

32nd International Conference on Machine Learning (ICML 2015)

Lille, France
6 – 11 July 2015

Volume 1 of 3

Editors:

Francis Bach
David Blei

ISBN: 978-1-5108-1058-7

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© (2015) 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

| | |
|--|-----|
| STOCHASTIC OPTIMIZATION WITH IMPORTANCE SAMPLING FOR REGULARIZED LOSS MINIMIZATION | 1 |
| <i>Peilin Zhao, Tong Zhang</i> | |
| APPROVAL VOTING AND INCENTIVES IN CROWDSOURCING | 10 |
| <i>Nihar Shah, Dengyong Zhou, Yuval Peres</i> | |
| A LOW VARIANCE CONSISTENT TEST OF RELATIVE DEPENDENCY | 20 |
| <i>Wacha Bounliphone, Arthur Gretton, Arthur Tenenhaus, Matthew Blaschko</i> | |
| AN ALIGNED SUBTREE KERNEL FOR WEIGHTED GRAPHS | 30 |
| <i>Lu Bai, Luca Rossi, Zhihong Zhang, Edwin Hancock</i> | |
| SPECTRAL CLUSTERING VIA THE POWER METHOD - PROVABLY | 40 |
| <i>Christos Boutsidis, Prabhanjan Kambadur, Alex Gittens</i> | |
| INFORMATION GEOMETRY AND MINIMUM DESCRIPTION LENGTH NETWORKS | 49 |
| <i>Ke Sun, Jun Wang, Alexandros Kalousis, Stephan Marchand-Maillet</i> | |
| EFFICIENT TRAINING OF LDA ON A GPU BY MEAN-FOR-MODE ESTIMATION | 59 |
| <i>Jean-Baptiste Tristan, Joseph Tassarotti, Guy Steele</i> | |
| ADAPTIVE STOCHASTIC ALTERNATING DIRECTION METHOD OF MULTIPLIERS | 69 |
| <i>Peilin Zhao, Jimwei Yang, Tong Zhang, Ping Li</i> | |
| A LOWER BOUND FOR THE OPTIMIZATION OF FINITE SUMS | 78 |
| <i>Alekh Agarwal, Leon Bottou</i> | |
| LEARNING WORD REPRESENTATIONS WITH HIERARCHICAL SPARSE CODING | 87 |
| <i>Dani Yogatama, Manaal Faruqui, Chris Dyer, Noah Smith</i> | |
| LEARNING TRANSFERABLE FEATURES WITH DEEP ADAPTATION NETWORKS | 97 |
| <i>Mingsheng Long, Yue Cao, Jianmin Wang, Michael Jordan</i> | |
| ROBUST PARTIALLY OBSERVABLE MARKOV DECISION PROCESS | 106 |
| <i>Takayuki Osogami</i> | |
| ON THE RELATIONSHIP BETWEEN SUM-PRODUCT NETWORKS AND BAYESIAN NETWORKS | 116 |
| <i>Han Zhao, Mazen Melibari, Pascal Poupart</i> | |
| LEARNING FROM CORRUPTED BINARY LABELS VIA CLASS-PROBABILITY ESTIMATION | 125 |
| <i>Aditya Menon, Brendan Van Rooyen, Cheng Soon Ong, Bob Williamson</i> | |
| AN EXPLICIT SAMPLING DEPENDENT SPECTRAL ERROR BOUND FOR COLUMN SUBSET SELECTION | 135 |
| <i>Tianbao Yang, Lijun Zhang, Rong Jin, Shenghuo Zhu</i> | |
| A STOCHASTIC PCA AND SVD ALGORITHM WITH AN EXPONENTIAL CONVERGENCE RATE | 144 |
| <i>Ohad Shamir</i> | |
| ATTRIBUTE EFFICIENT LINEAR REGRESSION WITH DISTRIBUTION-DEPENDENT SAMPLING | 153 |
| <i>Doron Kukliansky, Ohad Shamir</i> | |
| LEARNING LOCAL INVARIANT MAHALANOBIS DISTANCES | 162 |
| <i>Ethan Fetaya, Shimon Ullman</i> | |
| FINDING LINEAR STRUCTURE IN LARGE DATASETS WITH SCALABLE CANONICAL CORRELATION ANALYSIS | 169 |
| <i>Zhuang Ma, Yichao Lu, Dean Foster</i> | |
| ABSTRACTION SELECTION IN MODEL-BASED REINFORCEMENT LEARNING | 179 |
| <i>Nan Jiang, Alex Kulesza, Satinder Singh</i> | |
| SURROGATE FUNCTIONS FOR MAXIMIZING PRECISION AT THE TOP | 189 |
| <i>Purushottam Kar, Harikrishna Narasimhan, Prateek Jain</i> | |
| OPTIMIZING NON-DECOMPOSABLE PERFORMANCE MEASURES: A TALE OF TWO CLASSES | 199 |
| <i>Harikrishna Narasimhan, Purushottam Kar, Prateek Jain</i> | |
| CORESETS FOR NONPARAMETRIC ESTIMATION - THE CASE OF DP-MEANS | 209 |
| <i>Olivier Bachem, Mario Lucic, Andreas Krause</i> | |
| A RELATIVE EXPONENTIAL WEIGHING ALGORITHM FOR ADVERSARIAL UTILITY-BASED DUELING BANDITS | 218 |
| <i>Pratik Gajane, Tanguy Urvoy, Fabrice Clérot</i> | |

| | |
|---|-----|
| FUNCTIONAL SUBSPACE CLUSTERING WITH APPLICATION TO TIME SERIES | 228 |
| <i>Mohammad Taha Bahadori, David Kale, Yingying Fan, Yan Liu</i> | |
| ACCELERATED ONLINE LOW RANK TENSOR LEARNING FOR MULTIVARIATE SPATIOTEMPORAL STREAMS | 238 |
| <i>Rose Yu, Dehua Cheng, Yan Liu</i> | |
| ATOMIC SPATIAL PROCESSES | 248 |
| <i>Sean Jewell, Neil Spencer, Alexandre Bouchard-Côté</i> | |
| CLASSIFICATION WITH LOW RANK AND MISSING DATA | 257 |
| <i>Elad Hazan, Roi Livni, Yishay Mansour</i> | |
| DYNAMIC SENSING: BETTER CLASSIFICATION UNDER ACQUISITION CONSTRAINTS | 267 |
| <i>Oran Richman, Shie Mannor</i> | |
| A MODIFIED ORTHANT-WISE LIMITED MEMORY QUASI-NEWTON METHOD WITH CONVERGENCE ANALYSIS | 276 |
| <i>Pinghua Gong, Jieping Ye</i> | |
| TELLING CAUSE FROM EFFECT IN DETERMINISTIC LINEAR DYNAMICAL SYSTEMS | 285 |
| <i>Naji Shajarisales, Dominik Janzing, Bernhard Schoelkopf, Michel Besserve</i> | |
| HIGH DIMENSIONAL BAYESIAN OPTIMISATION AND BANDITS VIA ADDITIVE MODELS | 295 |
| <i>Kirthevasan Kandasamy, Jeff Schneider, Barnabas Poczos</i> | |
| THEORY OF DUAL-SPARSE REGULARIZED RANDOMIZED REDUCTION | 305 |
| <i>Tianbao Yang, Lijun Zhang, Rong Jin, Shenghuo Zhu</i> | |
| GENERALIZATION ERROR BOUNDS FOR LEARNING TO RANK: DOES THE LENGTH OF DOCUMENT LISTS MATTER? | 315 |
| <i>Ambuj Tewari, Sougata Chaudhuri</i> | |
| PEAKSEG: CONSTRAINED OPTIMAL SEGMENTATION AND SUPERVISED PENALTY LEARNING FOR PEAK DETECTION IN COUNT DATA | 324 |
| <i>Toby Hocking, Guillem Rigaill, Guillaume Bourque</i> | |
| MIND THE DUALITY GAP: SAFER RULES FOR THE LASSO | 333 |
| <i>Olivier Fercoq, Alexandre Gramfort, Joseph Salmon</i> | |
| A GENERAL ANALYSIS OF THE CONVERGENCE OF ADMM | 343 |
| <i>Robert Nishihara, Laurent Lessard, Ben Recht, Andrew Packard, Michael Jordan</i> | |
| STOCHASTIC PRIMAL-DUAL COORDINATE METHOD FOR REGULARIZED EMPIRICAL RISK MINIMIZATION | 353 |
| <i>Yuchen Zhang, Xiao Lin</i> | |
| DISCO: DISTRIBUTED OPTIMIZATION FOR SELF-CONCORDANT EMPIRICAL LOSS | 362 |
| <i>Yuchen Zhang, Xiao Lin</i> | |
| SPECTRAL MLE: TOP-K RANK AGGREGATION FROM PAIRWISE COMPARISONS | 371 |
| <i>Yuxin Chen, Changho Suh</i> | |
| PAIRED-DUAL LEARNING FOR FAST TRAINING OF LATENT VARIABLE HINGE-LOSS MRFS | 381 |
| <i>Stephen Bach, Bert Huang, Jordan Boyd-Graber, Lise Getoor</i> | |
| STRUCTURAL MAXENT MODELS | 391 |
| <i>Corinna Cortes, Vitaly Kuznetsov, Mehryar Mohri, Umar Syed</i> | |
| A PROVABLE GENERALIZED TENSOR SPECTRAL METHOD FOR UNIFORM HYPERGRAPH PARTITIONING | 400 |
| <i>Debarghya Ghoshdastidar, Ambedkar Dukkipati</i> | |
| THE BENEFITS OF LEARNING WITH STRONGLY CONVEX APPROXIMATE INFERENCE | 410 |
| <i>Ben London, Bert Huang, Lise Getoor</i> | |
| PUSHING THE LIMITS OF AFFINE RANK MINIMIZATION BY ADAPTING PROBABILISTIC PCA | 419 |
| <i>Bo Xin, David Wipf</i> | |
| BUDGET ALLOCATION PROBLEM WITH MULTIPLE ADVERTISERS: A GAME THEORETIC VIEW | 428 |
| <i>Takanori Maehara, Akihiro Yabe, Ken-Ichi Kawarabayashi</i> | |
| TRACKING APPROXIMATE SOLUTIONS OF PARAMETERIZED OPTIMIZATION PROBLEMS OVER MULTI-DIMENSIONAL (HYPER-)PARAMETER DOMAINS | 438 |
| <i>Katharina Blechschmidt, Joachim Giesen, Soeren Laue</i> | |
| BATCH NORMALIZATION: ACCELERATING DEEP NETWORK TRAINING BY REDUCING INTERNAL COVARIATE SHIFT | 448 |
| <i>Sergey Ioffe, Christian Szegedy</i> | |
| DISTRIBUTED ESTIMATION OF GENERALIZED MATRIX RANK: EFFICIENT ALGORITHMS AND LOWER BOUNDS | 457 |
| <i>Yuchen Zhang, Martin Wainwright, Michael Jordan</i> | |

| | |
|---|-----|
| LANDMARKING MANIFOLDS WITH GAUSSIAN PROCESSES | 466 |
| <i>Dawen Liang, John Paisley</i> | |
| MARKOV MIXED MEMBERSHIP MODELS | 475 |
| <i>Aonan Zhang, John Paisley</i> | |
| A UNIFIED FRAMEWORK FOR OUTLIER-ROBUST PCA-LIKE ALGORITHMS | 484 |
| <i>Wenzhuo Yang, Huan Xu</i> | |
| STREAMING SPARSE PRINCIPAL COMPONENT ANALYSIS | 494 |
| <i>Wenzhuo Yang, Huan Xu</i> | |
| A DIVIDE AND CONQUER FRAMEWORK FOR DISTRIBUTED GRAPH CLUSTERING | 504 |
| <i>Wenzhuo Yang, Huan Xu</i> | |
| HOW CAN DEEP RECTIFIER NETWORKS ACHIEVE LINEAR SEPARABILITY AND PRESERVE DISTANCES? | 514 |
| <i>Senjian An, Farid Boussaid, Mohammed Bennamoun</i> | |
| IMPROVED REGRET BOUNDS FOR UNDISCOUNTED CONTINUOUS REINFORCEMENT LEARNING | 524 |
| <i>K. Lakshmanan, Ronald Ortner, Daniil Ryabko</i> | |
| THE FUNDAMENTAL INCOMPATIBILITY OF SCALABLE HAMILTONIAN MONTE CARLO AND NAIVE DATA SUBSAMPLING | 533 |
| <i>Michael Betancourt</i> | |
| FASTER RATES FOR THE FRANK-WOLFE METHOD OVER STRONGLY-CONVEX SETS | 541 |
| <i>Dan Garber, Elad Hazan</i> | |
| ORDERED STICK-BREAKING PRIOR FOR SEQUENTIAL MCMC INFERENCE OF BAYESIAN NONPARAMETRIC MODELS | 550 |
| <i>Mrinal Das, Trapit Bansal, Chiranjib Bhattacharyya</i> | |
| ONLINE LEARNING OF EIGENVECTORS | 560 |
| <i>Dan Garber, Elad Hazan, Tengyu Ma</i> | |
| A UNIFYING FRAMEWORK OF ANYTIME SPARSE GAUSSIAN PROCESS REGRESSION MODELS WITH STOCHASTIC VARIATIONAL INFERENCE FOR BIG DATA | 569 |
| <i>Trong Nghia Hoang, Quang Minh Hoang, Bryan Kian Hsiang Low</i> | |
| YINYANG K-MEANS: A DROP-IN REPLACEMENT OF THE CLASSIC K-MEANS WITH CONSISTENT SPEEDUP | 579 |
| <i>Yufei Ding, Yue Zhao, Xipeng Shen, Madanlal Musuwathi, Todd Mytkowicz</i> | |
| ORDINAL MIXED MEMBERSHIP MODELS | 588 |
| <i>Seppo Virtanen, Mark Girolami</i> | |
| ONLINE TRACKING BY LEARNING DISCRIMINATIVE SALIENCY MAP WITH CONVOLUTIONAL NEURAL NETWORK | 597 |
| <i>Seunghoon Hong, Tackgeun You, Suha Kwak, Bohyung Han</i> | |
| FAST KRONECKER INFERENCE IN GAUSSIAN PROCESSES WITH NON-GAUSSIAN LIKELIHOODS | 607 |
| <i>Seth Flaxman, Andrew Wilson, Daniel Neill, Hannes Nickisch, Alex Smola</i> | |
| STATISTICAL AND ALGORITHMIC PERSPECTIVES ON RANDOMIZED SKETCHING FOR ORDINARY LEAST-SQUARES | 617 |
| <i>Garvesh Raskutti, Michael Mahoney</i> | |
| ON TD(0) WITH FUNCTION APPROXIMATION: CONCENTRATION BOUNDS AND A CENTERED VARIANT WITH EXPONENTIAL CONVERGENCE | 626 |
| <i>Nathaniel Korda, Prashanth La</i> | |
| LEARNING PARAMETRIC-OUTPUT HMMS WITH TWO ALIASED STATES | 635 |
| <i>Roi Weiss, Boaz Nadler</i> | |
| LATENT GAUSSIAN PROCESSES FOR DISTRIBUTION ESTIMATION OF MULTIVARIATE CATEGORICAL DATA | 645 |
| <i>Yarin Gal, Yutian Chen, Zoubin Ghahramani</i> | |
| IMPROVING THE GAUSSIAN PROCESS SPARSE SPECTRUM APPROXIMATION BY REPRESENTING UNCERTAINTY IN FREQUENCY INPUTS | 655 |
| <i>Yarin Gal, Richard Turner</i> | |
| RANKING FROM STOCHASTIC PAIRWISE PREFERENCES: RECOVERING CONDORCET WINNERS AND TOURNAMENT SOLUTION SETS AT THE TOP | 665 |
| <i>Arun Rajkumar, Suprovat Ghoshal, Lek-Heng Lim, Shivani Agarwal</i> | |
| STOCHASTIC DUAL COORDINATE ASCENT WITH ADAPTIVE PROBABILITIES | 674 |
| <i>Dominik Csiba, Zheng Qu, Peter Richtarik</i> | |
| VECTOR-SPACE MARKOV RANDOM FIELDS VIA EXPONENTIAL FAMILIES | 684 |
| <i>Wesley Tansey, Oscar Hernan Madrid Padilla, Arun Sai Suggala, Pradeep Ravikumar</i> | |

| | |
|---|-----|
| JUMP-MEANS: SMALL-VARIANCE ASYMPTOTICS FOR MARKOV JUMP PROCESSES | 693 |
| <i>Jonathan Huggins, Karthik Narasimhan, Ardavan Saeedi, Vikash Mansinghka</i> | |
| LOW RANK APPROXIMATION USING ERROR CORRECTING CODING MATRICES | 702 |
| <i>Shashanka Ubaru, Arya Mazumdar, Yousef Saad</i> | |
| OFF-POLICY MODEL-BASED LEARNING UNDER UNKNOWN FACTORED DYNAMICS | 711 |
| <i>Assaf Hallak, Francois Schnitzler, Timothy Mann, Shie Mannor</i> | |
| LOG-EUCLIDEAN METRIC LEARNING ON SYMMETRIC POSITIVE DEFINITE MANIFOLD WITH APPLICATION TO IMAGE SET CLASSIFICATION | 720 |
| <i>Zhiwu Huang, Ruiping Wang, Shiguang Shan, Xianqiu Li, Xilin Chen</i> | |
| ASYMMETRIC TRANSFER LEARNING WITH DEEP GAUSSIAN PROCESSES | 730 |
| <i>Melih Kandemir</i> | |
| TOWARDS A LOWER SAMPLE COMPLEXITY FOR ROBUST ONE-BIT COMPRESSED SENSING | 739 |
| <i>Rongda Zhu, Quanquan Gu</i> | |
| BILBOWA: FAST BILINGUAL DISTRIBUTED REPRESENTATIONS WITHOUT WORD ALIGNMENTS | 748 |
| <i>Stephan Gouws, Yoshua Bengio, Greg Corrado</i> | |
| MULTI-VIEW SPARSE CO-CLUSTERING VIA PROXIMAL ALTERNATING LINEARIZED MINIMIZATION | 757 |
| <i>Jiangwen Sun, Jin Lu, Tingyang Xu, Jinbo Bi</i> | |
| CASCADING BANDITS: LEARNING TO RANK IN THE CASCADE MODEL | 767 |
| <i>Branislav Kveton, Csaba Szepesvari, Zheng Wen, Azin Ashkan</i> | |
| LATENT TOPIC NETWORKS: A VERSATILE PROBABILISTIC PROGRAMMING FRAMEWORK FOR TOPIC MODELS | 777 |
| <i>James Foulds, Shachi Kumar, Lise Getoor</i> | |
| RANDOM COORDINATE DESCENT METHODS FOR MINIMIZING DECOMPOSABLE SUBMODULAR FUNCTIONS | 787 |
| <i>Alina Ene, Huy Nguyen</i> | |
| ALPHA-BETA DIVERGENCES DISCOVER MICRO AND MACRO STRUCTURES IN DATA | 796 |
| <i>Karthik Narayan, Ali Punjani, Pieter Abbeel</i> | |
| FICTITIOUS SELF-PLAY IN EXTENSIVE-FORM GAMES | 805 |
| <i>Johannes Heinrich, Marc Lanctot, David Silver</i> | |
| COUNTERFACTUAL RISK MINIMIZATION: LEARNING FROM LOGGED BANDIT FEEDBACK | 814 |
| <i>Adith Swaminathan, Thorsten Joachims</i> | |
| THE HEDGE ALGORITHM ON A CONTINUUM | 824 |
| <i>Walid Krichene, Maximilian Balandat, Claire Tomlin, Alexandre Bayen</i> | |
| A LINEAR DYNAMICAL SYSTEM MODEL FOR TEXT | 833 |
| <i>David Belanger, Sham Kakade</i> | |
| UNSUPERVISED LEARNING OF VIDEO REPRESENTATIONS USING LSTMS | 843 |
| <i>Nitish Srivastava, Elman Mansimov, Ruslan Salakhudinov</i> | |

VOLUME 2

| | |
|---|-----|
| MESSAGE PASSING FOR COLLECTIVE GRAPHICAL MODELS | 853 |
| <i>Tao Sun, Dan Sheldon, Akshat Kumar</i> | |
| DP-SPACE: BAYESIAN NONPARAMETRIC SUBSPACE CLUSTERING WITH SMALL-VARIANCE ASYMPTOTICS | 862 |
| <i>Yining Wang, Jun Zhu</i> | |
| HAWKESTOPIC: A JOINT MODEL FOR NETWORK INFERENCE AND TOPIC MODELING FROM TEXT-BASED CASCADES | 871 |
| <i>Xinran He, Theodoros Rekatsinas, James Foulds, Lise Getoor, Yan Liu</i> | |
| MADE: MASKED AUTOENCODER FOR DISTRIBUTION ESTIMATION | 881 |
| <i>Mathieu Germain, Karol Gregor, Iain Murray, Hugo Larochelle</i> | |
| AN ONLINE LEARNING ALGORITHM FOR BILINEAR MODELS | 890 |
| <i>Yuanbin Wu, Shiliang Sun</i> | |
| ADAPTIVE BELIEF PROPAGATION | 899 |
| <i>Georgios Papachristoudis, John Fisher</i> | |
| LARGE-SCALE LOG-DETERMINANT COMPUTATION THROUGH STOCHASTIC CHEBYSHEV EXPANSIONS | 908 |
| <i>Insu Han, Dmitry Malioutov, Jinwoo Shin</i> | |

| | |
|---|------|
| DIFFERENTIALLY PRIVATE BAYESIAN OPTIMIZATION | 918 |
| <i>Matt Kusner, Jacob Gardner, Roman Garnett, Kilian Weinberger</i> | |
| A NEARLY-LINEAR TIME FRAMEWORK FOR GRAPH-STRUCTURED SPARSITY | 928 |
| <i>Chinmay Hegde, Piotr Indyk, Ludwig Schmidt</i> | |
| SUPPORT MATRIX MACHINES | 938 |
| <i>Luo Luo, Yubo Xie, Zhihua Zhang, Wu-Jun Li</i> | |
| RADEMACHER OBSERVATIONS, PRIVATE DATA, AND BOOSTING | 948 |
| <i>Richard Nock, Giorgio Patrini, Arik Friedman</i> | |
| FROM WORD EMBEDDINGS TO DOCUMENT DISTANCES | 957 |
| <i>Matt Kusner, Yu Sun, Nicholas Kolkin, Kilian Weinberger</i> | |
| BAYESIAN AND EMPIRICAL BAYESIAN FORESTS | 967 |
| <i>Taddy Matthew, Chun-Sheng Chen, Jun Yu, Mitch Wyle</i> | |
| INFERRING GRAPHS FROM CASCADES: A SPARSE RECOVERY FRAMEWORK | 977 |
| <i>Jean Pouget-Abadie, Thibaut Horel</i> | |
| DISTRIBUTED BOX-CONSTRAINED QUADRATIC OPTIMIZATION FOR DUAL LINEAR SVM | 987 |
| <i>Ching-Pei Lee, Dan Roth</i> | |
| SAFE EXPLORATION FOR OPTIMIZATION WITH GAUSSIAN PROCESSES | 997 |
| <i>Yanan Sui, Alkis Gotovos, Joel Burdick, Andreas Krause</i> | |
| THE LADDER: A RELIABLE LEADERBOARD FOR MACHINE LEARNING COMPETITIONS | 1006 |
| <i>Avrim Blum, Moritz Hardt</i> | |
| ENABLING SCALABLE STOCHASTIC GRADIENT-BASED INFERENCE FOR GAUSSIAN PROCESSES BY EMPLOYING THE UNBIASED LINEAR SYSTEM SOLVER (ULISSE) | 1015 |
| <i>Maurizio Filippone, Raphael Engler</i> | |
| FINDING GALAXIES IN THE SHADOWS OF QUASARS WITH GAUSSIAN PROCESSES | 1025 |
| <i>Roman Garnett, Shirley Ho, Jeff Schneider</i> | |
| FOLLOWING THE PERTURBED LEADER FOR ONLINE STRUCTURED LEARNING | 1034 |
| <i>Alon Cohen, Tamir Hazan</i> | |
| REIFIED CONTEXT MODELS | 1043 |
| <i>Jacob Steinhardt, Percy Liang</i> | |
| LARGE-SCALE MARKOV DECISION PROBLEMS WITH KL CONTROL COST AND ITS APPLICATION TO CROWDSOURCING | 1053 |
| <i>Yasin Abbasi-Yadkori, Peter Bartlett, Xi Chen, Alan Malek</i> | |
| LEARNING FAST-MIXING MODELS FOR STRUCTURED PREDICTION | 1063 |
| <i>Jacob Steinhardt, Percy Liang</i> | |
| A PROBABILISTIC MODEL FOR DIRTY MULTI-TASK FEATURE SELECTION | 1073 |
| <i>Daniel Hernandez-Lobato, Jose Miguel Hernandez-Lobato, Zoubin Ghahramani</i> | |
| ON DEEP MULTI-VIEW REPRESENTATION LEARNING | 1083 |
| <i>Weiran Wang, Raman Arora, Karen Livescu, Jeff Bilmes</i> | |
| LEARNING PROGRAM EMBEDDINGS TO PROPAGATE FEEDBACK ON STUDENT CODE | 1093 |
| <i>Chris Piech, Jonathan Huang, Andy Nguyen, Mike Phulsuksombati, Mehran Sahami, Leonidas Guibas</i> | |
| SAFE SUBSPACE SCREENING FOR NUCLEAR NORM REGULARIZED LEAST SQUARES PROBLEMS | 1103 |
| <i>Qiang Zhou, Qi Zhao</i> | |
| EFFICIENT LEARNING IN LARGE-SCALE COMBINATORIAL SEMI-BANDITS | 1113 |
| <i>Zheng Wen, Branislav Kveton, Azin Ashkan</i> | |
| SWEPT APPROXIMATE MESSAGE PASSING FOR SPARSE ESTIMATION | 1123 |
| <i>Andre Manoel, Florent Krzakala, Eric Tramel, Lenka Zdeborová</i> | |
| SIMPLE REGRET FOR INFINITELY MANY ARMED BANDITS | 1133 |
| <i>Alexandra Carpentier, Michal Valko</i> | |
| EXPONENTIAL INTEGRATION FOR HAMILTONIAN MONTE CARLO | 1142 |
| <i>Wei-Lun Chao, Justin Solomon, Dominik Michels, Fei Sha</i> | |
| OPTIMAL REGRET ANALYSIS OF THOMPSON SAMPLING IN STOCHASTIC MULTI-ARMED BANDIT PROBLEM WITH MULTIPLE PLAYS | 1152 |
| <i>Junpei Komiyama, Junya Honda, Hiroshi Nakagawa</i> | |
| FASTER COVER TREES | 1162 |
| <i>Mike Izbicki, Christian Shelton</i> | |
| BLITZ: A PRINCIPLED META-ALGORITHM FOR SCALING SPARSE OPTIMIZATION | 1171 |
| <i>Tyler Johnson, Carlos Guestrin</i> | |
| UNSUPERVISED DOMAIN ADAPTATION BY BACKPROPAGATION | 1180 |
| <i>Yaroslav Ganin, Victor Lempitsky</i> | |

| | |
|---|------|
| NON-LINEAR CROSS-DOMAIN COLLABORATIVE FILTERING VIA HYPER-STRUCTURE TRANSFER | 1190 |
| <i>Yan-Fu Liu, Cheng-Yu Hsu, Shan-Hung Wu</i> | |
| MANIFOLD-VALUED DIRICHLET PROCESSES | 1199 |
| <i>Hyunwoo Kim, Jia Xu, Baba Vemuri, Vikas Singh</i> | |
| MULTI-TASK LEARNING FOR SUBSPACE SEGMENTATION | 1209 |
| <i>Yu Wang, David Wipf, Qing Ling, Wei Chen, Ian Wassell</i> | |
| MARKOV CHAIN MONTE CARLO AND VARIATIONAL INFERENCE: BRIDGING THE GAP | 1218 |
| <i>Tim Salimans, Diederik Kingma, Max Welling</i> | |
| SCALABLE MODEL SELECTION FOR LARGE-SCALE FACTORIAL RELATIONAL MODELS | 1227 |
| <i>Chunchen Liu, Lu Feng, Ryohei Fujimaki, Yusuke Muraoka</i> | |
| THE POWER OF RANDOMIZATION: DISTRIBUTED SUBMODULAR MAXIMIZATION ON MASSIVE DATASETS | 1236 |
| <i>Rafael Barbosa, Alina Ene, Huy Nguyen, Justin Ward</i> | |
| DEALING WITH SMALL DATA: ON THE GENERALIZATION OF CONTEXT TREES | 1245 |
| <i>Ralf Eggeling, Mikko Koivisto, Ivo Grosse</i> | |
| NON-GAUSSIAN DISCRIMINATIVE FACTOR MODELS VIA THE MAX-MARGIN RANK-LIKELIHOOD | 1254 |
| <i>Xin Yuan, Ricardo Henao, Ephraim Tsalik, Raymond Langley, Lawrence Carin</i> | |
| A BAYESIAN NONPARAMETRIC PROCEDURE FOR COMPARING ALGORITHMS | 1264 |
| <i>Alessio Benavoli, Giorgio Corani, Francesca Mangili, Marco Zaffalon</i> | |
| CONVERGENCE RATE OF BAYESIAN TENSOR ESTIMATOR AND ITS MINIMAX OPTIMALITY | 1273 |
| <i>Taiji Suzuki</i> | |
| ON IDENTIFYING GOOD OPTIONS UNDER COMBINATORIALLY STRUCTURED FEEDBACK IN FINITE NOISY ENVIRONMENTS | 1283 |
| <i>Yifan Wu, Andras Gyorgy, Csaba Szepesvari</i> | |
| NESTED SEQUENTIAL MONTE CARLO METHODS | 1292 |
| <i>Christian Naeseth, Fredrik Lindsten, Thomas Schon</i> | |
| SPARSE VARIATIONAL INFERENCE FOR GENERALIZED GP MODELS | 1302 |
| <i>Rishit Sheth, Yuyang Wang, Roni Khardon</i> | |
| UNIVERSAL VALUE FUNCTION APPROXIMATORS | 1312 |
| <i>Tom Schaul, Daniel Horgan, Karol Gregor, David Silver</i> | |
| APPROXIMATE DYNAMIC PROGRAMMING FOR TWO-PLAYER ZERO-SUM MARKOV GAMES | 1321 |
| <i>Julien Perolat, Bruno Scherrer, Bilal Piot, Olivier Pietquin</i> | |
| ON GREEDY MAXIMIZATION OF ENTROPY | 1330 |
| <i>Dravyansh Sharma, Ashish Kapoor, Amit Deshpande</i> | |
| METADATA DEPENDENT MONDRIAN PROCESSES | 1339 |
| <i>Yi Wang, Bin Li, Yang Wang, Fang Chen</i> | |
| COMPLEX EVENT DETECTION USING SEMANTIC SALIENCY AND NEARLY-ISOTONIC SVM | 1348 |
| <i>Xiaojun Chang, Yi Yang, Eric Xing, Yaoliang Yu</i> | |
| REBUILDING FACTORIZED INFORMATION CRITERION: ASYMPTOTICALLY ACCURATE MARGINAL LIKELIHOOD | 1358 |
| <i>Kohei Hayashi, Shin-Ichi Maeda, Ryohei Fujimaki</i> | |
| DOUBLE NYSTRÖM METHOD: AN EFFICIENT AND ACCURATE NYSTRÖM SCHEME FOR LARGE-SCALE DATA SETS | 1367 |
| <i>Woosang Lim, Minhwan Kim, Haesun Park, Kyomin Jung</i> | |
| THE COMPOSITION THEOREM FOR DIFFERENTIAL PRIVACY | 1376 |
| <i>Peter Kairouz, Sewoong Oh, Pramod Viswanath</i> | |
| CONVEX FORMULATION FOR LEARNING FROM POSITIVE AND UNLABELED DATA | 1386 |
| <i>Marthinus Du Plessis, Gang Niu, Masashi Sugiyama</i> | |
| THRESHOLD INFLUENCE MODEL FOR ALLOCATING ADVERTISING BUDGETS | 1395 |
| <i>Atsushi Miyaochi, Yuni Iwamasa, Takuro Fukunaga, Naonori Kakimura</i> | |
| STRONGLY ADAPTIVE ONLINE LEARNING | 1405 |
| <i>Amit Daniely, Alon Gonen, Shai Shalev-Shwartz</i> | |
| CUR ALGORITHM FOR PARTIALLY OBSERVED MATRICES | 1412 |
| <i>Miao Xu, Rong Jin, Zhi-Hua Zhou</i> | |
| A DETERMINISTIC ANALYSIS OF NOISY SPARSE SUBSPACE CLUSTERING FOR DIMENSIONALITY-REDUCED DATA | 1422 |
| <i>Yining Wang, Yu-Xiang Wang, Aarti Singh</i> | |

| | |
|--|------|
| MRA-BASED STATISTICAL LEARNING FROM INCOMPLETE RANKINGS | 1432 |
| <i>Eric Sibony, Stéphane Cléménçon, Jérémie Jakubowicz</i> | |
| RISK AND REGRET OF HIERARCHICAL BAYESIAN LEARNERS | 1442 |
| <i>Jonathan Huggins, Josh Tenenbaum</i> | |
| TOWARDS A LEARNING THEORY OF CAUSE-EFFECT INFERENCE | 1452 |
| <i>David Lopez-Paz, Krikamol Muandet, Bernhard Schölkopf, Iliya Tolstikhin</i> | |
| DRAW: A RECURRENT NEURAL NETWORK FOR IMAGE GENERATION | 1462 |
| <i>Karol Gregor, Ivo Danihelka, Alex Graves, Danilo Rezende, Daan Wierstra</i> | |
| MULTIVIEW TRIPLET EMBEDDING: LEARNING ATTRIBUTES IN MULTIPLE MAPS | 1472 |
| <i>Ehsan Amid, Antti Ukkonen</i> | |
| DISTRIBUTED GAUSSIAN PROCESSES | 1481 |
| <i>Marc Deisenroth, Jun Wei Ng</i> | |
| GUARANTEED TENSOR DECOMPOSITION: A MOMENT APPROACH | 1491 |
| <i>Gongguo Tang, Parikshit Shah</i> | |
| $L_{1,p}$-NORM REGULARIZATION: ERROR BOUNDS AND CONVERGENCE RATE ANALYSIS OF FIRST-ORDER METHODS | 1501 |
| <i>Zirui Zhou, Qi Zhang, Anthony Man-Cho So</i> | |
| CONSISTENT ESTIMATION OF DYNAMIC AND MULTI-LAYER BLOCK MODELS | 1511 |
| <i>Qiuyi Han, Kevin Xu, Edoardo Airoldi</i> | |
| ON THE RATE OF CONVERGENCE AND ERROR BOUNDS FOR LSTD(λ) | 1521 |
| <i>Manel Tagorti, Bruno Scherrer</i> | |
| VARIATIONAL INFERENCE WITH NORMALIZING FLOWS | 1530 |
| <i>Danilo Rezende, Shakir Mohamed</i> | |
| CONTROVERSY IN MECHANISTIC MODELLING WITH GAUSSIAN PROCESSES | 1539 |
| <i>Benn Macdonald, Catherine Higham, Dirk Husmeier</i> | |
| CONVEX LEARNING OF MULTIPLE TASKS AND THEIR STRUCTURE | 1548 |
| <i>Carlo Ciliberto, Youssef Mroueh, Tomaso Poggio, Lorenzo Rosasco</i> | |
| K-HYPERPLANE HINGE-MINIMAX CLASSIFIER | 1558 |
| <i>Margarita Osadchy, Tamir Hazan, Daniel Keren</i> | |
| NON-STATIONARY APPROXIMATE MODIFIED POLICY ITERATION | 1567 |
| <i>Boris Lesner, Bruno Scherrer</i> | |
| ENTROPY EVALUATION BASED ON CONFIDENCE INTERVALS OF FREQUENCY ESTIMATES : APPLICATION TO THE LEARNING OF DECISION TREES | 1576 |
| <i>Mathieu Serrurier, Henri Prade</i> | |
| GEOMETRIC CONDITIONS FOR SUBSPACE-SPARSE RECOVERY | 1585 |
| <i>Chong You, Rene Vidal</i> | |
| AN EMPIRICAL STUDY OF STOCHASTIC VARIATIONAL INFERENCE ALGORITHMS FOR THE BETA BERNOULLI PROCESS | 1594 |
| <i>Amar Shah, David Knowles, Zoubin Ghahramani</i> | |
| LONG SHORT-TERM MEMORY OVER RECURSIVE STRUCTURES | 1604 |
| <i>Xiaodan Zhu, Parinaz Sobhani, Hongyu Guo</i> | |
| WEIGHT UNCERTAINTY IN NEURAL NETWORK | 1613 |
| <i>Charles Blundell, Julien Cornebise, Koray Kavukcuoglu, Daan Wierstra</i> | |
| LEARNING SUBMODULAR LOSSES WITH THE LOVASZ HINGE | 1623 |
| <i>Jiaqian Yu, Matthew Blaschko</i> | |
| COORDINATE DESCENT CONVERGES FASTER WITH THE GAUSS-SOUTHWELL RULE THAN RANDOM SELECTION | 1632 |
| <i>Julie Nutini, Mark Schmidt, Issam Laradji, Michael Friedlander, Hoyt Koepke</i> | |
| HASHING FOR DISTRIBUTED DATA | 1642 |
| <i>Cong Leng, Jiayang Wu, Jian Cheng, Xi Zhang, Hanqing Lu</i> | |
| LARGE-SCALE DISTRIBUTED DEPENDENT NONPARAMETRIC TREES | 1651 |
| <i>Zhiting Hu, Ho Qirong, Avinava Dubey, Eric Xing</i> | |
| QUALITATIVE MULTI-ARMED BANDITS: A QUANTILE-BASED APPROACH | 1660 |
| <i>Balazs Szorenyi, Robert Busa-Fekete, Paul Weng, Eyke Hüllermeier</i> | |
| DEEP EDGE-AWARE FILTERS | 1669 |
| <i>Li Xu, Jimmy Ren, Qiong Yan, Renjie Liao, Jiaya Jia</i> | |
| A CONVEX OPTIMIZATION FRAMEWORK FOR BI-CLUSTERING | 1679 |
| <i>Shiau Hong Lim, Yudong Chen, Huan Xu</i> | |
| IS FEATURE SELECTION SECURE AGAINST TRAINING DATA POISONING? | 1689 |
| <i>Huang Xiao, Battista Biggio, Gavin Brown, Giorgio Fumera, Claudia Eckert, Fabio Roli</i> | |

| | |
|--|------|
| PREDICTIVE ENTROPY SEARCH FOR BAYESIAN OPTIMIZATION WITH UNKNOWN CONSTRAINTS | 1699 |
| <i>Jose Miguel Hernandez-Lobato, Michael Gelbart, Matthew Hoffman, Ryan Adams, Zoubin Ghahramani</i> | |

VOLUME 3

| | |
|--|------|
| A THEORETICAL ANALYSIS OF METRIC HYPOTHESIS TRANSFER LEARNING | 1708 |
| <i>Michaël Perrot, Amaury Habrard</i> | |
| GENERATIVE MOMENT MATCHING NETWORKS | 1718 |
| <i>Yujia Li, Kevin Swersky, Rich Zemel</i> | |
| STAY ON PATH: PCA ALONG GRAPH PATHS | 1728 |
| <i>Megasthenis Asteris, Anastastos Kyriellidis, Alex Dimakis, Han-Gyol Yi, Bharath Chandrasekaran</i> | |
| DEEP LEARNING WITH LIMITED NUMERICAL PRECISION | 1737 |
| <i>Suyog Gupta, Ankur Agrawal, Kailash Gopalakrishnan, Pritish Narayanan</i> | |
| SAFE SCREENING FOR MULTI-TASK FEATURE LEARNING WITH MULTIPLE DATA MATRICES | 1747 |
| <i>Jie Wang, Jieping Ye</i> | |
| HARMONIC EXPONENTIAL FAMILIES ON MANIFOLDS | 1757 |
| <i>Taco Cohen, Max Welling</i> | |
| TRAINING DEEP CONVOLUTIONAL NEURAL NETWORKS TO PLAY GO | 1766 |
| <i>Christopher Clark, Amos Storkey</i> | |
| KERNEL INTERPOLATION FOR SCALABLE STRUCTURED GAUSSIAN PROCESSES (KISS-GP) | 1775 |
| <i>Andrew Wilson, Hannes Nickisch</i> | |
| LEARNING DEEP STRUCTURED MODELS | 1785 |
| <i>Liang-Chieh Chen, Alexander Schwing, Alan Yuille, Raquel Urtasun</i> | |
| COMMUNITY DETECTION USING TIME-DEPENDENT PERSONALIZED PAGERANK | 1795 |
| <i>Haim Avron, Lior Horesh</i> | |
| SCALABLE VARIATIONAL INFERENCE IN LOG-SUPERMODULAR MODELS | 1804 |
| <i>Josip Djolonga, Andreas Krause</i> | |
| VARIATIONAL INFERENCE FOR GAUSSIAN PROCESS MODULATED POISSON PROCESSES | 1814 |
| <i>Chris Lloyd, Tom Gunter, Michael Osborne, Stephen Roberts</i> | |
| SCALABLE DEEP POISSON FACTOR ANALYSIS FOR TOPIC MODELING | 1823 |
| <i>Zhe Gan, Changyou Chen, Ricardo Henao, David Carlson, Lawrence Carin</i> | |
| HIDDEN MARKOV ANOMALY DETECTION | 1833 |
| <i>Nico Goernitz, Mikio Braun, Marius Kloft</i> | |
| ROBUST ESTIMATION OF TRANSITION MATRICES IN HIGH DIMENSIONAL HEAVY-TAILED VECTOR AUTOREGRESSIVE PROCESSES | 1843 |
| <i>Huitong Qiu, Sheng Xu, Fang Han, Han Liu, Brian Caffo</i> | |
| CONVEX CALIBRATED SURROGATES FOR HIERARCHICAL CLASSIFICATION | 1852 |
| <i>Harish Ramaswamy, Ambuj Tewari, Shivani Agarwal</i> | |
| PROBABILISTIC BACKPROPAGATION FOR SCALABLE LEARNING OF BAYESIAN NEURAL NETWORKS | 1861 |
| <i>Jose Miguel Hernandez-Lobato, Ryan Adams</i> | |
| ACTIVE NEAREST NEIGHBORS IN CHANGING ENVIRONMENTS | 1870 |
| <i>Christopher Berling, Ruth Uerner</i> | |
| BIPARTITE EDGE PREDICTION VIA TRANSDUCTIVE LEARNING OVER PRODUCT GRAPHS | 1880 |
| <i>Hanxiao Liu, Yiming Yang</i> | |
| TRUST REGION POLICY OPTIMIZATION | 1889 |
| <i>John Schulman, Sergey Levine, Pieter Abbeel, Michael Jordan, Philipp Moritz</i> | |
| DISCOVERING TEMPORAL CAUSAL RELATIONS FROM SUBSAMPLED DATA | 1898 |
| <i>Mingming Gong, Kun Zhang, Bernhard Schoelkopf, Dacheng Tao, Philipp Geiger</i> | |
| PREFERENCE COMPLETION: LARGE-SCALE COLLABORATIVE RANKING FROM PAIRWISE COMPARISONS | 1907 |
| <i>Dohyung Park, Joe Neeman, Jin Zhang, Sujay Sanghavi, Inderjit Dhillon</i> | |
| CAUSAL INFERENCE BY IDENTIFICATION OF VECTOR AUTOREGRESSIVE PROCESSES WITH HIDDEN COMPONENTS | 1917 |
| <i>Philipp Geiger, Kun Zhang, Bernhard Schoelkopf, Mingming Gong, Dominik Janzing</i> | |

| | |
|--|------|
| ON SYMMETRIC AND ASYMMETRIC LSHS FOR INNER PRODUCT SEARCH..... | 1926 |
| <i>Behnam Neyshabur, Nathan Srebro</i> | |
| THE KENDALL AND MALLOWS KERNELS FOR PERMUTATIONS | 1935 |
| <i>Yunlong Jiao, Jean-Philippe Vert</i> | |
| BAYESIAN MULTIPLE TARGET LOCALIZATION..... | 1945 |
| <i>Purnima Rajan, Weidong Han, Raphael Sznitman, Peter Frazier, Bruno Jedynak</i> | |
| SUBMODULARITY IN DATA SUBSET SELECTION AND ACTIVE LEARNING | 1954 |
| <i>Kai Wei, Rishabh Iyer, Jeff Bilmes</i> | |
| VARIATIONAL GENERATIVE STOCHASTIC NETWORKS WITH COLLABORATIVE SHAPING..... | 1964 |
| <i>Philip Bachman, Doina Precup</i> | |
| ADDING VS. AVERAGING IN DISTRIBUTED PRIMAL-DUAL OPTIMIZATION..... | 1973 |
| <i>Chenxin Ma, Virginia Smith, Martin Jaggi, Michael Jordan, Peter Richtarik, Martin Takac</i> | |
| FEATURE-BUDGETED RANDOM FOREST..... | 1983 |
| <i>Feng Nan, Joseph Wang, Venkatesh Saligrama</i> | |
| ENTROPIC GRAPH-BASED POSTERIOR REGULARIZATION | 1992 |
| <i>Maxwell Libbrecht, Michael Hoffman, Jeff Bilmes, William Noble</i> | |
| UNSUPERVISED RIEMANNIAN METRIC LEARNING FOR HISTOGRAMS USING AITCHISON TRANSFORMATIONS..... | 2002 |
| <i>Tam Le, Marco Cuturi</i> | |
| LOW-RANK MATRIX RECOVERY FROM ROW-AND-COLUMN AFFINE MEASUREMENTS..... | 2012 |
| <i>Or Zuk, Avishai Wagner</i> | |
| ALGORITHMS FOR THE HARD PRE-IMAGE PROBLEM OF STRING KERNELS AND THE GENERAL PROBLEM OF STRING PREDICTION..... | 2021 |
| <i>Sébastien Giguère, Amélie Rolland, Francois Laviolette, Mario Marchand</i> | |
| A MULTITASK POINT PROCESS PREDICTIVE MODEL..... | 2030 |
| <i>Wenzhao Lian, Ricardo Henao, Vinayak Rao, Joseph Lucas, Lawrence Carin</i> | |
| A HYBRID APPROACH FOR PROBABILISTIC INFERENCE USING RANDOM PROJECTIONS | 2039 |
| <i>Michael Zhu, Stefano Ermon</i> | |
| SHOW, ATTEND AND TELL: NEURAL IMAGE CAPTION GENERATION WITH VISUAL ATTENTION | 2048 |
| <i>Kelvin Xu, Jimmy Ba, Ryan Kiros, Kyunghyun Cho, Aaron Courville, Ruslan Salakhudinov, Rich Zemel, Yoshua Bengio</i> | |
| LEARNING TO SEARCH BETTER THAN YOUR TEACHER..... | 2058 |
| <i>Kai-Wei Chang, Akshay Krishnamurthy, Alekh Agarwal, Hal Daume, John Langford</i> | |
| GATED FEEDBACK RECURRENT NEURAL NETWORKS..... | 2067 |
| <i>Junyoung Chung, Caglar Gulcehre, Kyunghyun Cho, Yoshua Bengio</i> | |
| CONTEXT-BASED UNSUPERVISED DATA FUSION FOR DECISION MAKING | 2076 |
| <i>Erfan Soltanmohammadi, Mort Naraghi-Pour, Mihaela Van Der Schaer</i> | |
| PHRASE-BASED IMAGE CAPTIONING | 2085 |
| <i>Remi Lebreton, Pedro Pinheiro, Ronan Collobert</i> | |
| CELESTE: VARIATIONAL INFERENCE FOR A GENERATIVE MODEL OF ASTRONOMICAL IMAGES..... | 2095 |
| <i>Jeffrey Regier, Andrew Miller, Jon McAuliffe, Ryan Adams, Matt Hoffman, Dustin Lang, David Schlegel, Mr Prabhat</i> | |
| DISTRIBUTIONAL RANK AGGREGATION, AND AN AXIOMATIC ANALYSIS | 2104 |
| <i>Adarsh Prasad, Harsh Pareek, Pradeep Ravikumar</i> | |
| GRADIENT-BASED HYPERPARAMETER OPTIMIZATION THROUGH REVERSIBLE LEARNING..... | N/A |
| <i>Dougal Maclaurin, David Duvenaud, Ryan Adams</i> | |
| BIMODAL MODELLING OF SOURCE CODE AND NATURAL LANGUAGE | 2113 |
| <i>Miltos Allamanis, Daniel Tarlow, Andrew Gordon, Yi Wei</i> | |
| CHEAP BANDITS | 2123 |
| <i>Manjesh Hanawal, Venkatesh Saligrama, Michal Valko, Remi Munos</i> | |
| SUBSAMPLING METHODS FOR PERSISTENT HOMOLOGY | 2133 |
| <i>Frederic Chazal, Brittany Fasy, Fabrizio Lecci, Bertrand Michel, Alessandro Rinaldo, Larry Wasserman</i> | |
| AN EMBARRASSINGLY SIMPLE APPROACH TO ZERO-SHOT LEARNING..... | 2142 |
| <i>Bernardino Romera-Paredes, Philip Torr</i> | |
| BINARY EMBEDDING: FUNDAMENTAL LIMITS AND FAST ALGORITHM..... | 2152 |
| <i>Xinyang Yi, Constantine Caramanis, Eric Price</i> | |

| | |
|---|------|
| SCALABLE BAYESIAN OPTIMIZATION USING DEEP NEURAL NETWORKS | 2161 |
| <i>Jasper Snoek, Oren Rippel, Kevin Swersky, Ryan Kiros, Nadathur Satish, Narayanan Sundaram, Mostofa Patwary, Mr Prabhat, Ryan Adams</i> | |
| HOW HARD IS INFERENCE FOR STRUCTURED PREDICTION? | 2171 |
| <i>Amir Globerson, Tim Roughgarden, David Sontag, Cafer Yildirim</i> | |
| ONLINE TIME SERIES PREDICTION WITH MISSING DATA | 2181 |
| <i>Oren Anava, Elad Hazan, Assaf Zeevi</i> | |
| PROTEINS, PARTICLES, AND PSEUDO-MAX-MARGINALS: A SUBMODULAR APPROACH | 2190 |
| <i>Jason Pacheco, Erik Sudderth</i> | |
| A FAST VARIATIONAL APPROACH FOR LEARNING MARKOV RANDOM FIELD LANGUAGE MODELS | 2199 |
| <i>Yacine Jernite, Alexander Rush, David Sontag</i> | |
| REMOVING SYSTEMATIC ERRORS FOR EXOPLANET SEARCH VIA LATENT CAUSES | 2208 |
| <i>Bernhard Schölkopf, David Hogg, Dun Wang, Dan Foreman-Mackey, Dominik Janzing, Carl-Johann Simon-Gabriel, Jonas Peters</i> | |
| SCALABLE NONPARAMETRIC BAYESIAN INFERENCE ON POINT PROCESSES WITH GAUSSIAN PROCESSES | 2217 |
| <i>Yves-Laurent Kom Samo, Stephen Roberts</i> | |
| CORRELATION CLUSTERING IN DATA STREAMS | 2227 |
| <i>Kookjin Ahn, Graham Cormode, Sudipto Guha, Andrew McGregor, Anthony Wirth</i> | |
| LEARNING SCALE-FREE NETWORKS BY DYNAMIC NODE SPECIFIC DEGREE PRIOR | 2237 |
| <i>Qingming Tang, Siqi Sun, Jinbo Xu</i> | |
| DEEP UNSUPERVISED LEARNING USING NONEQUILIBRIUM THERMODYNAMICS | 2246 |
| <i>Jascha Sohl-Dickstein, Eric Weiss, Niru Maheswaranathan, Surya Ganguli</i> | |
| MODELING ORDER IN NEURAL WORD EMBEDDINGS AT SCALE | 2256 |
| <i>Andrew Trask, David Gilmore, Matthew Russell</i> | |
| DISTRIBUTED INFERENCE FOR DIRICHLET PROCESS MIXTURE MODELS | 2266 |
| <i>Hong Ge, Yutian Chen, Moquan Wan, Zoubin Ghahramani</i> | |
| COMPRESSING NEURAL NETWORKS WITH THE HASHING TRICK | 2275 |
| <i>Wenlin Chen, James Wilson, Stephen Tyree, Kilian Weinberger, Yixin Chen</i> | |
| INTERSECTING FACES: NON-NEGATIVE MATRIX FACTORIZATION WITH NEW GUARANTEES | 2285 |
| <i>Rong Ge, James Zou</i> | |
| SCALING UP NATURAL GRADIENT BY SPARSELY FACTORIZING THE INVERSE FISHER MATRIX | 2294 |
| <i>Roger Grosse, Ruslan Salakhudinov</i> | |
| A DEEPER LOOK AT PLANNING AS LEARNING FROM REPLAY | 2304 |
| <i>Harm Vanseijen, Rich Sutton</i> | |
| OPTIMAL AND ADAPTIVE ALGORITHMS FOR ONLINE BOOSTING | 2313 |
| <i>Alina Beygelzimer, Satyen Kale, Haipeng Luo</i> | |
| GLOBAL CONVERGENCE OF STOCHASTIC GRADIENT DESCENT FOR SOME NON-CONVEX MATRIX PROBLEMS | 2322 |
| <i>Christopher De Sa, Christopher Re, Kunle Olukotun</i> | |
| AN EMPIRICAL EXPLORATION OF RECURRENT NETWORK ARCHITECTURES | 2332 |
| <i>Rafal Jozefowicz, Wojciech Zaremba, Ilya Sutskever</i> | |
| COMPLETE DICTIONARY RECOVERY USING NONCONVEX OPTIMIZATION | 2341 |
| <i>Ju Sun, Qing Qu, John Wright</i> | |
| SAFE POLICY SEARCH FOR LIFELONG REINFORCEMENT LEARNING WITH SUBLINEAR REGRET | 2351 |
| <i>Haitham Bou Ammar, Rasul Tutunov, Eric Eaton</i> | |
| PASSCODE: PARALLEL ASYNCHRONOUS STOCHASTIC DUAL CO-ORDINATE DESCENT | 2360 |
| <i>Cho-Jui Hsieh, Hsiang-Fu Yu, Inderjit Dhillon</i> | |
| HIGH CONFIDENCE POLICY IMPROVEMENT | 2370 |
| <i>Philip Thomas, Georgios Theodorou, Mohammad Ghavamzadeh</i> | |
| FIXED-POINT ALGORITHMS FOR LEARNING DETERMINANTAL POINT PROCESSES | 2379 |
| <i>Zelda Mariet, Suvrit Sra</i> | |
| CONSISTENT MULTICLASS ALGORITHMS FOR COMPLEX PERFORMANCE MEASURES | 2388 |
| <i>Harikrishna Narasimhan, Harish Ramaswamy, Aadirupa Saha, Shivani Agarwal</i> | |
| OPTIMIZING NEURAL NETWORKS WITH KRONECKER-FACTORED APPROXIMATE CURVATURE | 2398 |
| <i>James Martens, Roger Grosse</i> | |

| | |
|---|------|
| A CONVEX EXEMPLAR-BASED APPROACH TO MAD-BAYES DIRICHLET PROCESS MIXTURE MODELS..... | 2408 |
| <i>En-Hsu Yen, Xin Lin, Kai Zhong, Pradeep Ravikumar, Inderjit Dhillon</i> | |
| MULTI-INSTANCE MULTI-LABEL LEARNING IN THE PRESENCE OF NOVEL CLASS INSTANCES | 2417 |
| <i>Anh Pham, Raviv Raich, Xiaoli Fern, Jesús Pérez Arriaga</i> | |
| ENTROPY-BASED CONCENTRATION INEQUALITIES FOR DEPENDENT VARIABLES..... | 2426 |
| <i>Liva Ralaivola, Massih-Reza Amini</i> | |
| PU LEARNING FOR MATRIX COMPLETION | 2435 |
| <i>Cho-Jui Hsieh, Nagarajan Natarajan, Inderjit Dhillon</i> | |
| AN ASYNCHRONOUS DISTRIBUTED PROXIMAL GRADIENT METHOD FOR COMPOSITE CONVEX OPTIMIZATION | 2444 |
| <i>Necdet Aybat, Zi Wang, Garud Iyengar</i> | |
| SPARSE SUBSPACE CLUSTERING WITH MISSING ENTRIES | 2453 |
| <i>Congyuan Yang, Daniel Robinson, Rene Vidal</i> | |
| MODERATED AND DRIFTING LINEAR DYNAMICAL SYSTEMS..... | 2463 |
| <i>Jinyan Guan, Kyle Simek, Ernesto Brau, Clayton Morrison, Emily Butler, Kobus Barnard</i> | |
| BOOSTED CATEGORICAL RESTRICTED BOLTZMANN MACHINE FOR COMPUTATIONAL PREDICTION OF SPLICE JUNCTIONS..... | 2473 |
| <i>Taehoon Lee, Sungroh Yoon</i> | |
| PRIVACY FOR FREE: POSTERIOR SAMPLING AND STOCHASTIC GRADIENT MONTE CARLO..... | 2483 |
| <i>Yu-Xiang Wang, Stephen Fienberg, Alex Smola</i> | |
| A TRUST-REGION METHOD FOR STOCHASTIC VARIATIONAL INFERENCE WITH APPLICATIONS TO STREAMING DATA..... | 2493 |
| <i>Lucas Theis, Matt Hoffman</i> | |
| INFERENCE IN A PARTIALLY OBSERVED QUEUING MODEL WITH APPLICATIONS IN ECOLOGY..... | 2502 |
| <i>Kevin Winner, Garrett Bernstein, Dan Sheldon</i> | |
| DETERMINISTIC INDEPENDENT COMPONENT ANALYSIS..... | 2511 |
| <i>Ruitong Huang, Andras Gyorgy, Csaba Szepesvári</i> | |
| ON THE OPTIMALITY OF MULTI-LABEL CLASSIFICATION UNDER SUBSET ZERO-ONE LOSS FOR DISTRIBUTIONS SATISFYING THE COMPOSITION PROPERTY | 2521 |
| <i>Maxime Gasse, Alexandre Aussem, Haytham Elghazel</i> | |
| UN-REGULARIZING: APPROXIMATE PROXIMAL POINT AND FASTER STOCHASTIC ALGORITHMS FOR EMPIRICAL RISK MINIMIZATION..... | 2530 |
| <i>Roy Frostig, Rong Ge, Sham Kakade, Aaron Sidford</i> | |
| A NEW GENERALIZED ERROR PATH ALGORITHM FOR MODEL SELECTION..... | 2539 |
| <i>Bin Gu, Charles Ling</i> | |
| Author Index | |