach to the Subspace Clustering Problem	4398
<i>Mostafa Rahmani, George Atia</i>	
High Dimensional Bayesian Optimization with Elastic Gaussian Process	4407
<i>Santu Rana, Cheng Li, Sunil Gupta, Vu Nguyen, Svetha Venkatesh</i>	
Equivariance Through Parameter-Sharing	4416
<i>Siamak Ravanbakhsh, Jeff Schneider, Barnabás Póczos</i>	
Large-Scale Evolution of Image Classifiers	4429
<i>Esteban Real, Sherry Moore, Andrew Selle, Saurabh Saxena, Yutaka Leon Suematsu, Jie Tan, Quoc V. Le, Alexey Kurakin</i>	
Parallel Multiscale Autoregressive Density Estimation	4447
<i>Scott Reed, Aïron Oord, Nal Kalchbrenner, Sergio Gómez Colmenarejo, Ziyu Wang, Yutian Chen, Dan Belov, Nando Freitas</i>	
Real-Time Adaptive Image Compression	4457
<i>Oren Rippel, Lubomir Bourdev</i>	
Active Learning for Accurate Estimation of Linear Models	4474
<i>Carlos Riquelme, Mohammad Ghavamzadeh, Alessandro Lazaric</i>	
Cognitive Psychology for Deep Neural Networks: A Shape Bias Case Study	4510
<i>Samuel Ritter, David G. T. Barrett, Adam Santoro, Matt M. Botvinick</i>	
Pain-Free Random Differential Privacy with Sensitivity Sampling	4520
<i>Benjamin I. P. Rubinstein, Francesco Aldà</i>	
Enumerating Distinct Decision Trees	4530
<i>Salvatore Ruggieri</i>	
Bayesian Boolean Matrix Factorisation	4539
<i>Tammo Rukat, Chris C. Holmes, Michalis K. Titsias, Christopher Yau</i>	
Depth-Width Tradeoffs in Approximating Natural Functions with Neural Networks	4550
<i>Itay Safran, Ohad Shamir</i>	
Asymmetric Tri-training for Unsupervised Domain Adaptation	4573
<i>Kuniaki Saito, Yoshitaka Ushiku, Tatsuya Harada</i>	
Semi-Supervised Classification Based on Classification from Positive and Unlabeled Data	4586
<i>Tomoya Sakai, Marthinus Christoffel Plessis, Gang Niu, Masashi Sugiyama</i>	
Analytical Guarantees on Numerical Precision of Deep Neural Networks	4603
<i>Charbel Sakr, Yongjune Kim, Naresh Shanbhag</i>	
Hierarchy Through Composition with Multitask LMDPs	4616
<i>Andrew M. Saxe, Adam C. Earle, Benjamin Rosman</i>	
Optimal Algorithms for Smooth and Strongly Convex Distributed Optimization in Networks	4630
<i>Kevin Scaman, Francis Bach, Sébastien Bubeck, Yin Tat Lee, Laurent Massoulié</i>	
Adapting Kernel Representations Online Using Submodular Maximization	4643
<i>Matthew Schlegel, Yangchen Pan, Jiecao Chen, Martha White</i>	
Developing Bug-Free Machine Learning Systems With Formal Mathematics	4661
<i>Daniel Selsam, Percy Liang, David L. Dill</i>	
Identifying Best Interventions through Online Importance Sampling	4671
<i>Rajat Sen, Karthikeyan Shanmugam, Alexandros G. Dimakis, Sanjay Shakkottai</i>	
Failures of Gradient-Based Deep Learning	4694
<i>Shai Shalev-Shwartz, Ohad Shamir, Shaked Shammah</i>	
Estimating Individual Treatment Effect: Generalization Bounds and Algorithms	4709
<i>Uri Shalit, Fredrik D. Johansson, David Sontag</i>	
Online Learning with Local Permutations and Delayed Feedback	4719
<i>Ohad Shamir, Liran Szlak</i>	
Orthogonalized ALS: A Theoretically Principled Tensor Decomposition Algorithm for Practical Use	4738
<i>Vatsal Sharan, Gregory Valiant</i>	

VOLUME 7

Differentially Private Ordinary Least Squares	4774
<i>Or Sheffet</i>	
On the Iteration Complexity of Support Recovery via Hard Thresholding Pursuit	4802
<i>Jie Shen, Ping Li</i>	
GSOS: Gauss-Seidel Operator Splitting Algorithm for Multi-Term Nonsmooth Convex Composite Optimization	4824
<i>Li Shen, Wei Liu, Ganzhao Yuan, Shiqian Ma</i>	
World of Bits: An Open-Domain Platform for Web-Based Agents	4834
<i>Tianlin Shi, Andrej Karpathy, Linxi Fan, Jonathan Hernandez, Percy Liang</i>	
Learning Important Features Through Propagating Activation Differences	4844
<i>Avanti Shrikumar, Peyton Greenside, Anshul Kundaje</i>	
Optimal Densification for Fast and Accurate Minwise Hashing	4867
<i>Anshumali Shrivastava</i>	
Bottleneck Conditional Density Estimation	4879
<i>Rui Shu, Hung H. Bui, Mohammad Ghavamzadeh</i>	
Attentive Recurrent Comparators	4890
<i>Pranav Shyam, Shubham Gupta, Ambedkar Dukkipati</i>	

Gradient Boosted Decision Trees for High Dimensional Sparse Output	4899
<i>Si Si, Huan Zhang, S. Sathiya Keerthi, Dhruv Mahajan, Inderjit S. Dhillon, Cho-Jui Hsieh</i>	
The Predictron: End-To-End Learning and Planning	4909
<i>David Silver, Hado Hasselt, Matteo Hessel, Tom Schaul, Arthur Guez, Tim Harley, Gabriel Dulac-Arnold, David Reichert, Neil Rabinowitz, Andre Barreto, Thomas Degris</i>	
Fractional Langevin Monte Carlo: Exploring Levy Driven Stochastic Differential Equations for Markov Chain Monte Carlo	4921
<i>Umut Simsekli</i>	
Nonparanormal Information Estimation	4939
<i>Shashank Singh, Barnabás Póczos</i>	
High-Dimensional Structured Quantile Regression	4953
<i>Vidyashankar Sivakumar, Arindam Banerjee</i>	
Robust Budget Allocation via Continuous Submodular Functions	4963
<i>Matthew Staib, Stefanie Jegelka</i>	
Probabilistic Submodular Maximization in Sub-Linear Time	4981
<i>Serban Stan, Morteza Zadimoghaddam, Andreas Krause, Amin Karbasi</i>	
Approximate Steepest Coordinate Descent	4993
<i>Sebastian U. Stich, Anant Raj, Martin Jaggi</i>	
Ordinal Graphical Models: A Tale of Two Approaches	5008
<i>Arun Sai Suggala, Eunho Yang, Pradeep Ravikumar</i>	
Tensor Balancing on Statistical Manifold	5027
<i>Mahito Sugiyama, Hiroyuki Nakahara, Koji Tsuda</i>	
Safety-Aware Algorithms for Adversarial Contextual Bandit	5037
<i>Wen Sun, Debadeepta Dey, Ashish Kapoor</i>	
Relative Fisher Information and Natural Gradient for Learning Large Modular Models	5058
<i>Ke Sun, Frank Nielsen</i>	
meProp: Sparsified Back Propagation for Accelerated Deep Learning with Reduced Overfitting	5080
<i>Xu Sun, Xuancheng Ren, Shuming Ma, Houfeng Wang</i>	
Deeply AggreVaTeD: Differentiable Imitation Learning for Sequential Prediction	5090
<i>Wen Sun, Arun Venkatraman, Geoffrey J. Gordon, Byron Boots, J. Andrew Bagnell</i>	
Axiomatic Attribution for Deep Networks	5109
<i>Mukund Sundararajan, Ankur Taly, Qiqi Yan</i>	
Distributed Mean Estimation with Limited Communication	5119
<i>Ananda Theertha Suresh, Felix X. Yu, Sanjiv Kumar, H. Brendan McMahan</i>	
Selective Inference for Sparse High-Order Interaction Models	5129
<i>Shinya Suzumura, Kazuya Nakagawa, Yuta Umezū, Koji Tsuda, Ichiro Takeuchi</i>	
Coherent Probabilistic Forecasts for Hierarchical Time Series	5143
<i>Souhaib Ben Taieb, James W. Taylor, Rob J. Hyndman</i>	
Partitioned Tensor Factorizations for Learning Mixed Membership Models	5156
<i>Zilong Tan, Sayan Mukherjee</i>	
Gradient Coding: Avoiding Stragglers in Distributed Learning	5166
<i>Rashish Tandon, Qi Lei, Alexandros G. Dimakis, Nikos Karampatziakis</i>	
Gradient Projection Iterative Sketch for Large-Scale Constrained Least-Squares	5179
<i>Junqi Tang, Mohammad Golbabae, Mike E. Davies</i>	
Neural Networks and Rational Functions	5195
<i>Matus Telgarsky</i>	
Stochastic DCA for the Large-sum of Non-convex Functions Problem and its Application to Group Variable Selection in Classification	5211
<i>Hoi An Le Thi, Hoi Minh Le, Duy Nhat Phan, Bach Tran</i>	
An Analytical Formula of Population Gradient for Two-layered ReLU Network and its Applications in Convergence and Critical Point Analysis	5221
<i>Yuandong Tian</i>	
Evaluating the Variance of Likelihood-Ratio Gradient Estimators	5244
<i>Seiya Tokui, Issei Sato</i>	
Accelerating Eulerian Fluid Simulation With Convolutional Networks	5258
<i>Jonathan Tompson, Kristofer Schlachter, Pablo Sprechmann, Ken Perlin</i>	
Boosted Fitted Q-Iteration	5268
<i>Samuele Tosatto, Matteo Pirodda, Carlo D'Eramo, Marcello Restelli</i>	
Diameter-Based Active Learning	5283
<i>Christopher Tosh, Sanjoy Dasgupta</i>	
Magnetic Hamiltonian Monte Carlo	5292
<i>Nilesh Tripuraneni, Mark Rowland, Zoubin Ghahramani, Richard Turner</i>	
Know-Evolve: Deep Temporal Reasoning for Dynamic Knowledge Graphs	5313
<i>Rakshit Trivedi, Hanjun Dai, Yichen Wang, Le Song</i>	
Hyperplane Clustering via Dual Principal Component Pursuit	5328
<i>Manolis C. Tsakiris, René Vidal</i>	
Breaking Locality Accelerates Block Gauss-Seidel	5338
<i>Stephen Tu, Shivaram Venkataraman, Ashia C. Wilson, Alex Gittens, Michael I. Jordan, Benjamin Recht</i>	
Multilabel Classification with Group Testing and Codes	5348
<i>Shashanka Ubaru, Arya Mazumdar</i>	

Learning Stable Stochastic Nonlinear Dynamical Systems	5361
<i>Jonas Umlauf, Sandra Hirche</i>	
Learning Determinantal Point Processes with Moments and Cycles	5370
<i>John Urschel, Victor-Emmanuel Brunel, Ankur Moitra, Philippe Rigollet</i>	
Automatic Discovery of the Statistical Types of Variables in a Dataset.....	5380
<i>Isabel Valera, Zoubin Ghahramani</i>	
Model-Independent Online Learning for Influence Maximization	5389
<i>Sharan Vaswani, Branislav Kveton, Zheng Wen, Mohammad Ghavamzadeh, Laks V. S. Lakshmanan, Mark Schmidt</i>	
FeUdal Networks for Hierarchical Reinforcement Learning.....	5409
<i>Alexander Sasha Vezhnevets, Simon Osindero, Tom Schaul, Nicolas Heess, Max Jaderberg, David Silver, Koray Kavukcuoglu</i>	
Scalable Multi-Class Gaussian Process Classification using Expectation Propagation	5419
<i>Carlos Villacampa-Calvo, Daniel Hernández-Lobato</i>	
Learning to Generate Long-term Future Via Hierarchical Prediction.....	5429
<i>Ruben Villegas, Jimei Yang, Yuliang Zou, Sungryull Sohn, Xunyu Lin, Honglak Lee</i>	
On Orthogonality and Learning Recurrent Networks with Long Term Dependencies.....	5450
<i>Eugene Vorontsov, Chiheb Trabelsi, Samuel Kadoury, Chris Pal</i>	
Fast Bayesian Intensity Estimation for the Permenantal Process.....	5459
<i>Christian J. Walder, Adrian N. Bishop</i>	
Optimal and Adaptive Off-policy Evaluation in Contextual Bandits.....	5472
<i>Yu-Xiang Wang, Alekh Agarwal, Miroslav Dudik</i>	
Capacity Releasing Diffusion for Speed and Locality	5495
<i>Di Wang, Kimon Fountoulakis, Monika Henzinger, Michael W. Mahoney, Satish Rao</i>	
Sketched Ridge Regression: Optimization Perspective, Statistical Perspective, and Model Averaging	5511
<i>Shusen Wang, Alex Gittens, Michael W. Mahoney</i>	
Robust Gaussian Graphical Model Estimation with Arbitrary Corruption	5520
<i>Lingxiao Wang, Quanquan Gu</i>	
Max-value Entropy Search for Efficient Bayesian Optimization	5530
<i>Zi Wang, Stefanie Jegelka</i>	
Efficient Distributed Learning with Sparsity.....	5544
<i>Jialei Wang, Mladen Kolar, Nathan Srebro, Tong Zhang</i>	
Robust Probabilistic Modeling with Bayesian Data Reweighting.....	5564
<i>Yixin Wang, Alp Kucukelbir, David M. Blei</i>	

VOLUME 8

Batched High-dimensional Bayesian Optimization via Structural Kernel Learning	5590
<i>Zi Wang, Chengtao Li, Stefanie Jegelka, Pushmeet Kohli</i>	
Tensor Decomposition via Simultaneous Power Iteration	5604
<i>Po-An Wang, Chi-Jen Lu</i>	
Sequence Modeling via Segmentations.....	5622
<i>Chong Wang, Yining Wang, Po-Sen Huang, Abdelrahman Mohamed, Dengyong Zhou, Li Deng</i>	
Variational Policy for Guiding Point Processes	5632
<i>Yichen Wang, Grady Williams, Evangelos Theodorou, Le Song</i>	
Exploiting Strong Convexity from Data with Primal-Dual First-Order Algorithms	5648
<i>Jialei Wang, Lin Xiao</i>	
Beyond Filters: Compact Feature Map for Portable Deep Model.....	5672
<i>Yunhe Wang, Chang Xu, Chao Xu, Dacheng Tao</i>	
A Unified Variance Reduction-Based Framework for Nonconvex Low-Rank Matrix Recovery	5681
<i>Lingxiao Wang, Xiao Zhang, Quanquan Gu</i>	
Source-Target Similarity Modelings for Multi-Source Transfer Gaussian Process Regression	5708
<i>Pengfei Wei, Ramon Sagarna, Yiping Ke, Yew-Soon Ong, Chi-Keong Goh</i>	
Latent Intention Dialogue Models	5718
<i>Tsung-Hsien Wen, Yishu Miao, Phil Blunsom, Steve Young</i>	
Unifying Task Specification in Reinforcement Learning.....	5728
<i>Martha White</i>	
Learned Optimizers that Scale and Generalize	5744
<i>Olga Wichrowska, Niru Maheswaranathan, Matthew W. Hoffman, Sergio Gómez Colmenarejo, Misha Denil, Nando Freitas, Jascha Sohl-Dickstein</i>	
Exact Inference for Integer Latent-Variable Models	5754
<i>Kevin Winner, Debora Sujono, Dan Sheldon</i>	
Tensor Belief Propagation	5765
<i>Andrew Wrigley, Wee Sun Lee, Nan Ye</i>	
A Unified View of Multi-Label Performance Measures.....	5778
<i>Xi-Zhu Wu, Zhi-Hua Zhou</i>	
Dual Supervised Learning	5792
<i>Yingce Xia, Tao Qin, Wei Chen, Jiang Bian, Nenghai Yu, Tie-Yan Liu</i>	
Learning Latent Space Models with Angular Constraints	5804
<i>Pengtao Xie, Yuntian Deng, Yi Zhou, Abhimanu Kumar, Yaoliang Yu, James Zou, Eric P. Xing</i>	

Uncorrelation and Evenness: A New Diversity-Promoting Regularizer	5816
<i>Pengtao Xie, Aarti Singh, Eric P. Xing</i>	
Stochastic Convex Optimization: Faster Local Growth Implies Faster Global Convergence.....	5827
<i>Yi Xu, Qihang Lin, Tianbao Yang</i>	
Learning Hawkes Processes from Short Doubly-Censored Event Sequences	5843
<i>Hongteng Xu, Dixin Luo, Hongyuan Zha</i>	
Adaptive Consensus ADMM for Distributed Optimization	5864
<i>Zheng Xu, Gavin Taylor, Hao Li, Mário A. T. Figueiredo, Xiaoming Yuan, Tom Goldstein</i>	
High-dimensional Non-Gaussian Single Index Models via Thresholded Score Function Estimation	5878
<i>Zhuoran Yang, Krishnakumar Balasubramanian, Han Liu</i>	
Towards K-means-friendly Spaces: Simultaneous Deep Learning and Clustering	5888
<i>Bo Yang, Xiao Fu, Nicholas D. Sidiropoulos, Mingyi Hong</i>	
On The Projection Operator to A Three-view Cardinality Constrained Set.....	5902
<i>Haichuan Yang, Shupeng Gui, Chuyang Ke, Daniel Stefankovic, Ryohei Fujimaki, Ji Liu</i>	
Improved Variational Autoencoders for Text Modeling using Dilated Convolutions.....	5917
<i>Zichao Yang, Zhiting Hu, Ruslan Salakhutdinov, Taylor Berg-Kirkpatrick</i>	
Tensor-Train Recurrent Neural Networks for Video Classification	5929
<i>Yinchong Yang, Denis Krompass, Volker Tresp</i>	
A Richer Theory of Convex Constrained Optimization with Reduced Projections and Improved Rates	5939
<i>Tianbao Yang, Qihang Lin, Lijun Zhang</i>	
Sparse + Group-Sparse Dirty Models: Statistical Guarantees without Unreasonable Conditions and a Case for Non-Convexity	5949
<i>Eunho Yang, Aurélie C. Lozano</i>	
Scalable Bayesian Rule Lists	5971
<i>Hongyu Yang, Cynthia Rudin, Margo Seltzer</i>	
Approximate Newton Methods and Their Local Convergence	5981
<i>Haishan Ye, Luo Luo, Zhihua Zhang</i>	
A Simulated Annealing Based Inexact Oracle for Wasserstein Loss Minimization	6005
<i>Jianbo Ye, James Z. Wang, Jia Li</i>	
Latent Feature Lasso.....	6018
<i>Ian En-Hsu Yen, Wei-Cheng Lee, Sung-En Chang, Arun Sai Suggala, Shou-De Lin, Pradeep Ravikumar</i>	
Combined Group and Exclusive Sparsity for Deep Neural Networks	6031
<i>Jaehong Yoon, Sung Ju Hwang</i>	
Latent LSTM Allocation: Joint Clustering and Non-Linear Dynamic Modeling of Sequence Data	6040
<i>Manzil Zaheer, Amr Ahmed, Alexander J. Smola</i>	
Canopy Fast Sampling with Cover Trees	6050
<i>Manzil Zaheer, Satwik Kottur, Amr Ahmed, José Moura, Alex Smola</i>	
Continual Learning Through Synaptic Intelligence.....	6072
<i>Friedemann Zenke, Ben Poole, Surya Ganguli</i>	
Stochastic Gradient Monomial Gamma Sampler	6083
<i>Yizhe Zhang, Changyou Chen, Zhe Gan, Ricardo Henao, Lawrence Carin</i>	
Adversarial Feature Matching for Text Generation	6093
<i>Yizhe Zhang, Zhe Gan, Kai Fan, Zhi Chen, Ricardo Henao, Dinghan Shen, Lawrence Carin</i>	
Scaling Up Sparse Support Vector Machines by Simultaneous Feature and Sample Reduction	6103
<i>Weizhong Zhang, Bin Hong, Wei Liu, Jieping Ye, Deng Cai, Xiaofei He, Jie Wang</i>	
Re-visiting Learning on Hypergraphs: Confidence Interval and Subgradient Method	6123
<i>Chenzi Zhang, Shuguang Hu, Zhihao Gavin Tang, T-H. Hubert Chan</i>	
ZipML: Training Linear Models with End-to-End Low Precision, and a Little Bit of Deep Learning	6132
<i>Hantian Zhang, Jerry Li, Kaan Kara, Dan Alistarh, Ji Liu, Ce Zhang</i>	
Convexified Convolutional Neural Networks.....	6141
<i>Yuchen Zhang, Percy Liang, Martin J. Wainwright</i>	
Projection-free Distributed Online Learning in Networks	6155
<i>Wenpeng Zhang, Peilin Zhao, Wenwu Zhu, Steven C. H. Hoi, Tong Zhang</i>	
Multi-Class Optimal Margin Distribution Machine.....	6167
<i>Teng Zhang, Zhi-Hua Zhou</i>	
Leveraging Node Attributes for Incomplete Relational Data.....	6176
<i>He Zhao, Lan Du, Wray Buntine</i>	
Theoretical Properties for Neural Networks with Weight Matrices of Low Displacement Rank.....	6186
<i>Liang Zhao, Siyu Liao, Yanzhi Wang, Zhe Li, Jian Tang, Bo Yuan</i>	
Learning Hierarchical Features from Deep Generative Models.....	6195
<i>Shengjia Zhao, Jiaming Song, Stefano Ermon</i>	
Learning Sleep Stages from Radio Signals: A Conditional Adversarial Architecture.....	6205
<i>Mingmin Zhao, Shichao Yue, Dina Katabi, Tommi S. Jaakkola, Matt T. Bianchi</i>	
Follow the Moving Leader in Deep Learning.....	6215
<i>Shuai Zheng, James T. Kwok</i>	
Asynchronous Stochastic Gradient Descent with Delay Compensation.....	6229
<i>Shuxin Zheng, Qi Meng, Taifeng Wang, Wei Chen, Nenghai Yu, Zhi-Ming Ma, Tie-Yan Liu</i>	
Collect at Once, Use Effectively: Making Non-interactive Locally Private Learning Possible	6249
<i>Kai Zheng, Wenlong Mou, Liwei Wang</i>	

Recovery Guarantees for One-hidden-layer Neural Networks	6263
<i>Kai Zhong, Zhao Song, Prateek Jain, Peter L. Bartlett, Inderjit S. Dhillon</i>	
Stochastic Adaptive Quasi-Newton Methods for Minimizing Expected Values.....	6273
<i>Chaoxu Zhou, Wenbo Gao, Donald Goldfarb</i>	
Identify the Nash Equilibrium in Static Games with Random Payoffs.....	6297
<i>Yichi Zhou, Jialian Li, Jun Zhu</i>	
When can Multi-Site Datasets be Pooled for Regression? Hypothesis Tests, ℓ_2 - Consistency and Neuroscience Applications.....	6311
<i>Hao Henry Zhou, Yilin Zhang, Vamsi K. Ithapu, Sterling C. Johnson, Grace Wahba, Vikas Singh</i>	
High-Dimensional Variance-Reduced Stochastic Gradient Expectation-Maximization Algorithm.....	6337
<i>Rongda Zhu, Lingxiao Wang, Chengxiang Zhai, Quanquan Gu</i>	
Recurrent Highway Networks.....	6346
<i>Julian Georg Zilly, Rupesh Kumar Srivastava, Jan Koutník, Jürgen Schmidhuber</i>	
Online Learning to Rank in Stochastic Click Models	6358
<i>Masrour Zoghi, Tomas Tunys, Mohammad Ghavamzadeh, Branislav Kveton, Csaba Szepesvari, Zheng Wen</i>	
Author Index	