

# **33rd International Conference on Machine Learning (ICML 2016)**

New York City, New York, USA  
19-24 June 2016

Volume 1 of 6

## **Editors:**

**Maria Florina Balcan**  
**Kilian Q. Weinberger**

ISBN: 978-1-5108-2900-8

**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© (2016) by International Machine Learning Society (IMLS)  
All rights reserved.

Printed by Curran Associates, Inc. (2016)

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: 845-758-2634  
Email: curran@proceedings.com  
Web: www.proceedings.com

# TABLE OF CONTENTS

## VOLUME 1

<b>NO OOPS, YOU WON'T DO IT AGAIN: MECHANISMS FOR SELF-CORRECTION IN CROWDSOURCING</b> .....	1
<i>Nihar Shah, Dengyong Zhou</i>	
<b>STOCHASTICALLY TRANSITIVE MODELS FOR PAIRWISE COMPARISONS: STATISTICAL AND COMPUTATIONAL ISSUES</b> .....	22
<i>Nihar Shah, Sivaraman Balakrishnan, Aditya Guntuboyina, Martin Wainwright</i>	
<b>UPROOTING AND REROOTING GRAPHICAL MODELS</b> .....	50
<i>Adrian Weller</i>	
<b>A DEEP LEARNING APPROACH TO UNSUPERVISED ENSEMBLE LEARNING</b> .....	72
<i>Uri Shaham, Xiuyuan Cheng, Omer Dror, Ariel Jaffe, Boaz Nadler, Joseph Chang, Yuval Kluger</i>	
<b>REVISITING SEMI-SUPERVISED LEARNING WITH GRAPH EMBEDDINGS</b> .....	86
<i>Zhilin Yang, William Cohen, Ruslan Salakhudinov</i>	
<b>GUIDED COST LEARNING: DEEP INVERSE OPTIMAL CONTROL VIA POLICY OPTIMIZATION</b> .....	95
<i>Chelsea Finn, Sergey Levine, Pieter Abbeel</i>	
<b>DIVERSITY-PROMOTING BAYESIAN LEARNING OF LATENT VARIABLE MODELS</b> .....	118
<i>Pengtao Xie, Jun Zhu, Eric Xing</i>	
<b>ADDITIVE APPROXIMATIONS IN HIGH DIMENSIONAL NONPARAMETRIC REGRESSION VIA THE SALSA</b> .....	122
<i>Kirthevasan Kandasamy, Yaoliang Yu</i>	
<b>HAWKES PROCESSES WITH STOCHASTIC EXCITATIONS</b> .....	142
<i>Young Lee, Kar Wai Lim, Cheng Soon Ong</i>	
<b>DATA-DRIVEN RANK BREAKING FOR EFFICIENT RANK AGGREGATION</b> .....	146
<i>Ashish Khetan, Sewoong Oh</i>	
<b>DROPOUT DISTILLATION</b> .....	156
<i>Samuel Rota Bulò, Lorenzo Porzi, Peter Kotschieder</i>	
<b>METADATA-CONSCIOUS ANONYMOUS MESSAGING</b> .....	166
<i>Giulia Fanti, Peter Kairouz, Sewoong Oh, Kannan Ramchandran, Pramod Viswanath</i>	
<b>THE TEACHING DIMENSION OF LINEAR LEARNERS</b> .....	181
<i>Ji Liu, Xiaojin Zhu, Hrag Ohannessian</i>	
<b>TRUTHFUL UNIVARIATE ESTIMATORS</b> .....	200
<i>Ioannis Caragiannis, Ariel Procaccia, Nisarg Shah</i>	
<b>WHY REGULARIZED AUTO-ENCODERS LEARN SPARSE REPRESENTATION?</b> .....	211
<i>Devansh Arpit, Yingbo Zhou, Hung Ngo, Venu Govindaraju</i>	
<b>K-VARIATES++: MORE PLUSES IN THE K-MEANS++</b> .....	224
<i>Richard Nock, Raphael Canyasse, Roksana Boreli, Frank Nielsen</i>	
<b>MULTI-PLAYER BANDITS – A MUSICAL CHAIRS APPROACH</b> .....	276
<i>Jonathan Rosenski, Ohad Shamir, Liran Szlak</i>	
<b>THE INFORMATION SIEVE</b> .....	299
<i>Greg Ver Steeg, Aram Galstyan</i>	
<b>DEEP SPEECH 2 : END-TO-END SPEECH RECOGNITION IN ENGLISH AND MANDARIN</b> .....	312
<i>Dario Amodei, Rishita Anubhai, Eric Battenberg, Carl Case, Jared Casper, Bryan Catanzaro, Jingdong Chen, Mike Chrzanowski, Adam Coates, Greg Diamos, Erich Elsen, Jesse Engel, Linxi Fan, Christopher Fougner, Awni Hannun, Billy Jun, Tony Han, Patrick Legresley, Xiangang Li, Libby Lin, Sharan Narang, Andrew Ng, Sherjil Ozair, Ryan Prenger, Sheng Qian, Jonathan Raiman, Sanjeev Satheesh, David Seetapun, Shubho Sengupta, Chong Wang, Yi Wang, Zhiqian Wang, Bo Xiao, Yan Xie, Dani Yogatama, Jun Zhan, Zhenyao Zhu</i>	
<b>ON THE CONSISTENCY OF FEATURE SELECTION WITH LASSO FOR NON-LINEAR TARGETS</b> .....	322
<i>Yue Zhang, Weihong Guo, Soumya Ray</i>	
<b>MINIMUM REGRET SEARCH FOR SINGLE- AND MULTI-TASK OPTIMIZATION</b> .....	331
<i>Jan Hendrik Metzen</i>	
<b>CRYPTONETS: APPLYING NEURAL NETWORKS TO ENCRYPTED DATA WITH HIGH THROUGHPUT AND ACCURACY</b> .....	342
<i>Ran Gilad-Bachrach, Nathan Dowlin, Kim Laine, Kristin Lauter, Michael Naehrig, John Wernsing</i>	
<b>THE VARIATIONAL NYSTROM METHOD FOR LARGE-SCALE SPECTRAL PROBLEMS</b> .....	352
<i>Max Vladymyrov, Miguel Carreira-Perpinan</i>	

<b>MULTI-BIAS NON-LINEAR ACTIVATION IN DEEP NEURAL NETWORKS</b> .....	365
<i>Hongyang Li, Wanli Ouyang, Xiaogang Wang</i>	
<b>ASYMMETRIC MULTI-TASK LEARNING BASED ON TASK RELATEDNESS AND LOSS</b> .....	374
<i>Giwoong Lee, Eunho Yang, Sung Ju Hwang</i>	
<b>ACCURATE ROBUST AND EFFICIENT ERROR ESTIMATION FOR DECISION TREES</b> .....	383
<i>Lixin Fan</i>	
<b>FAST STOCHASTIC ALGORITHMS FOR SVD AND PCA: CONVERGENCE PROPERTIES AND CONVEXITY</b> .....	392
<i>Ohad Shamir</i>	
<b>CONVERGENCE OF STOCHASTIC GRADIENT DESCENT FOR PCA</b> .....	420
<i>Ohad Shamir</i>	
<b>DEALBREAKER: A NONLINEAR LATENT VARIABLE MODEL FOR EDUCATIONAL DATA</b> .....	438
<i>Andrew Lan, Tom Goldstein, Richard Baraniuk, Christoph Studer</i>	
<b>A KERNELIZED STEIN DISCREPANCY FOR GOODNESS-OF-FIT TESTS</b> .....	448
<i>Qiang Liu, Jason Lee, Michael Jordan</i>	
<b>VARIABLE ELIMINATION IN THE FOURIER DOMAIN</b> .....	462
<i>Yexiang Xue, Stefano Ermon, Ronan Le Bras, Carla, Bart Selman</i>	
<b>LOW-RANK MATRIX APPROXIMATION WITH STABILITY</b> .....	475
<i>Dongsheng Li, Chao Chen, Qin Lv, Junchi Yan, Li Shang, Stephen Chu</i>	
<b>LINKING LOSSES FOR DENSITY RATIO AND CLASS-PROBABILITY ESTIMATION</b> .....	484
<i>Aditya Menon, Cheng Soon Ong</i>	
<b>STOCHASTIC VARIANCE REDUCTION FOR NONCONVEX OPTIMIZATION</b> .....	505
<i>Sashank J. Reddi, Ahmed Hefny, Suvrit Sra, Barnabas Poczos, Alex Smola</i>	
<b>HIERARCHICAL VARIATIONAL MODELS</b> .....	515
<i>Rajesh Ranganath, Dustin Tran, David Blei</i>	
<b>HIERARCHICAL SPAN-BASED CONDITIONAL RANDOM FIELDS FOR LABELING AND SEGMENTING EVENTS IN WEARABLE SENSOR DATA STREAMS</b> .....	529
<i>Roy Adams, Nazir Saleheen, Edison Thomaz, Abhinav Parate, Santosh Kumar, Benjamin Marlin</i>	
<b>BINARY EMBEDDINGS WITH STRUCTURED HASHED PROJECTIONS</b> .....	539
<i>Anna Choromanska, Krzysztof Choromanski, Mariusz Bojarski, Tony Jebara, Sanjiv Kumar, Yann Lecun</i>	
<b>A VARIATIONAL ANALYSIS OF STOCHASTIC GRADIENT ALGORITHMS</b> .....	555
<i>Stephan Mandt, Matthew Hoffman, David Blei</i>	
<b>ADAPTIVE SAMPLING FOR SGD BY EXPLOITING SIDE INFORMATION</b> .....	567
<i>Siddharth Gopal</i>	
<b>LEARNING FROM MULTIWAY DATA: SIMPLE AND EFFICIENT TENSOR REGRESSION</b> .....	576
<i>Rose Yu, Yan Liu</i>	
<b>A DISTRIBUTED VARIATIONAL INFERENCE FRAMEWORK FOR UNIFYING PARALLEL SPARSE GAUSSIAN PROCESS REGRESSION MODELS</b> .....	585
<i>Trong Nghia Hoang, Quang Minh Hoang, Bryan Kian Hsiang Low</i>	
<b>ONLINE STOCHASTIC LINEAR OPTIMIZATION UNDER ONE-BIT FEEDBACK</b> .....	602
<i>Lijun Zhang, Tianbao Yang, Rong Jin, Yichi Xiao, Zhi-Hua Zhou</i>	
<b>ADAPTIVE ALGORITHMS FOR ONLINE CONVEX OPTIMIZATION WITH LONG-TERM CONSTRAINTS</b> .....	620
<i>Rodolphe Jenatton, Jim Huang, Cedric Archambeau</i>	
<b>ACTIVELY LEARNING HEMIMETRICS WITH APPLICATIONS TO ELICITING USER PREFERENCES</b> .....	630
<i>Adish Singla, Sebastian Tschiatschek, Andreas Krause</i>	
<b>LEARNING SIMPLE ALGORITHMS FROM EXAMPLES</b> .....	639
<i>Wojciech Zaremba, Tomas Mikolov, Armand Joulin, Rob Fergus</i>	
<b>LEARNING PHYSICAL INTUITION OF BLOCK TOWERS BY EXAMPLE</b> .....	648
<i>Adam Lerer, Sam Gross, Rob Fergus</i>	
<b>STRUCTURE LEARNING OF PARTITIONED MARKOV NETWORKS</b> .....	657
<i>Song Liu, Taiji Suzuki, Masashi Sugiyama, Kenji Fukumizu</i>	
<b>TRACKING SLOWLY MOVING CLAIRVOYANT: OPTIMAL DYNAMIC REGRET OF ONLINE LEARNING WITH TRUE AND NOISY GRADIENT</b> .....	681
<i>Tianbao Yang, Lijun Zhang, Rong Jin, Jinfeng Yi</i>	
<b>BEYOND CCA: MOMENT MATCHING FOR MULTI-VIEW MODELS</b> .....	703
<i>Anastasia Podosinnikova, Francis Bach, Simon Lacoste-Julien</i>	
<b>FAST METHODS FOR ESTIMATING THE NUMERICAL RANK OF LARGE MATRICES</b> .....	725
<i>Shashanka Ubaru, Yousef Saad</i>	
<b>UNSUPERVISED DEEP EMBEDDING FOR CLUSTERING ANALYSIS</b> .....	740
<i>Junyuan Xie, Ross Girshick, Ali Farhadi</i>	

<b>EFFICIENT PRIVATE EMPIRICAL RISK MINIMIZATION FOR HIGH-DIMENSIONAL LEARNING</b> .....	750
<i>Shiva Prasad Kasiviswanathan, Hongxia Jin</i>	
<b>PARAMETER ESTIMATION FOR GENERALIZED THURSTONE CHOICE MODELS</b> .....	762
<i>Milan Vojnovic, Seyoung Yun</i>	

## VOLUME 2

<b>A RANDOM MATRIX APPROACH TO ECHO-STATE NEURAL NETWORKS</b> .....	785
<i>Romain Couillet, Gilles Wainrib, Hafiz Tiomoko Ali, Harry Sevi</i>	
<b>SUPERVISED AND SEMI-SUPERVISED TEXT CATEGORIZATION USING LSTM FOR REGION EMBEDDINGS</b> .....	794
<i>Rie Johnson, Tong Zhang</i>	
<b>OPTIMALITY OF BELIEF PROPAGATION FOR CROWDSOURCED CLASSIFICATION</b> .....	803
<i>Jungseul Ok, Sewoong Oh, Jinwoo Shin, Yung Yi</i>	
<b>STABILITY OF CONTROLLERS FOR GAUSSIAN PROCESS FORWARD MODELS</b> .....	819
<i>Julia Vinogradska, Bastian Bischoff, Duy Nguyen-Tuong, Anne Romer, Henner Schmidt, Jan Peters</i>	
<b>LEARNING PRIVATELY FROM MULTIPARTY DATA</b> .....	829
<i>Jihun Hamm, Yingjun Cao, Mikhail Belkin</i>	
<b>NETWORK MORPHISM</b> .....	842
<i>Tao Wei, Changhu Wang, Yong Rui, Chang Wen Chen</i>	
<b>A KRONECKER-FACTORED APPROXIMATE FISHER MATRIX FOR CONVOLUTION LAYERS</b> .....	851
<i>Roger Grosse, James Martens</i>	
<b>EXPERIMENTAL DESIGN ON A BUDGET FOR SPARSE LINEAR MODELS AND APPLICATIONS</b> .....	875
<i>Sathya Narayanan Ravi, Vamsi Ithapu, Sterling Johnson, Vikas Singh</i>	
<b>MINDING THE GAPS FOR BLOCK FRANK-WOLFE OPTIMIZATION OF STRUCTURED SVMs</b> .....	885
<i>Anton Osokin, Jean-Baptiste Alayrac, Isabella Lukasewitz, Puneet Dokania, Simon Lacoste-Julien</i>	
<b>EXACT EXPONENT IN OPTIMAL RATES FOR CROWDSOURCING</b> .....	926
<i>Chao Gao, Yu Lu, Dengyong Zhou</i>	
<b>AUGMENTING SUPERVISED NEURAL NETWORKS WITH UNSUPERVISED OBJECTIVES FOR LARGE-SCALE IMAGE CLASSIFICATION</b> .....	939
<i>Yuting Zhang, Kibok Lee, Honglak Lee</i>	
<b>ONLINE LOW-RANK SUBSPACE CLUSTERING BY BASIS DICTIONARY PURSUIT</b> .....	958
<i>Jie Shen, Ping Li, Huan Xu</i>	
<b>A SELF-CORRECTING VARIABLE-METRIC ALGORITHM FOR STOCHASTIC OPTIMIZATION</b> .....	982
<i>Frank Curtis</i>	
<b>STOCHASTIC QUASI-NEWTON LANGEVIN MONTE CARLO</b> .....	1004
<i>Umut Simsekli, Roland Badeau, Taylan Cemgil, Gaël Richard</i>	
<b>DOUBLY ROBUST OFF-POLICY VALUE EVALUATION FOR REINFORCEMENT LEARNING</b> .....	1022
<i>Nan Jiang, Lihong Li</i>	
<b>FAST RATE ANALYSIS OF SOME STOCHASTIC OPTIMIZATION ALGORITHMS</b> .....	1036
<i>Chao Qu, Huan Xu, Chong Jin Ong</i>	
<b>FAST K-NEAREST NEIGHBOUR SEARCH VIA DYNAMIC CONTINUOUS INDEXING</b> .....	1054
<i>Ke Li, Jitendra Malik</i>	
<b>SMOOTH IMITATION LEARNING FOR ONLINE SEQUENCE PREDICTION</b> .....	1067
<i>Hoang Le, Andrew Kang, Yisong Yue, Peter Carr</i>	
<b>COMMUNITY RECOVERY IN GRAPHS WITH LOCALITY</b> .....	1083
<i>Yuxin Chen, Govinda Kamath, Changho Suh, David Tse</i>	
<b>VARIANCE REDUCTION FOR FASTER NON-CONVEX OPTIMIZATION</b> .....	1093
<i>Zeyuan Allen-Zhu, Elad Hazan</i>	
<b>LOSS FACTORIZATION, WEAKLY SUPERVISED LEARNING AND LABEL NOISE ROBUSTNESS</b> .....	1102
<i>Giorgio Patrini, Frank Nielsen, Richard Nock, Marcello Carioni</i>	
<b>ANALYSIS OF DEEP NEURAL NETWORKS WITH EXTENDED DATA JACOBIAN MATRIX</b> .....	1127
<i>Shengjie Wang, Abdel-Rahman Mohamed, Rich Caruana, Jeff Bilmes, Matthai Philipose, Matthew Richardson, Krzysztof Geras, Gregor Urban, Ozlem Aslan</i>	

<b>DOUBLY DECOMPOSING NONPARAMETRIC TENSOR REGRESSION</b> .....	1136
<i>Masaaki Imaizumi, Kohei Hayashi</i>	
<b>HYPERPARAMETER OPTIMIZATION WITH APPROXIMATE GRADIENT</b> .....	1150
<i>Fabian Pedregosa</i>	
<b>SDCA WITHOUT DUALITY, REGULARIZATION, AND INDIVIDUAL CONVEXITY</b> .....	1160
<i>Shai Shalev-Shwartz</i>	
<b>HETEROSCEDASTIC SEQUENCES: BEYOND GAUSSIANTITY</b> .....	1168
<i>Oren Anava, Shie Mannor</i>	
<b>A NEURAL AUTOREGRESSIVE APPROACH TO COLLABORATIVE FILTERING</b> .....	1179
<i>Yin Zheng, Bangsheng Tang, Wenkui Ding, Hanning Zhou</i>	
<b>ON THE QUALITY OF THE INITIAL BASIN IN OVERSPECIFIED NEURAL NETWORKS</b> .....	1191
<i>Itay Safran, Ohad Shamir</i>	
<b>PRIMAL-DUAL RATES AND CERTIFICATES</b> .....	1220
<i>Celestine Dünnner, Simone Forte, Martin Takac, Martin Jaggi</i>	
<b>MINIMIZING THE MAXIMAL LOSS: HOW AND WHY</b> .....	1253
<i>Shai Shalev-Shwartz, Yonatan Wexler</i>	
<b>THE INFORMATION-THEORETIC REQUIREMENTS OF SUBSPACE CLUSTERING WITH MISSING DATA</b> .....	1266
<i>Daniel Pimentel-Alarcon, Robert Nowak</i>	
<b>PAC LEARNING OF PROBABILISTIC AUTOMATON BASED ON THE METHOD OF MOMENTS</b> .....	1275
<i>Hadrien Glaude, Olivier</i>	
<b>ESTIMATING STRUCTURED VECTOR AUTOREGRESSIVE MODELS</b> .....	1297
<i>Igor Melnyk, Arindam Banerjee</i>	
<b>MIXING RATES FOR THE ALTERNATING GIBBS SAMPLER OVER RESTRICTED BOLTZMANN MACHINES AND FRIENDS</b> .....	1331
<i>Christopher Tosh</i>	
<b>POLYNOMIAL NETWORKS AND FACTORIZATION MACHINES: NEW INSIGHTS AND EFFICIENT TRAINING ALGORITHMS</b> .....	1351
<i>Mathieu Blondel, Masakazu Ishihata, Akinori Fujino, Naonori Ueda</i>	
<b>A NEW PAC-BAYESIAN PERSPECTIVE ON DOMAIN ADAPTATION</b> .....	1368
<i>Pascal Germain, Amaury Habrard, François Laviolette, Emilie Morvant</i>	
<b>CORRELATION CLUSTERING AND BICLUSTERING WITH LOCALLY BOUNDED ERRORS</b> .....	1380
<i>Gregory Puleo, Olgica Milenkovic</i>	
<b>PAC LOWER BOUNDS AND EFFICIENT ALGORITHMS FOR THE MAX K-ARMED BANDIT PROBLEM</b> .....	1389
<i>Yahel David, Nahum Shimkin</i>	
<b>A COMPARATIVE ANALYSIS AND STUDY OF MULTIVIEW CNN MODELS FOR JOINT OBJECT CATEGORIZATION AND POSE ESTIMATION</b> .....	1402
<i>Mohamed Elhoseiny, Tarek El-Gaaly, Amr Bakry, Ahmed Elgammal</i>	
<b>BASC: APPLYING BAYESIAN OPTIMIZATION TO THE SEARCH FOR GLOBAL MINIMA ON POTENTIAL ENERGY SURFACES</b> .....	1423
<i>Shane Carr, Roman Garnett, Cynthia Lo</i>	
<b>ON THE ITERATION COMPLEXITY OF OBLIVIOUS FIRST-ORDER OPTIMIZATION ALGORITHMS</b> .....	1433
<i>Yossi Arjevani, Ohad Shamir</i>	
<b>STOCHASTIC VARIANCE REDUCED OPTIMIZATION FOR NONCONVEX SPARSE LEARNING</b> .....	1448
<i>Xingguo Li, Tuo Zhao, Raman Arora, Han Liu, Jarvis Haupt</i>	
<b>ANALYSIS OF VARIATIONAL BAYESIAN FACTORIZATIONS FOR SPARSE AND LOW-RANK ESTIMATION</b> .....	1461
<i>David Wipf</i>	
<b>FAST K-MEANS WITH ACCURATE BOUNDS</b> .....	1471
<i>James Newling, Francois Fleuret</i>	
<b>BOOLEAN MATRIX FACTORIZATION AND NOISY COMPLETION VIA MESSAGE PASSING</b> .....	1486
<i>Siamak Ravanbakhsh, Barnabas Poczos, Russell Greiner</i>	

<b>CONVOLUTIONAL RECTIFIER NETWORKS AS GENERALIZED TENSOR DECOMPOSITIONS</b> .....	1500
<i>Nadav Cohen, Amnon Shashua</i>	

### VOLUME 3

<b>LOW-RANK SOLUTIONS OF LINEAR MATRIX EQUATIONS VIA PROCRUSTES FLOW</b> .....	1520
<i>Stephen Tu, Ross Boczar, Max Simchowitz, Mahdi Solimanolkotabi, Ben Recht</i>	
<b>ANYTIME EXPLORATION FOR MULTI-ARMED BANDITS USING CONFIDENCE INFORMATION</b> .....	1530
<i>Kwang-Sung Jun, Robert Nowak</i>	
<b>STRUCTURED PREDICTION ENERGY NETWORKS</b> .....	1545
<i>David Belanger, Andrew McCallum</i>	
<b><math>l_1</math>-REGULARIZED NEURAL NETWORKS ARE IMPROPERLY LEARNABLE IN POLYNOMIAL TIME</b> .....	1555
<i>Yuchen Zhang, Jason D. Lee, Michael I. Jordan</i>	
<b>COMPRESSIVE SPECTRAL CLUSTERING</b> .....	1564
<i>Nicolas Tremblay, Gilles Puy, Remi Gribonval, Pierre Vandergheynst</i>	
<b>LOW-RANK TENSOR COMPLETION: A RIEMANNIAN MANIFOLD PRECONDITIONING APPROACH</b> .....	1576
<i>Hiroyuki Kasai, Bamdev Mishra</i>	
<b>PROVABLE NON-CONVEX PHASE RETRIEVAL WITH OUTLIERS: MEDIAN TRUNCATED WIRTINGER FLOW</b> .....	1607
<i>Huishuai Zhang, Yuejie Chi, Yingbin Liang</i>	
<b>ESTIMATING MAXIMUM EXPECTED VALUE THROUGH GAUSSIAN APPROXIMATION</b> .....	1628
<i>Carlo D'Eramo, Marcello Restelli, Alessandro Nuara</i>	
<b>REPRESENTATIONAL SIMILARITY LEARNING WITH APPLICATION TO BRAIN NETWORKS</b> .....	1639
<i>Urvashi Oswal, Christopher Cox, Matthew Lambon-Ralph, Timothy Rogers, Robert Nowak</i>	
<b>DROPOUT AS A BAYESIAN APPROXIMATION: REPRESENTING MODEL UNCERTAINTY IN DEEP LEARNING</b> .....	1651
<i>Yarin Gal, Zoubin Ghahramani</i>	
<b>GENERATIVE ADVERSARIAL TEXT TO IMAGE SYNTHESIS</b> .....	1681
<i>Scott Reed, Zeynep Akata, Xinchun Yan, Lajanugen Logeswaran, Bernt Schiele, Honglak Lee</i>	
<b>DIRICHLET PROCESS MIXTURE MODEL FOR CORRECTING TECHNICAL VARIATION IN SINGLE-CELL GENE EXPRESSION DATA</b> .....	1691
<i>Sandhya Prabhakaran, Elham Azizi, Ambrose Carr, Dana Pe'Er</i>	
<b>IMPROVED SVRG FOR NON-STRONGLY-CONVEX OR SUM-OF-NON-CONVEX OBJECTIVES</b> .....	1716
<i>Zeyuan Allen-Zhu, Yang Yuan</i>	
<b>SPARSE PARAMETER RECOVERY FROM AGGREGATED DATA</b> .....	1726
<i>Avrudeep Bhowmik, Joydeep Ghosh, Oluwasanmi Koyejo</i>	
<b>DEEP STRUCTURED ENERGY BASED MODELS FOR ANOMALY DETECTION</b> .....	1742
<i>Shuangfei Zhai, Yu Cheng, Weining Lu, Zhongfei Zhang</i>	
<b>MARKOV LATENT FEATURE MODELS</b> .....	1752
<i>Aonan Zhang, John Paisley</i>	
<b>THE KNOWLEDGE GRADIENT FOR SEQUENTIAL DECISION MAKING WITH STOCHASTIC BINARY FEEDBACKS</b> .....	1761
<i>Yingfei Wang, Chu Wang, Warren Powell</i>	
<b>A SIMPLE AND PROVABLE ALGORITHM FOR SPARSE DIAGONAL CCA</b> .....	1774
<i>Megasthenis Asteris, Anastasios Kyriklidis, Oluwasanmi Koyejo, Russell Poldrack</i>	
<b>QUADRATIC OPTIMIZATION WITH ORTHOGONALITY CONSTRAINTS: EXPLICIT LOJASIEWICZ EXPONENT AND LINEAR CONVERGENCE OF LINE-SEARCH METHODS</b> .....	1784
<i>Huikang Liu, Weijie Wu, Anthony Man-Cho So</i>	
<b>NORMALIZATION PROPAGATION: A PARAMETRIC TECHNIQUE FOR REMOVING INTERNAL COVARIATE SHIFT IN DEEP NETWORKS</b> .....	1800
<i>Devansh Arpit, Yingbo Zhou, Bhargava Kota, Venu Govindaraju</i>	
<b>LEARNING TO GENERATE WITH MEMORY</b> .....	1811
<i>Chongxuan Li, Jun Zhu, Bo Zhang</i>	
<b>LEARNING END-TO-END VIDEO CLASSIFICATION WITH RANK-POOLING</b> .....	1823
<i>Basura Fernando, Stephen Gould</i>	

<b>LEARNING TO FILTER WITH PREDICTIVE STATE INFERENCE MACHINES</b> .....	1833
<i>Wen Sun, Arun Venkatraman, Byron Boots, J. Andrew Bagnell</i>	
<b>A SUBSPACE LEARNING APPROACH FOR HIGH DIMENSIONAL MATRIX DECOMPOSITION WITH EFFICIENT COLUMN/ROW SAMPLING</b> .....	1846
<i>Mostafa Rahmani, Geroge Atia</i>	
<b>DCM BANDITS: LEARNING TO RANK WITH MULTIPLE CLICKS</b> .....	1855
<i>Sumeet Katariya, Branislav Kveton, Csaba Szepesvari, Zheng Wen</i>	
<b>TRAIN FASTER, GENERALIZE BETTER: STABILITY OF STOCHASTIC GRADIENT DESCENT</b> .....	1868
<i>Moritz Hardt, Ben Recht, Yoram Singer</i>	
<b>COPELAND DUELING BANDIT PROBLEM: REGRET LOWER BOUND, OPTIMAL ALGORITHM, AND COMPUTATIONALLY EFFICIENT ALGORITHM</b> .....	1878
<i>Junpei Komiyama, Junya Honda, Hiroshi Nakagawa</i>	
<b>CONTEXTUAL COMBINATORIAL CASCADING BANDITS</b> .....	1900
<i>Shuai Li, Baoxiang Wang, Shengyu Zhang, Wei Chen</i>	
<b>CONSERVATIVE BANDITS</b> .....	1917
<i>Yifan Wu, Roshan Shariff, Tor Lattimore, Csaba Szepesvari</i>	
<b>VARIANCE-REDUCED AND PROJECTION-FREE STOCHASTIC OPTIMIZATION</b> .....	1926
<i>Elad Hazan, Haipeng Luo</i>	
<b>FACTORED TEMPORAL SIGMOID BELIEF NETWORKS FOR SEQUENCE LEARNING</b> .....	1937
<i>Jiaming Song, Zhe Gan, Lawrence Carin</i>	
<b>FALSE DISCOVERY RATE CONTROL AND STATISTICAL QUALITY ASSESSMENT OF ANNOTATORS IN CROWDSOURCED RANKING</b> .....	1947
<i>Qianqian Xu, Jiechao Xiong, Xiaochun Cao, Yuan Yao</i>	
<b>STRONGLY-TYPED RECURRENT NEURAL NETWORKS</b> .....	1957
<i>David Balduzzi, Muhammad Ghifary</i>	
<b>DISTRIBUTED CLUSTERING OF LINEAR BANDITS IN PEER TO PEER NETWORKS</b> .....	1966
<i>Nathan Korda, Balazs Szorenyi, Shuai Li</i>	
<b>COLLAPSED VARIATIONAL INFERENCE FOR SUM-PRODUCT NETWORKS</b> .....	1981
<i>Han Zhao, Tameem Adel, Geoff Gordon, Brandon Amos</i>	
<b>ON THE ANALYSIS OF COMPLEX BACKUP STRATEGIES IN MONTE CARLO TREE SEARCH</b> .....	1991
<i>Piyush Khandelwal, Elad Liebman, Scott, Peter Stone</i>	
<b>BENCHMARKING DEEP REINFORCEMENT LEARNING FOR CONTINUOUS CONTROL</b> .....	2001
<i>Yan Duan, Xi Chen, Rein Houthoof, John Schulman, Pieter Abbeel</i>	
<b>K-MEANS CLUSTERING WITH DISTRIBUTED DIMENSIONS</b> .....	2015
<i>Hu Ding, Yu Liu, Lingxiao Huang, Jian Li</i>	
<b>TEXTURE NETWORKS: FEED-FORWARD SYNTHESIS OF TEXTURES AND STYLIZED IMAGES</b> .....	2027
<i>Dmitry Ulyanov, Vadim Lebedev, Andrea, Victor Lempitsky</i>	
<b>FAST CONSTRAINED SUBMODULAR MAXIMIZATION: PERSONALIZED DATA SUMMARIZATION</b> .....	2042
<i>Baharan Mirzasoleiman, Ashwinkumar Badanidiyuru, Amin Karbasi</i>	
<b>ON THE STATISTICAL LIMITS OF CONVEX RELAXATIONS</b> .....	2055
<i>Zhaoran Wang, Quanquan Gu, Han Liu</i>	
<b>ASK ME ANYTHING: DYNAMIC MEMORY NETWORKS FOR NATURAL LANGUAGE PROCESSING</b> .....	2068
<i>Ankit Kumar, Ozan Irsoy, Peter Ondruska, Mohit Iyyer, James Bradbury, Ishaan Gulrajani, Victor Zhong, Romain Paulus, Richard Socher</i>	
<b>GOSSIP DUAL AVERAGING FOR DECENTRALIZED OPTIMIZATION OF PAIRWISE FUNCTIONS</b> .....	2079
<i>Igor Colin, Aurelien Bellet, Joseph Salmon, Stéphan Cléménçon</i>	
<b>SOLVING RIDGE REGRESSION USING SKETCHED PRECONDITIONED SVRG</b> .....	2102
<i>Alon Gonen, Francesco Orabona, Shai Shalev-Shwartz</i>	
<b>CUMULATIVE PROSPECT THEORY MEETS REINFORCEMENT LEARNING: PREDICTION AND CONTROL</b> .....	2112
<i>L. A. Prashanth, Cheng Jie, Michael Fu, Steve Marcus, Csaba Szepesvari</i>	
<b>ESTIMATING ACCURACY FROM UNLABELED DATA: A BAYESIAN APPROACH</b> .....	2122
<i>Emmanouil Antonios Platanios, Avinava Dubey, Tom Mitchell</i>	
<b>NON-NEGATIVE MATRIX FACTORIZATION UNDER HEAVY NOISE</b> .....	2134
<i>Chiranjib Bhattacharya, Navin Goyal, Ravindran Kannan, Jagdeep Pani</i>	



<b>EXTREME F-MEASURE MAXIMIZATION USING SPARSE PROBABILITY ESTIMATES</b> .....	2143
<i>Kalina Jasinska, Krzysztof Dembczynski, Robert Busa-Fekete, Karlson Pfannschmidt, Timo Klerx, Eyke Hullermeier</i>	
<b>IMPORTANCE SAMPLING TREE FOR LARGE-SCALE EMPIRICAL EXPECTATION</b> .....	2157
<i>Olivier Canevet, Cijo Jose, Francois Fleuret</i>	
<b>STARTING SMALL - LEARNING WITH ADAPTIVE SAMPLE SIZES</b> .....	2167
<i>Hadi Daneshmand, Aurelien Lucchi, Thomas Hofmann</i>	
<b>DEEP GAUSSIAN PROCESSES FOR REGRESSION USING APPROXIMATE EXPECTATION PROPAGATION</b> .....	2187
<i>Thang Bui, Daniel Hernandez-Lobato, Jose Miguel Hernandez-Lobato, Yingzhen Li, Richard Turner</i>	
<b>DR-ABC: APPROXIMATE BAYESIAN COMPUTATION WITH KERNEL-BASED DISTRIBUTION REGRESSION</b> .....	2209
<i>Jovana Mitrovic, Dino Sejdinovic, Yee-Whye Teh</i>	
<b>PREDICTIVE ENTROPY SEARCH FOR MULTI-OBJECTIVE BAYESIAN OPTIMIZATION</b> .....	2219
<i>Daniel Hernandez-Lobato, Jose Miguel Hernandez-Lobato, Amar Shah, Ryan Adams</i>	
<b>RICH COMPONENT ANALYSIS</b> .....	2238
<i>Rong Ge, James Zou</i>	

## VOLUME 4

<b>BLACK-BOX ALPHA DIVERGENCE MINIMIZATION</b> .....	2256
<i>Jose Miguel Hernandez-Lobato, Yingzhen Li, Mark Rowland, Thang Bui, Daniel Hernandez-Lobato, Richard Turner</i>	
<b>ONE-SHOT GENERALIZATION IN DEEP GENERATIVE MODELS</b> .....	2274
<i>Danilo Rezende, Shakir, Ivo Danihelka, Karol Gregor, Daan Wierstra</i>	
<b>OPTIMAL CLASSIFICATION WITH MULTIVARIATE LOSSES</b> .....	2283
<i>Nagarajan Natarajan, Oluwasanmi Koyejo, Pradeep Ravikumar, Inderjit Dhillon</i>	
<b>A RANKING APPROACH TO GLOBAL OPTIMIZATION</b> .....	2296
<i>Cedric Malherbe, Emile Contal, Nicolas Vayatis</i>	
<b>PARALLEL AND DISTRIBUTED BLOCK-COORDINATE FRANK-WOLFE ALGORITHMS</b> .....	2317
<i>Yu-Xiang Wang, Veeranjaneyulu Sadhanala, Wei Dai, Willie Neiswanger, Suvrit Sra, Eric Xing</i>	
<b>AUTOENCODING BEYOND PIXELS USING A LEARNED SIMILARITY METRIC</b> .....	2341
<i>Anders Boesen Lindbo Larsen, Søren Kaae Sønderby, Hugo Larochelle, Ole Winther</i>	
<b>ENSURING RAPID MIXING AND LOW BIAS FOR ASYNCHRONOUS GIBBS SAMPLING</b> .....	2350
<i>Christopher De Sa, Chris Re, Kunle Olukotun</i>	
<b>SIMULTANEOUS SAFE SCREENING OF FEATURES AND SAMPLES IN DOUBLY SPARSE MODELING</b> .....	2381
<i>Atsushi Shibagaki, Masayuki Karasuyama, Kohei Hatano, Ichiro Takeuchi</i>	
<b>ANYTIME OPTIMAL ALGORITHMS IN STOCHASTIC MULTI-ARMED BANDITS</b> .....	2391
<i>Rémy Degenne, Vianney Perchet</i>	
<b>BOUNDED OFF-POLICY EVALUATION WITH MISSING DATA FOR COURSE RECOMMENDATION AND CURRICULUM DESIGN</b> .....	2410
<i>William Hoiles, Mihaela Van Der Schaar</i>	
<b>ON COLLAPSED REPRESENTATION OF HIERARCHICAL COMPLETELY RANDOM MEASURES</b> .....	2420
<i>Gaurav Pandey, Ambedkar Dukkipati</i>	
<b>FROM SOFTMAX TO SPARSEMAX: A SPARSE MODEL OF ATTENTION AND MULTI-LABEL CLASSIFICATION</b> .....	2432
<i>Andre Martins, Ramon Astudillo</i>	
<b>BLACK-BOX OPTIMIZATION WITH A POLITICIAN</b> .....	2444
<i>Sebastien Bubeck, Yin Tat Lee</i>	
<b>GAUSSIAN PROCESS NONPARAMETRIC TENSOR ESTIMATOR AND ITS MINIMAX OPTIMALITY</b> .....	2452
<i>Heishiro Kanagawa, Taiji Suzuki, Hayato Kobayashi, Nobuyuki Shimizu, Yukihiko Tagami</i>	
<b>NO-REGRET ALGORITHMS FOR HEAVY-TAILED LINEAR BANDITS</b> .....	2474
<i>Andres Munoz Medina, Scott Yang</i>	
<b>EXTENDED AND UNSCENTED KITCHEN SINKS</b> .....	2483
<i>Edwin Bonilla, Daniel Steinberg, Alistair Reid</i>	
<b>MATRIX EIGEN-DECOMPOSITION VIA DOUBLY STOCHASTIC RIEMANNIAN OPTIMIZATION</b> .....	2496
<i>Zhiqiang Xu, Peilin Zhao, Jianneng Cao, Xiaoli Li</i>	

<b>RECOMMENDATIONS AS TREATMENTS: DEBIASING LEARNING AND EVALUATION</b> .....	2512
<i>Tobias Schnabel, Adith Swaminathan, Ashudeep Singh, Navin Chandak, Thorsten Joachims</i>	
<b>FORECASTICU: A PROGNOSTIC DECISION SUPPORT SYSTEM FOR TIMELY PREDICTION OF INTENSIVE CARE UNIT ADMISSION</b> .....	2524
<i>Jinsung Yoon, Ahmed Alaa, Scott Hu, Mihaela Van Der Schaar</i>	
<b>AN OPTIMAL ALGORITHM FOR THE THRESHOLDING BANDIT PROBLEM</b> .....	2539
<i>Andrea Locatelli, Maurilio Gutzeit, Alexandra Carpentier</i>	
<b>FAST PARAMETER INFERENCE IN NONLINEAR DYNAMICAL SYSTEMS USING ITERATIVE GRADIENT MATCHING</b> .....	2555
<i>Mu Niu, Simon Rogers, Maurizio Filippone, Dirk Husmeier</i>	
<b>STRUCTURED AND EFFICIENT VARIATIONAL DEEP LEARNING WITH MATRIX GAUSSIAN POSTERiors</b> .....	2564
<i>Christos Louizos, Max Welling</i>	
<b>LEARNING GRANGER CAUSALITY FOR HAWKES PROCESSES</b> .....	2576
<i>Hongteng Xu, Mehrdad Farajtabar, Hongyuan Zha</i>	
<b>NEURAL VARIATIONAL INFERENCE FOR TEXT PROCESSING</b> .....	2589
<i>Yishu Miao, Lei Yu, Phil Blunsom</i>	
<b>DICTIONARY LEARNING FOR MASSIVE MATRIX FACTORIZATION</b> .....	2601
<i>Arthur Mensch, Julien Mairal, Bertrand Thirion, Gael Varoquaux</i>	
<b>PIXEL RECURRENT NEURAL NETWORKS</b> .....	2611
<i>Aaron Van Den Oord, Nal Kalchbrenner, Koray Kavukcuoglu</i>	
<b>WHY MOST DECISIONS ARE EASY IN TETRIS—AND PERHAPS IN OTHER SEQUENTIAL DECISION PROBLEMS, AS WELL</b> .....	2621
<i>Ozgur Simsek, Simon Algorta, Amit Kothiyal</i>	
<b>GAUSSIAN QUADRATURE FOR MATRIX INVERSE FORMS WITH APPLICATIONS</b> .....	2630
<i>Chengtao Li, Suvrit Sra, Stefanie Jegelka</i>	
<b>TRAIN AND TEST TIGHTNESS OF LP RELAXATIONS IN STRUCTURED PREDICTION</b> .....	2652
<i>Ofer Meshi, Mehrdad Mahdavi, Adrian Weller, David Sontag</i>	
<b>HIERARCHICAL COMPOUND POISSON FACTORIZATION</b> .....	2665
<i>Mehmet Basbug, Barbara Engelhardt</i>	
<b>OPPONENT MODELING IN DEEP REINFORCEMENT LEARNING</b> .....	2675
<i>He He, Jordan Boyd-Graber, Kevin Kwok, Hal Daumé III</i>	
<b>NO PENALTY NO TEARS: LEAST SQUARES IN HIGH-DIMENSIONAL LINEAR MODELS</b> .....	2685
<i>Xiangyu Wang, David Dunson, Chenlei Leng</i>	
<b>SDNA: STOCHASTIC DUAL NEWTON ASCENT FOR EMPIRICAL RISK MINIMIZATION</b> .....	2707
<i>Zheng Qu, Peter Richtarik, Martin Takac, Olivier Fercoq</i>	
<b>ON GRADUATED OPTIMIZATION FOR STOCHASTIC NON-CONVEX PROBLEMS</b> .....	2726
<i>Elad Hazan, Kfir Yehuda Levy, Shai Shalev-Shwartz</i>	
<b>META-LEARNING WITH MEMORY-AUGMENTED NEURAL NETWORKS</b> .....	2740
<i>Adam Santoro, Sergey Bartunov, Matthew Botvinick, Daan Wierstra, Timothy Lillicrap</i>	
<b>THE KNOCKOFF FILTER FOR FDR CONTROL IN GROUP-SPARSE AND MULTITASK REGRESSION</b> .....	2752
<i>Ran Dai, Rina Barber</i>	
<b>SOFTENED APPROXIMATE POLICY ITERATION FOR MARKOV GAMES</b> .....	2761
<i>Julien Pérolat, Bilal Piot, Matthieu Geist, Bruno Scherrer, Olivier Pietquin</i>	
<b>STOCHASTIC BLOCK BFGS: SQUEEZING MORE CURVATURE OUT OF DATA</b> .....	2774
<i>Robert Gower, Donald Goldfarb, Peter Richtarik</i>	
<b>DIFFERENTIAL GEOMETRIC REGULARIZATION FOR SUPERVISED LEARNING OF CLASSIFIERS</b> .....	2784
<i>Qinxun Bai, Steven Rosenberg, Zheng Wu, Stan Sclaroff</i>	
<b>EXPLOITING CYCLIC SYMMETRY IN CONVOLUTIONAL NEURAL NETWORKS</b> .....	2799
<i>Sander Dieleman, Jeffrey De Fauw, Koray Kavukcuoglu</i>	
<b>GRAYING THE BLACK BOX: UNDERSTANDING DQNS</b> .....	2809
<i>Tom Zahavy, Nir Ben-Zrihem, Shie Mannor</i>	
<b>THE SUM-PRODUCT THEOREM: A FOUNDATION FOR LEARNING TRACTABLE MODELS</b> .....	2823
<i>Abram Friesen, Pedro Domingos</i>	
<b>PARETO FRONTIER LEARNING WITH EXPENSIVE CORRELATED OBJECTIVES</b> .....	2839
<i>Amar Shah, Zoubin Ghahramani</i>	
<b>ASYNCHRONOUS METHODS FOR DEEP REINFORCEMENT LEARNING</b> .....	2850
<i>Volodymyr Mnih, Adria Puigdomenech Badia, Mehdi Mirza, Alex Graves, Timothy Lillicrap, Tim Harley, David Silver, Koray Kavukcuoglu</i>	

<b>A SIMPLE AND STRONGLY-LOCAL FLOW-BASED METHOD FOR CUT IMPROVEMENT</b> .....	2870
<i>Nate Veldt, David Gleich, Michael Mahoney</i>	
<b>NONLINEAR STATISTICAL LEARNING WITH TRUNCATED GAUSSIAN GRAPHICAL MODELS</b> .....	2884
<i>Qinliang Su, Xuejun Liao, Changyou Chen, Lawrence Carin</i>	
<b>BARRON AND COVER'S THEORY IN SUPERVISED LEARNING AND ITS APPLICATION TO LASSO</b> .....	2896
<i>Masanori Kawakita, Jun'Ichi Takeuchi</i>	
<b>NONPARAMETRIC CANONICAL CORRELATION ANALYSIS</b> .....	2906
<i>Tomer Michaeli, Weiran Wang, Karen Livescu</i>	
<b>BISTRO: AN EFFICIENT RELAXATION-BASED METHOD FOR CONTEXTUAL BANDITS</b> .....	2917
<i>Alexander Rakhlin, Karthik Sridharan</i>	
<b>ASSOCIATIVE LONG SHORT-TERM MEMORY</b> .....	2929
<i>Ivo Danihelka, Greg Wayne, Benigno Uria, Nal Kalchbrenner, Alex Graves</i>	
<b>DUELING NETWORK ARCHITECTURES FOR DEEP REINFORCEMENT LEARNING</b> .....	2939
<i>Ziyu Wang, Tom Schaul, Matteo Hessel, Hado Van Hasselt, Marc Lanctot, Nando De Freitas</i>	
<b>PERSISTENCE WEIGHTED GAUSSIAN KERNEL FOR TOPOLOGICAL DATA ANALYSIS</b> .....	2948
<i>Genki Kusano, Yasuaki Hiraoka, Kenji Fukumizu</i>	
<b>LEARNING CONVOLUTIONAL NEURAL NETWORKS FOR GRAPHS</b> .....	2958
<i>Mathias Niepert, Mohamed Ahmed, Konstantin Kutzkov</i>	
<b>PERSISTENT RNNs: STASHING RECURRENT WEIGHTS ON-CHIP</b> .....	2968
<i>Greg Diamos, Shubho Sengupta, Bryan Catanzaro, Mike Chrzanowski, Adam Coates, Erich Elsen, Jesse Engel, Awni Hannun, Sanjeev Satheesh</i>	

## VOLUME 5

<b>RECURRENT ORTHOGONAL NETWORKS AND LONG-MEMORY TASKS</b> .....	2978
<i>Mikael Henaff, Arthur Szlam, Yann Lecun</i>	
<b>THE ARROW OF TIME IN MULTIVARIATE TIME SERIES</b> .....	2987
<i>Stefan Bauer, Bernhard Schölkopf, Jonas Peters</i>	
<b>MIXTURE PROPORTION ESTIMATION VIA KERNEL EMBEDDINGS OF DISTRIBUTIONS</b> .....	2996
<i>Harish Ramaswamy, Clayton Scott, Ambuj Tewari</i>	
<b>FAST DPP SAMPLING FOR NYSTROM WITH APPLICATION TO KERNEL METHODS</b> .....	3005
<i>Chengtao Li, Stefanie Jegelka, Suvrit Sra</i>	
<b>COMPLEX EMBEDDINGS FOR SIMPLE LINK PREDICTION</b> .....	3021
<i>Théo Trouillon, Johannes Welbl, Sebastian Riedel, Eric Gaussier, Guillaume Bouchard</i>	
<b>INTERACTIVE BAYESIAN HIERARCHICAL CLUSTERING</b> .....	3033
<i>Sharad Vikram, Sanjoy Dasgupta</i>	
<b>A CONVOLUTIONAL ATTENTION NETWORK FOR EXTREME SUMMARIZATION OF SOURCE CODE</b> .....	3043
<i>Miltiadis Allamanis, Hao Peng, Charles Sutton</i>	
<b>HOW TO FAKE MULTIPLY BY A GAUSSIAN MATRIX</b> .....	3053
<i>Michael Kapralov, Vamsi Potluru, David Woodruff</i>	
<b>DIFFERENTIALLY PRIVATE CHI-SQUARED HYPOTHESIS TESTING: GOODNESS OF FIT AND INDEPENDENCE TESTING</b> .....	3086
<i>Ryan Rogers, Salil Vadhan, Hyun Lim, Marco Gaboardi</i>	
<b>PLIABLE REJECTION SAMPLING</b> .....	3122
<i>Akram Erraqabi, Michal Valko, Alexandra Carpentier, Odalric Maillard</i>	
<b>DIFFERENTIALLY PRIVATE POLICY EVALUATION</b> .....	3138
<i>Borja Balle, Maziar Gomrokchi, Doina Precup</i>	
<b>DATA-EFFICIENT OFF-POLICY POLICY EVALUATION FOR REINFORCEMENT LEARNING</b> .....	3158
<i>Philip Thomas, Emma Brunskill</i>	
<b>DISCRETE DEEP FEATURE EXTRACTION: A THEORY AND NEW ARCHITECTURES</b> .....	3168
<i>Thomas Wiatowski, Michael Tschannen, Aleksandar Stanic, Philipp Grohs, Helmut Boelskei</i>	
<b>EFFICIENT ALGORITHMS FOR ADVERSARIAL CONTEXTUAL LEARNING</b> .....	3184
<i>Vasilis Syrgkanis, Akshay Krishnamurthy, Robert Schapire</i>	
<b>TRAINING DEEP NEURAL NETWORKS VIA DIRECT LOSS MINIMIZATION</b> .....	3207
<i>Yang Song, Alexander Schwing, Richard, Raquel Urtasun</i>	
<b>SEQUENCE TO SEQUENCE TRAINING OF CTC-RNNs WITH PARTIAL WINDOWING</b> .....	3222
<i>Kyuyeon Hwang, Wonyong Sung</i>	

<b>VARIATIONAL INFERENCE FOR MONTE CARLO OBJECTIVES</b> .....	3237
<i>Andriy Mnih, Danilo Rezende</i>	
<b>HIERARCHICAL DECISION MAKING IN ELECTRICITY GRID MANAGEMENT</b> .....	3249
<i>Gal Dalal, Elad Gilboa, Shie Mannor</i>	
<b>LEARNING SPARSE COMBINATORIAL REPRESENTATIONS VIA TWO-STAGE SUBMODULAR MAXIMIZATION</b> .....	3259
<i>Eric Balkanski, Baharan Mirzasoleiman, Andreas Krause, Yaron Singer</i>	
<b>UNDERSTANDING AND IMPROVING CONVOLUTIONAL NEURAL NETWORKS VIA CONCATENATED RECTIFIED LINEAR UNITS</b> .....	3276
<i>Wenling Shang, Kihyuk Sohn, Diogo Almeida, Honglak Lee</i>	
<b>ISOTONIC HAWKES PROCESSES</b> .....	3285
<i>Yichen Wang, Bo Xie, Nan Du, Le Song</i>	
<b>CROSS-GRAPH LEARNING OF MULTI-RELATIONAL ASSOCIATIONS</b> .....	3302
<i>Hanxiao Liu, Yiming Yang</i>	
<b>MARKOV-MODULATED MARKED POISSON PROCESSES FOR CHECK-IN DATA</b> .....	3311
<i>Jiangwei Pan, Vinayak Rao, Pankaj Agarwal, Alan Gelfand</i>	
<b>BEYOND PARITY CONSTRAINTS: FOURIER ANALYSIS OF HASH FUNCTIONS FOR INFERENCE</b> .....	3321
<i>Tudor Achim, Ashish Sabharwal, Stefano Ermon</i>	
<b>ON THE POWER AND LIMITS OF DISTANCE-BASED LEARNING</b> .....	3332
<i>Periklis Papakonstantinou, Jia Xu, Guang Yang</i>	
<b>A CONVEX ATOMIC-NORM APPROACH TO MULTIPLE SEQUENCE ALIGNMENT AND MOTIF DISCOVERY</b> .....	3369
<i>Ian En-Hsu Yen, Xin Lin, Jiong Zhang, Pradeep Ravikumar, Inderjit Dhillon</i>	
<b>GENERALIZED DIRECT CHANGE ESTIMATION IN ISING MODEL STRUCTURE</b> .....	3381
<i>Farideh Fazayeli, Arindam Banerjee</i>	
<b>ROBUST PRINCIPAL COMPONENT ANALYSIS WITH SIDE INFORMATION</b> .....	3391
<i>Kai-Yang Chiang, Cho-Jui Hsieh, Inderjit Dhillon</i>	
<b>TOWARDS FASTER RATES AND ORACLE PROPERTY FOR LOW-RANK MATRIX ESTIMATION</b> .....	3405
<i>Huan Gui, Jiawei Han, Quanquan Gu</i>	
<b>EARLY AND RELIABLE EVENT DETECTION USING PROXIMITY SPACE REPRESENTATION</b> .....	3432
<i>Maxime Sangnier, Jerome Gauthier, Alain Rakotomamonjy</i>	
<b>STRATIFIED SAMPLING MEETS MACHINE LEARNING</b> .....	3442
<i>Edo Liberty, Kevin Lang, Konstantin Shmakov</i>	
<b>EFFICIENT MULTI-INSTANCE LEARNING FOR ACTIVITY RECOGNITION FROM TIME SERIES DATA USING AN AUTO-REGRESSIVE HIDDEN MARKOV MODEL</b> .....	3452
<i>Xinze Guan, Raviv Raich, Weng-Keen Wong</i>	
<b>GENERALIZATION PROPERTIES AND IMPLICIT REGULARIZATION FOR MULTIPLE PASSES SGM</b> .....	3474
<i>Junhong Lin, Raffaello Camoriano, Lorenzo Rosasco</i>	
<b>PRINCIPAL COMPONENT PROJECTION WITHOUT PRINCIPAL COMPONENT ANALYSIS</b> .....	3495
<i>Roy Frostig, Cameron Musco, Christopher Musco, Aaron Sidford</i>	
<b>RECOVERY GUARANTEE OF WEIGHTED LOW-RANK APPROXIMATION VIA ALTERNATING MINIMIZATION</b> .....	3504
<i>Yuanzhi Li, Yingyu Liang, Andrej Risteski</i>	
<b>DECONSTRUCTING THE LADDER NETWORK ARCHITECTURE</b> .....	3527
<i>Mohammad Pezeshki, Linxi Fan, Philemon Brakel, Aaron Courville, Yoshua Bengio</i>	
<b>GENERALIZATION AND EXPLORATION VIA RANDOMIZED VALUE FUNCTIONS</b> .....	3540
<i>Ian Osband, Benjamin Van Roy, Zheng Wen</i>	
<b>EVASION AND HARDENING OF TREE ENSEMBLE CLASSIFIERS</b> .....	3562
<i>Alex Kanchelian, J. D. Tygar, Anthony Joseph</i>	
<b>DYNAMIC MEMORY NETWORKS FOR VISUAL AND TEXTUAL QUESTION ANSWERING</b> .....	3574
<i>Caiming Xiong, Stephen Merity, Richard Socher</i>	
<b>ESTIMATING COSMOLOGICAL PARAMETERS FROM THE DARK MATTER DISTRIBUTION</b> .....	3584
<i>Siamak Ravanbakhsh, Junier Oliva, Sebastian Fromenteau, Layne Price, Shirley Ho, Jeff Schneider, Barnabas Poczos</i>	
<b>EXPRESSIVENESS OF RECTIFIER NETWORKS</b> .....	3595
<i>Xingyuan Pan, Vivek Srikumar</i>	

<b>DISCRETE DISTRIBUTION ESTIMATION UNDER LOCAL PRIVACY</b> .....	3607
<i>Peter Kairouz, Keith Bonawitz, Daniel Ramage</i>	
<b>SQUARE ROOT GRAPHICAL MODELS: MULTIVARIATE GENERALIZATIONS OF UNIVARIATE EXPONENTIAL FAMILIES THAT PERMIT POSITIVE DEPENDENCIES</b> .....	3634
<i>David Inouye, Pradeep Ravikumar, Inderjit Dhillon</i>	
<b>A BOX-CONSTRAINED APPROACH FOR HARD PERMUTATION PROBLEMS</b> .....	3645
<i>Cong Han Lim, Steve Wright</i>	
<b>GEOMETRIC MEAN METRIC LEARNING</b> .....	3660
<i>Pourya Zadeh, Reshad Hosseini, Suvrit Sra</i>	
<b>SPARSE NONLINEAR REGRESSION: PARAMETER ESTIMATION UNDER NONCONVEXITY</b> .....	3668
<i>Zhuoran Yang, Zhaoran Wang, Han Liu, Yonina Eldar, Tong Zhang</i>	
<b>CONDITIONAL BERNOULLI MIXTURES FOR MULTI-LABEL CLASSIFICATION</b> .....	3678
<i>Cheng Li, Bingyu Wang, Virgil Pavlu, Javed Aslam</i>	
<b>SCALABLE DISCRETE SAMPLING AS A MULTI-ARMED BANDIT PROBLEM</b> .....	3691
<i>Yutian Chen, Zoubin Ghahramani</i>	
<b>RECYCLING RANDOMNESS WITH STRUCTURE FOR SUBLINEAR TIME KERNEL EXPANSIONS</b> .....	3708
<i>Krzysztof Choromanski, Vikas Sindhwani</i>	

## VOLUME 6

<b>BIDIRECTIONAL HELMHOLTZ MACHINES</b> .....	3725
<i>Jorg Bornschein, Samira Shabanian, Asja Fischer, Yoshua Bengio</i>	
<b>FASTER CONVEX OPTIMIZATION: SIMULATED ANNEALING WITH AN EFFICIENT UNIVERSAL BARRIER</b> .....	3734
<i>Jacob Abernethy, Elad Hazan</i>	
<b>PRECONDITIONING KERNEL MATRICES</b> .....	3747
<i>Kurt Cutajar, Michael Osborne, John Cunningham, Maurizio Filippone</i>	
<b>GREEDY COLUMN SUBSET SELECTION: NEW BOUNDS AND DISTRIBUTED ALGORITHMS</b> .....	3761
<i>Jason Altschuler, Aditya Bhaskara, Gang Fu, Vahab Mirrokni, Afshin Rostamizadeh, Morteza Zadimoghaddam</i>	
<b>DYNAMIC CAPACITY NETWORKS</b> .....	3775
<i>Amjad Almahairi, Nicolas Ballas, Tim Cooijmans, Yin Zheng, Hugo Larochelle, Aaron Courville</i>	
<b>PRICING A LOW-REGRET SELLER</b> .....	3785
<i>Hoda Heidari, Mohammad Mahdian, Umar Syed, Sergei Vassilvitskii, Sadra Yazdanbod</i>	
<b>ESTIMATION FROM INDIRECT SUPERVISION WITH LINEAR MOMENTS</b> .....	3794
<i>Aditi Raghunathan, Roy Frostig, John Duchi, Percy Liang</i>	
<b>SPEEDING UP K-MEANS BY APPROXIMATING EUCLIDEAN DISTANCES VIA BLOCK VECTORS</b> .....	3805
<i>Thomas Bottesch, Thomas Bühler, Markus Kächele</i>	
<b>LEARNING AND INFERENCE VIA MAXIMUM INNER PRODUCT SEARCH</b> .....	3814
<i>Stephen Mussmann, Stefano Ermon</i>	
<b>A SUPERLINEARLY-CONVERGENT PROXIMAL NEWTON-TYPE METHOD FOR THE OPTIMIZATION OF FINITE SUMS</b> .....	3827
<i>Anton Rodomanov, Dmitry Kropotov</i>	
<b>A KERNEL TEST OF GOODNESS OF FIT</b> .....	3854
<i>Kacper Chwialkowski, Heiko Strathmann, Arthur Gretton</i>	
<b>INTERACTING PARTICLE MARKOV CHAIN MONTE CARLO</b> .....	3868
<i>Tom Rainforth, Christian Naesseth, Fredrik Lindsten, Brooks Paige, Jan-Willem Vandemeent, Arnaud Doucet, Frank Wood</i>	
<b>FASTER EIGENVECTOR COMPUTATION VIA SHIFT-AND-INVERT PRECONDITIONING</b> .....	3886
<i>Dan Garber, Elad Hazan, Chi Jin, Sham, Cameron Musco, Praneeth Netrapalli, Aaron Sidford</i>	
<b>A THEORY OF GENERATIVE CONVNET</b> .....	3895
<i>Jianwen Xie, Yang Lu, Song-Chun Zhu, Yingnian Wu</i>	
<b>EFFICIENT LEARNING WITH A FAMILY OF NONCONVEX REGULARIZERS BY REDISTRIBUTING NONCONVEXITY</b> .....	3905
<i>Quanming Yao, James Kwok</i>	
<b>COMPUTATIONALLY EFFICIENT NYSTRÖM APPROXIMATION USING FAST TRANSFORMS</b> .....	3915
<i>Si Si, Cho-Jui Hsieh, Inderjit Dhillon</i>	
<b>GROMOV-WASSERSTEIN AVERAGING OF KERNEL AND DISTANCE MATRICES</b> .....	3927
<i>Gabriel Peyré, Marco Cuturi, Justin Solomon</i>	

<b>ROBUST MONTE CARLO SAMPLING USING RIEMANNIAN NOSÉ-POINCARÉ HAMILTONIAN DYNAMICS</b> .....	3936
<i>Anirban Roychowdhury, Brian Kulis, Srinivasan Parthasarathy</i>	
<b>THE SEGMENTED IHMM: A SIMPLE, EFFICIENT HIERARCHICAL INFINITE HMM</b> .....	3949
<i>Ardavan Saeedi, Matthew Hoffman, Matthew Johnson, Ryan Adams</i>	
<b>META-GRADIENT BOOSTED DECISION TREE MODEL FOR WEIGHT AND TARGET LEARNING</b> .....	3960
<i>Yury Ustinovskiy, Valentina Fedorova, Gleb Gusev, Pavel Serdyukov</i>	
<b>DISCRIMINATIVE EMBEDDINGS OF LATENT VARIABLE MODELS FOR STRUCTURED DATA</b> .....	3970
<i>Hanjun Dai, Bo Dai, Le Song</i>	
<b>ROBUST RANDOM CUT FOREST BASED ANOMALY DETECTION ON STREAMS</b> .....	3987
<i>Sudipto Guha, Nina Mishra, Gourav Roy, Okke Schrijvers</i>	
<b>CLUSTERING HIGH DIMENSIONAL CATEGORICAL DATA VIA TOPOGRAPHICAL FEATURES</b> .....	4000
<i>Chao Chen, Novi Quadrianto</i>	
<b>EFFICIENT ALGORITHMS FOR LARGE-SCALE GENERALIZED EIGENVECTOR COMPUTATION AND CANONICAL CORRELATION ANALYSIS</b> .....	4009
<i>Rong Ge, Chi Jin, Sham, Praneeth Netrapalli, Aaron Sidford</i>	
<b>ALGORITHMS FOR OPTIMIZING THE RATIO OF SUBMODULAR FUNCTIONS</b> .....	4027
<i>Wenruo Bai, Rishabh Iyer, Kai Wei, Jeff Bilmes</i>	
<b>MODEL-FREE IMITATION LEARNING WITH POLICY OPTIMIZATION</b> .....	4036
<i>Jonathan Ho, Jayesh Gupta, Stefano Ermon</i>	
<b>ADIOS: ARCHITECTURES DEEP IN OUTPUT SPACE</b> .....	4047
<i>Moustapha Cisse, Maruan Al-Shedivat, Samy Bengio</i>	
<b>CONDITIONAL DEPENDENCE VIA SHANNON CAPACITY: AXIOMS, ESTIMATORS AND APPLICATIONS</b> .....	4057
<i>Weihao Gao, Sreeram Kannan, Sewoong Oh, Pramod Viswanath</i>	
<b>CONTROL OF MEMORY, ACTIVE PERCEPTION, AND ACTION IN MINECRAFT</b> .....	4067
<i>Junhyuk Oh, Valliappa Chockalingam, Satinder, Honglak Lee</i>	
<b>THE LABEL COMPLEXITY OF MIXED-INITIATIVE CLASSIFIER TRAINING</b> .....	4090
<i>Jina Suh, Xiaojin Zhu, Saleema Amershi</i>	
<b>BAYESIAN POISSON TUCKER DECOMPOSITION FOR LEARNING THE STRUCTURE OF INTERNATIONAL RELATIONS</b> .....	4100
<i>Aaron Schein, Mingyuan Zhou, David Blei, Hanna Wallach</i>	
<b>TENSOR DECOMPOSITION VIA JOINT MATRIX SCHUR DECOMPOSITION</b> .....	4126
<i>Nicolo Colombo, Nikos Vlassis</i>	
<b>CONTINUOUS DEEP Q-LEARNING WITH MODEL-BASED ACCELERATION</b> .....	4135
<i>Shixiang Gu, Timothy Lillicrap, Ilya Sutskever, Sergey Levine</i>	
<b>DOMAIN ADAPTATION WITH CONDITIONAL TRANSFERABLE COMPONENTS</b> .....	4149
<i>Mingming Gong, Kun Zhang, Tongliang Liu, Dacheng Tao, Clark Glymour, Bernhard Schölkopf</i>	
<b>FIXED POINT QUANTIZATION OF DEEP CONVOLUTIONAL NETWORKS</b> .....	4166
<i>Darryl Lin, Sachin Talathi, Sreekanth Annapureddy</i>	
<b>PROVABLE ALGORITHMS FOR INFERENCE IN TOPIC MODELS</b> .....	4176
<i>Sanjeev Arora, Rong Ge, Frederic Koehler, Tengyu Ma, Ankur Moitra</i>	
<b>EPIGRAPH PROJECTIONS FOR FAST GENERAL CONVEX PROGRAMMING</b> .....	4185
<i>Po-Wei Wang, Matt Wytock, Zico Kolter</i>	
<b>FAST ALGORITHMS FOR SEGMENTED REGRESSION</b> .....	4199
<i>Jayadev Acharya, Ilias Diakonikolas, Jerry Li, Ludwig Schmidt</i>	
<b>ENERGETIC NATURAL GRADIENT DESCENT</b> .....	4235
<i>Philip Thomas, Bruno Castro Da Silva, Christoph Dann, Emma Brunskill</i>	
<b>PARTITION FUNCTIONS FROM RAO-BLACKWELLIZED TEMPERED SAMPLING</b> .....	4248
<i>David Carlson, Patrick Stinson, Ari Pakman, Liam Paninski</i>	
<b>LEARNING MIXTURES OF PLACKETT-LUCE MODELS</b> .....	4263
<i>Zhibing Zhao, Peter Piech, Lirong Xia</i>	
<b>NEAR OPTIMAL BEHAVIOR VIA APPROXIMATE STATE ABSTRACTION</b> .....	4287
<i>David Abel, David Hershkowitz, Michael Littman</i>	
<b>POWER OF ORDERED HYPOTHESIS TESTING</b> .....	4296
<i>Lihua Lei, William Fithian</i>	
<b>PHOG: PROBABILISTIC MODEL FOR CODE</b> .....	4311
<i>Pavol Bielik, Veselin Raychev, Martin Vechev</i>	

<b>SHIFTING REGRET, MIRROR DESCENT, AND MATRICES</b> .....	4324
<i>Andras Gyorgy, Csaba Szepesvari</i>	
<b>SCALABLE GRADIENT-BASED TUNING OF CONTINUOUS REGULARIZATION HYPERPARAMETERS</b> .....	4333
<i>Jelena Luketina, Tapani Raiko, Mathias Berglund, Klaus Greff</i>	
<b>MODEL-FREE TRAJECTORY OPTIMIZATION FOR REINFORCEMENT LEARNING</b> .....	4342
<i>Riad Akrouf, Gerhard Neumann, Hany Abdulsamad, Abbas Abdolmaleki</i>	
<b>CONTROLLING THE DISTANCE TO A KEMENY CONSENSUS WITHOUT COMPUTING IT</b> .....	4353
<i>Yunlong Jiao, Anna Korba, Eric Sibony</i>	
<b>HORIZONTALLY SCALABLE SUBMODULAR MAXIMIZATION</b> .....	4364
<i>Mario Lucic, Olivier Bachem, Morteza Zadimoghaddam, Andreas Krause</i>	
<b>GROUP EQUIVARIANT CONVOLUTIONAL NETWORKS</b> .....	4375
<i>Taco Cohen, Max Welling</i>	
<b>STOCHASTIC DISCRETE CLENSHAW-CURTIS QUADRATURE</b> .....	4387
<i>Nico Piatkowski, Katharina Morik</i>	
<b>CORRECTING FORECASTS WITH MULTIFACTOR NEURAL ATTENTION</b> .....	4397
<i>Matthew Riemer, Aditya Vempaty, Flavio Calmon, Fenno Heath, Richard Hull, Elham Khabiri</i>	
<b>LEARNING REPRESENTATIONS FOR COUNTERFACTUAL INFERENCE</b> .....	4407
<i>Fredrik Johansson, Uri Shalit, David Sontag</i>	
<b>AUTOMATIC CONSTRUCTION OF NONPARAMETRIC RELATIONAL REGRESSION MODELS FOR MULTIPLE TIME SERIES</b> .....	4419
<i>Yunseong Hwang, Anh Tong, Jaesik Choi</i>	
<b>INFERENCE NETWORKS FOR SEQUENTIAL MONTE CARLO IN GRAPHICAL MODELS</b> .....	4434
<i>Brooks Paige, Frank Wood</i>	
<b>SLICE SAMPLING ON HAMILTONIAN TRAJECTORIES</b> .....	4445
<i>Benjamin Bloem-Reddy, John Cunningham</i>	
<b>NOISY ACTIVATION FUNCTIONS</b> .....	4457
<i>Caglar Gulcehre, Marcin Moczulski, Misha Denil, Yoshua Bengio</i>	
<b>PD-SPARSE : A PRIMAL AND DUAL SPARSE APPROACH TO EXTREME MULTICLASS AND MULTILABEL CLASSIFICATION</b> .....	4467
<i>Ian En-Hsu Yen, Xiangru Huang, Pradeep Ravikumar, Kai Zhong, Inderjit Dhillon</i>	
<b>Author Index</b>	