

Advances in Neural Information Processing Systems 28

29th Annual Conference on Neural
Information Processing Systems 2015

Montreal, Canada
7-12 December 2015

Volume 1 of 4

Editors:

Corinna Cortes
Daniel D. Lee
Roman Garnett

Neil D. Lawrence
Masashi Sugiyama

ISBN: 978-1-5108-2502-4

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© (2015) by Neural Information Processing Systems
All rights reserved.

Printed by Curran Associates, Inc. (2016)

For permission requests, please contact Neural Information Processing Systems
at the address below.

Neural Information Processing Systems
10010 North Torrey Pines Road
La Jolla, CA 92037
USA

Phone: (858) 453-1623
Fax: (858) 587-0417

info@nips.cc

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

Contents

Contents	iii
Preface	xxxix
Donors	xxxv
NIPS foundation	xxxvii
Committees	xxxviii
Reviewers	xli
Double or Nothing: Multiplicative Incentive Mechanisms for Crowdsourcing, NIHAR SHAH, UC Berkeley, DENG YONG ZHOU, Microsoft Research	1
Learning with Symmetric Label Noise: The Importance of Being Unhinged, BRENDAN ROOYEN, ADITYA MENON, and ROBERT WILLIAMSON, NICTA	10
Algorithmic Stability and Uniform Generalization, IBRAHIM ALABDULMOHSIN, King Abdullah University of Science and Technology	19
Adaptive Low-Complexity Sequential Inference for Dirichlet Process Mixture Models, THEODOROS TSILIGKARIDIS, KEITH FORSYTHE, MIT Lincoln Laboratory	28
Covariance-Controlled Adaptive Langevin Thermostat for Large-Scale Bayesian Sampling, XIAOCHENG SHANG, ZHANXING ZHU, and BENEDICT LEIMKUHLER, AMOS STORKEY, University of Edinburgh	37
Robust Portfolio Optimization, HUITONG QIU, FANG HAN, Johns Hopkins University, and HAN LIU, Princeton University, BRIAN CAFFO, Johns Hopkins University	46
Logarithmic Time Online Multiclass prediction, ANNA CHOROMANSKA, Courant Institute NYU, JOHN LANGFORD, Microsoft Research New York	55
Planar Ultrametrics for Image Segmentation, JULIAN YARKONY, Dr., CHARLESS FOWLKES, UC Irvine	64
Expressing an Image Stream with a Sequence of Natural Sentences, CESC PARK, GUNHEE KIM, Seoul National University	73
Parallel Correlation Clustering on Big Graphs, XINGHAO PAN, DIMITRIS PAPALIOPOULOS, and BENJAMIN RECHT, KANNAN RAMCHANDRAN, MICHAEL JORDAN, UC Berkeley	82

Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks , SHAOQING REN, USTC, KAIMING HE, Microsoft Research Asia, and ROSS GIRSHICK, Microsoft Research, JIAN SUN, Microsoft Research Asia	91
Space-Time Local Embeddings , KE SUN, University of Geneva, JUN WANG, Expedia Geneva, and ALEXANDROS KALOUSIS, STEPHANE MARCHAND-MAILLET, University of Geneva	100
A Convergent Gradient Descent Algorithm for Rank Minimization and Semidefinite Programming from Random Linear Measurements , QINQING ZHENG, JOHN LAFFERTY, University of Chicago	109
Smooth Interactive Submodular Set Cover , BRYAN HE, Stanford University, YISONG YUE, Caltech	118
Galileo: Perceiving Physical Object Properties by Integrating a Physics Engine with Deep Learning , JIAJUN WU, ILKER YILDIRIM, and JOSEPH LIM, WILLIAM FREEMAN, JOSH TENENBAUM, Massachusetts Institute of Technology	127
On the Pseudo-Dimension of Nearly Optimal Auctions , JAMIE MORGENSTERN, University of Pennsylvania, TIM ROUGHGARDEN, Stanford University	136
Unlocking neural population non-stationarities using hierarchical dynamics models , MIJUNG PARK, University College London, GERGO BOHNER, Gatsby Unit UCL, and JAKOB MACKE, research center caesar & BCCN Tubingen	145
Bayesian Manifold Learning: The Locally Linear Latent Variable Model (LL-LVM) , MIJUNG PARK, University College London, WITTAWAT JITKRITTUM, Gatsby Unit UCL, and AHMAD QAMAR, ZOLTAN SZABO, Gatsby Unit UCL, LARS BUESING, MANEESH SAHANI	154
Color Constancy by Learning to Predict Chromaticity from Luminance , AYAN CHAKRABARTI, TTI Chicago	163
Fast and Accurate Inference of Plackett-Luce Models , LUCAS MAYSTRE, MATTHIAS GROSSGLAUSER, École polytechnique fédérale de Lausanne	172
Probabilistic Line Searches for Stochastic Optimization , MAREN MAHSERECI, PHILIPP HENNIG, MPI for Intelligent Systems Tübingen	181
Inferring Algorithmic Patterns with Stack-Augmented Recurrent Nets , ARMAND JOULIN, Facebook AI research, TOMAS MIKOLOV, Facebook AI Research	190
Where are they looking? , ADRIA RECASENS, ADITYA KHOSLA, and CARL VONDRICK, ANTONIO TORRALBA, Massachusetts Institute of Technology	199
The Pareto Regret Frontier for Bandits , TOR LATTIMORE, University of Alberta	208
On the Limitation of Spectral Methods: From the Gaussian Hidden Clique Problem to Rank-One Perturbations of Gaussian Tensors , ANDREA MONTANARI, Stanford University, DANIEL REICHMAN,	

Cornell University, and OFER ZEITOUNI, Weizmann Institute and Courant Institute	217
Measuring Sample Quality with Stein’s Method, JACKSON GORHAM, LESTER MACKEY, Stanford University	226
Bidirectional Recurrent Convolutional Networks for Multi-Frame Super-Resolution, YAN HUANG, CRIPAC CASIA, WEI WANG, NLPRCASIA, and LIANG WANG	235
Bounding errors of Expectation-Propagation, GUILLAUME DEHAENE, University of Geneva, SIMON BARTHELMÉ, Gipsa-lab CNRS	244
A fast, universal algorithm to learn parametric nonlinear embeddings, MIGUEL CARREIRA-PERPINAN, UC Merced, MAX VLADYMYROV, Yahoo	253
Texture Synthesis Using Convolutional Neural Networks, LEON GATYS, University of Tübingen, ALEXANDER ECKER, University of Tuebingen, and MATTHIAS BETHGE, CIN University Tübingen	262
Extending Gossip Algorithms to Distributed Estimation of U-statistics, IGOR COLIN, Télécom ParisTech, AURÉLIEN BELLET, Telecom ParisTech, and JOSEPH SALMON, Télécom ParisTech, STÉPHAN CLÉMENÇON, Telecom ParisTech	271
Streaming, Distributed Variational Inference for Bayesian Nonparametrics, TREVOR CAMPBELL, JULIAN STRAUB, and JOHN FISHER III, JONATHAN HOW, Massachusetts Institute of Technology	280
Learning visual biases from human imagination, CARL VONDRICK, Massachusetts Institute of Technology, HAMED PIRSIYAVASH, UMBC, and AUDE OLIVA, ANTONIO TORRALBA, Massachusetts Institute of Technology	289
Smooth and Strong: MAP Inference with Linear Convergence, OFER MESHİ, MEHRDAD MAHDAVI, TTI Chicago, and ALEX SCHWING, University of Toronto	298
Copeland Dueling Bandits, MASROUR ZOGHI, University of Amsterdam, ZOHAR KARNIN, and SHIMON WHITESON, University of Oxford, MAARTEN RIJKE, University of Amsterdam	307
Optimal Ridge Detection using Coverage Risk, YEN-CHI CHEN, CHRISTOPHER GENOVESE, and SHIRLEY HO, LARRY WASSERMAN, Carnegie Mellon University	316
Top-k Multiclass SVM, MAKSİM LAPIN, Max Planck Institute for Informatics, MATTHIAS HEIN, Saarland University, and BERNT SCHIELE, Max Planck Institute for Informatics	325
Policy Evaluation Using the Ω-Return, PHILIP THOMAS, University of Massachusetts Amherst Carnegie Mellon University, SCOTT NIEKUM, UT Austin, and GEORGIOS THEOCHAROUS, Adobe, GEORGE KONIDARIS, Duke ..	334
Orthogonal NMF through Subspace Exploration, MEGASTHENIS ASTERIS, University of Texas at Austin, DIMITRIS PAPALIOPOULOS, UC Berkeley, and ALEX DIMAKIS, Utaustin	343
Stochastic Online Greedy Learning with Semi-bandit Feedbacks, TIAN LIN, JIAN LI, Tsinghua University, and WEI CHEN, Microsoft.com	352

Deeply Learning the Messages in Message Passing Inference, GUOSHENG LIN, The University of Adelaide, CHUNHUA SHEN, and IAN REID, ANTON HENGEL, University of Adelaide	361
Synaptic Sampling: A Bayesian Approach to Neural Network Plasticity and Rewiring, DAVID KAPPEL, Graz University of Technology, STEFAN HABENSCHUSS, and ROBERT LEGENSTEIN, WOLFGANG MAASS	370
Accelerated Proximal Gradient Methods for Nonconvex Programming, HUAN LI, ZHOUCHE LIN, Peking University	379
Approximating Sparse PCA from Incomplete Data, ABHISEK KUNDU, Rensselaer Polytechnic Institute, PETROS DRINEAS, and MALIK MAGDON-ISMAIL, Rensselaer Polytechnic Institute	388
Nonparametric von Mises Estimators for Entropies, Divergences and Mutual Informations, KIRTHEVASAN KANDASAMY, AKSHAY KRISHNAMURTHY, and BARNABAS POZOS, LARRY WASSERMAN, Carnegie Mellon University, JAMES ROBINS, Harvard University	397
Column Selection via Adaptive Sampling, SAURABH PAUL, Paypal Inc, MALIK MAGDON-ISMAIL, and PETROS DRINEAS, Rensselaer Polytechnic Institute	406
HONOR: Hybrid Optimization for Non-convex Regularized problems, PINGHUA GONG, University of Michigan-Ann Arbor, JIEPING YE, University of Michigan	415
3D Object Proposals for Accurate Object Class Detection, XIAOZHI CHEN, Tsinghua University, KAUSTAV KUNDU, and YUKUN ZHU, ANDREW BERNESHAWI, University of Toronto, HUIMIN MA, Tsinghua University, SANJA FIDLER, RAQUEL URTASUN, University of Toronto	424
Algorithms with Logarithmic or Sublinear Regret for Constrained Contextual Bandits, HUASEN WU, University of California at Davis, R. SRIKANT, University of Illinois at Urbana-Champaign, and XIN LIU, University of California Davis, CHONG JIANG, University of Illinois at Urbana-Champaign ..	433
Tensorizing Neural Networks, ALEXANDER NOVIKOV, DMITRY PODOPRIKHIN, Skolkovo Institute of Science and Technology, and ANTON OSOKIN, Inria, DMITRY VETROV, Skoltech Moscow	442
Parallelizing MCMC with Random Partition Trees, XIANGYU WANG, FANGJIAN GUO, and KATHERINE HELLER, DAVID DUNSON, Duke University	451
A Reduced-Dimension fMRI Shared Response Model, PO-HSUAN CHEN, Princeton, JANICE CHEN, and YAARA YESHURUN, URI HASSON, Princeton University, JAMES HAXBY, PETER RAMADGE, Princeton	460
Spectral Learning of Large Structured HMMs for Comparative Epigenomics, CHICHENG ZHANG, UC San Diego, JIMIN SONG, Rutgers, and KAMALIKA CHAUDHURI, University of California San Diego, KEVIN CHEN, Rutgers	469
Individual Planning in Infinite-Horizon Multiagent Settings: Inference, Structure and Scalability, XIA QU, Epic Systems, PRASHANT DOSHI, University of Georgia	478
Estimating Mixture Models via Mixtures of Polynomials, SIDA WANG, ARUN CHAGANTY, and PERCY LIANG, Stanford University	487

On the Global Linear Convergence of Frank-Wolfe Optimization Variants , SIMON LACOSTE-JULIEN, INRIA, MARTIN JAGGI, Eidgenössische Technische Hochschule Zürich	496
Deep Knowledge Tracing , CHRIS PIECH, Stanford University, JONATHAN BASSEN, stanford.edu, and JONATHAN HUANG, google.com, SURYA GANGULI, MEHRAN SAHAMI, LEONIDAS GUIBAS, JASCHA SOHL-DICKSTEIN, stanford.edu	505
Rethinking LDA: Moment Matching for Discrete ICA , ANASTASIA PODOSINNIKOVA, INRIA/ENS, FRANCIS BACH, INRIA - ENS, and SIMON LACOSTE-JULIEN, INRIA	514
Efficient Compressive Phase Retrieval with Constrained Sensing Vectors , SOHAIL BAHMANI, Georgia Tech., JUSTIN ROMBERG, Georgia Institute of Technology	523
Barrier Frank-Wolfe for Marginal Inference , RAHUL KRISHNAN, New York University, SIMON LACOSTE-JULIEN, INRIA, and DAVID SONTAG, New York University	532
Learning Theory and Algorithms for Forecasting Non-stationary Time Series , VITALY KUZNETSOV, Courant Institute, MEHRYAR MOHRI, Courant Institute and Google	541
Compressive spectral embedding: sidestepping the SVD , DINESH RAMASAMY, UPAMANYU MADHOW, UC Santa Barbara	550
A Nonconvex Optimization Framework for Low Rank Matrix Estimation , TUO ZHAO, Johns Hopkins University, ZHAORAN WANG, and HAN LIU, Princeton University	559
Automatic Variational Inference in Stan , ALP KUCUKELBIR, RAJESH RANGANATH, Princeton University, and ANDREW GELMAN, DAVID BLEI, Columbia University	568
Attention-Based Models for Speech Recognition , JAN CHOROWSKI, University of Wroclaw, DZMITRY BAHDANAU, Jacobs University Germany, and DMITRIY SERDYUK, Université de Montréal, KYUNGHYUN CHO, New York University, YOSHUA BENGIO, U. Montreal	577
Closed-form Estimators for High-dimensional Generalized Linear Models , EUNHO YANG, IBM Thomas J. Watson Research Center, AURELIE LOZANO, IBM Research, and PRADEEP RAVIKUMAR, University of Texas at Austin	586
Online F-Measure Optimization , RÓBERT BUSA-FEKETE, UPB, BALAZS SZORENYI, The Technion / University of Szeged, and KRZYSZTOF DEMBCZYNSKI, Poznan University of Technology, EYKE HULLERMEIER, Marburguniversity	595
Online Rank Elicitation for Plackett-Luce: A Dueling Bandits Approach , BALAZS SZORENYI, The Technion / University of Szeged, RÓBERT BUSA-FEKETE, and ADIL PAUL, UPB, EYKE HULLERMEIER, Marburguniversity	604
M-Best-Diverse Labelings for Submodular Energies and Beyond , ALEXANDER KIRILLOV, DMITRIJ SCHLESINGER, TU Dresden, and DMITRY VETROV, Skoltech Moscow, CARSTEN ROTHER, BOGDAN SAVCHYNSKY, TU Dresden	613

Tractable Bayesian Network Structure Learning with Bounded Vertex Cover Number, JANNE KORHONEN, University of Helsinki, PEKKA PARVIAINEN, Aalto University	622
Learning Large-Scale Poisson DAG Models based on OverDispersion Scoring, GUNWOONG PARK, UW Madison, GARVESH RASKUTTI, University of Wisconsin Madison	631
Training Restricted Boltzmann Machine via the Thouless-Anderson-Palmer free energy, MARYLOU GABRIÉ, École Normale Supérieure ICFP Laboratoire de Physique Statistique, ERIC TRAMEL, LPS École Normale Supérieure, and FLORENT KRZAKALA, Ecole Normale Supérieure CNRS	640
Character-level Convolutional Networks for Text Classification, XIANG ZHANG, JUNBO ZHAO, and YANN LECUN, New York University	649
Robust Feature-Sample Linear Discriminant Analysis for Brain Disorders Diagnosis, EHSAN ADELI-MOSABBEB, KIM-HAN THUNG, and LE AN, FENG SHI, DINGGANG SHEN, UNC-Chapel Hill	658
Black-box optimization of noisy functions with unknown smoothness, JEAN-BASTIEN GRILL, MICHAL VALKO, INRIA Lille - Nord Europe, and REMI MUNOS, Google DeepMind	667
Recovering Communities in the General Stochastic Block Model Without Knowing the Parameters, EMMANUEL ABBE, COLIN SANDON, Princeton University	676
Deep learning with Elastic Averaging SGD, SIXIN ZHANG, New York University, ANNA CHOROMANSKA, Courant Institute NYU, and YANN LECUN, New York University	685
Monotone k-Submodular Function Maximization with Size Constraints, NAOTO OHSAKA, The University of Tokyo, YUICHI YOSHIDA, National Institute of Informatics and Preferred Infrastructure Inc.	694
Active Learning from Weak and Strong Labelers, CHICHENG ZHANG, UC San Diego, KAMALIKA CHAUDHURI, University of California San Diego	703
On the Optimality of Classifier Chain for Multi-label Classification, WEIWEI LIU, UTS, IVOR TSANG, University of Technology Sydney	712
Robust Regression via Hard Thresholding, KUSH BHATIA, PRATEEK JAIN, Microsoft Research, and PURUSHOTTAM KAR, Microsoft Research India ..	721
Sparse Local Embeddings for Extreme Multi-label Classification, KUSH BHATIA, Microsoft Research, HIMANSHU JAIN, IIT Delhi, and PURUSHOTTAM KAR, MANIK VARMA, Microsoft Research India, PRATEEK JAIN, Microsoft Research	730
Solving Random Quadratic Systems of Equations Is Nearly as Easy as Solving Linear Systems, YUXIN CHEN, EMMANUEL CANDES, Stanford University	739
A Framework for Individualizing Predictions of Disease Trajectories by Exploiting Multi-Resolution Structure, PETER SCHULAM, SUCHI SARIA, Johns Hopkins University	748

Subspace Clustering with Irrelevant Features via Robust Dantzig Selector , CHAO QU, HUAN XU, National University of Singapore	757
Sparse PCA via Bipartite Matchings , MEGASTHENIS ASTERIS, University of Texas at Austin, DIMITRIS PAPALIOPOULOS, UC Berkeley, and ANASTASIOS KYRILIDIS, University of Texas at Austin, ALEX DIMAKIS, Utaustin	766
Fast Randomized Kernel Ridge Regression with Statistical Guarantees , AHMED ALAOU, MICHAEL MAHONEY, UC Berkeley	775
Online Learning for Adversaries with Memory: Price of Past Mistakes , OREN ANAVA, Technion, ELAD HAZAN, Princeton University, and SHIE MANNOR, Technion	784
Convolutional spike-triggered covariance analysis for neural subunit models , ANQI WU, Princeton University, IL PARK, Stony Brook University, and JONATHAN PILLOW, Princeton University	793
Convolutional LSTM Network: A Machine Learning Approach for Precipitation Nowcasting , XINGJIAN SHI, ZHOURONG CHEN, and HAO WANG, DIT-YAN YEUNG, HKUST, WAI-KIN WONG, WANG-CHUN WOO	802
GAP Safe screening rules for sparse multi-task and multi-class models , EUGENE NDIAYE, Institut Mines-Télécom Télécom ParisTech CNRS LTCI, OLIVIER FERCOQ, Telecom ParisTech, and ALEXANDRE GRAMFORT, Telecom Paristech, JOSEPH SALMON, Télécom ParisTech	811
Empirical Localization of Homogeneous Divergences on Discrete Sample Spaces , TAKASHI TAKENOUCI, Future University Hakodate, TAKAFUMI KANAMORI, Nagoya University	820
Statistical Model Criticism using Kernel Two Sample Tests , JAMES LLOYD, ZOUBIN GHAHRAMANI, University of Cambridge	829
Precision-Recall-Gain Curves: PR Analysis Done Right , PETER FLACH, MEELIS KULL, University of Bristol	838
A Generalization of Submodular Cover via the Diminishing Return Property on the Integer Lattice , TASUKU SOMA, University of Tokyo, YUICHI YOSHIDA, National Institute of Informatics and Preferred Infrastructure Inc.	847
Bidirectional Recurrent Neural Networks as Generative Models , MATHIAS BERGLUND, Aalto University, TAPANI RAIKO, Aalto University The Curious AI Company, and MIKKO HONKALA, LEO KÄRKKÄINEN, AKOS VETEK, Nokia Labs, JUHA KARHUNEN, Aalto University	856
Quartz: Randomized Dual Coordinate Ascent with Arbitrary Sampling , ZHENG QU, University of Hong Kong, PETER RICHTARIK, University of Edinburgh, and TONG ZHANG, Rutgers	865
Maximum Likelihood Learning With Arbitrary Treewidth via Fast-Mixing Parameter Sets , JUSTIN DOMKE, NICTA	874
Hessian-free Optimization for Learning Deep Multidimensional Recurrent Neural Networks , MINH YUNG CHO, CHANDRA DHIR, and JAEHYUNG LEE, Gracenote	883
Large-scale probabilistic predictors with and without guarantees of validity , VLADIMIR VOVK, Royal Holloway Univ of London, IVAN PETEJ, and VALENTINA FEDOROVA, Yandex	892

Shepard Convolutional Neural Networks , JIMMY REN, LI XU, and QIONG YAN, WENXIU SUN, SenseTime Group Limited	901
Matrix Manifold Optimization for Gaussian Mixtures , RESHAD HOSSEINI, University of Tehran, SUVRIT SRA, Massachusetts Institute of Technology	910
Semi-supervised Convolutional Neural Networks for Text Categorization via Region Embedding , RIE JOHNSON, RJ Research Consulting, TONG ZHANG, Rutgers	919
Parallel Recursive Best-First AND/OR Search for Exact MAP Inference in Graphical Models , AKIHIRO KISHIMOTO, IBM Research, RADU MARINESCU, IBM Research Ireland, and ADI BOTEA, IBM Research	928
Convolutional Neural Networks with Intra-Layer Recurrent Connections for Scene Labeling , MING LIANG, XIAOLIN HU, and BO ZHANG, Tsinghua University	937
Bounding the Cost of Search-Based Lifted Inference , DAVID SMITH, University of Texas at Dallas, VIBHAV GOGATE, UT Dallas	946
Gradient-free Hamiltonian Monte Carlo with Efficient Kernel Exponential Families , HEIKO STRATHMANN, University College London, DINO SEJGINOVIC, University of Oxford, and SAMUEL LIVINGSTONE, University College London, ZOLTAN SZABO, Gatsby Unit UCL, ARTHUR GRETTON, University Collage London	955
Linear Multi-Resource Allocation with Semi-Bandit Feedback , TOR LATTIMORE, University of Alberta, KOBY CRAMMER, Technion, and CSABA SZEPESVARI, University of Alberta	964
Unsupervised Learning by Program Synthesis , KEVIN ELLIS, ARMANDO SOLAR-LEZAMA, and JOSH TENENBAUM, Massachusetts Institute of Technology	973
Enforcing balance allows local supervised learning in spiking recurrent networks , RALPH BOURDOUKAN, Ecole Normale Superieure, SOPHIE DENÈVE, GNT Ecole Normale Superieure	982
Fast and Guaranteed Tensor Decomposition via Sketching , YINING WANG, HSIAO-YU TUNG, and ALEX SMOLA, Carnegie Mellon University, ANIMA ANANDKUMAR, UC Irvine	991
Differentially private subspace clustering , YINING WANG, YU-XIANG WANG, and AARTI SINGH, Carnegie Mellon University	1000
Predtron: A Family of Online Algorithms for General Prediction Problems , PRATEEK JAIN, Microsoft Research, NAGARAJAN NATARAJAN, UT Austin, and AMBUJ TEWARI, University of Michigan	1009
Weighted Theta Functions and Embeddings with Applications to Max-Cut, Clustering and Summarization , FREDRIK JOHANSSON, Chalmers University Sweden, ANKANI CHATTORAJ, Chalmers University, and CHIRANJIB BHATTACHARYYA, Indian Institute of Science, DEVDATT DUBHASHI, Chalmers University Sweden	1018
SGD Algorithms based on Incomplete U-statistics: Large-Scale Minimization of Empirical Risk , GUILLAUME PAPA, Telecom paristech, STÉPHAN CLÉMENÇON, and AURÉLIEN BELLET, Telecom ParisTech	1027

On Top-k Selection in Multi-Armed Bandits and Hidden Bipartite Graphs, WEI CAO, JIAN LI, Tsinghua University, and YUFEI TAO, The Chinese University of Hong Kong, ZHIZE LI, Tsinghua University	1036
The Brain Uses Reliability of Stimulus Information when Making Perceptual Decisions, SEBASTIAN BITZER, STEFAN KIEBEL, TU Dresden	1045
Fast Classification Rates for High-dimensional Gaussian Generative Models, TIANYANG LI, ADARSH PRASAD, UT Austin, and PRADEEP RAVIKUMAR, University of Texas at Austin	1054
Fast Distributed k-Center Clustering with Outliers on Massive Data, GUSTAVO MALKOMES, Washington University in St. Louis, MATT KUSNER, Washington University in STL, and WENLIN CHEN, KILIAN WEINBERGER, Washington University in St. Louis, BENJAMIN MOSELEY, Washington University in St Lo	1063
Human Memory Search as Initial-Visit Emitting Random Walk, KWANG-SUNG JUN, XIAOJIN ZHU, and TIMOTHY ROGERS, University of Wisconsin-Madison, ZHUORAN YANG, Tsinghua University, MING YUAN, University of Wisconsin - Madison	1072
Non-convex Statistical Optimization for Sparse Tensor Graphical Model, WEI SUN, Yahoo Labs, ZHAORAN WANG, and HAN LIU, Princeton University, GUANG CHENG, Purdue University	1081
Convergence Rates of Active Learning for Maximum Likelihood Estimation, KAMALIKA CHAUDHURI, University of California San Diego, SHAM KAKADE, University of Washington, and PRANEETH NETRAPALLI, Microsoft Research, SUJAY SANGHAVI, UTexas-Austin	1090
Weakly-supervised Disentangling with Recurrent Transformations for 3D View Synthesis, JIMEI YANG, UC Merced, SCOTT REED, University of Michigan, and MING-HSUAN YANG, UC Merced, HONGLAK LEE, U. Michigan	1099
Efficient Exact Gradient Update for training Deep Networks with Very Large Sparse Targets, PASCAL VINCENT, U. Montreal, ALEXANDRE BRÉBISSE, Université de Montréal, and XAVIER BOUTHILLIER, Universit de Montréal	1108
Backpropagation for Energy-Efficient Neuromorphic Computing, STEVE ESSER, RATHINAKUMAR APPUSWAMY, and PAUL MEROLLA, JOHN ARTHUR, DHARMENDRA MODHA, IBM Research-Almaden	1117
Alternating Minimization for Regression Problems with Vector-valued Outputs, PRATEEK JAIN, Microsoft Research, AMBUJ TEWARI, University of Michigan	1126
Learning both Weights and Connections for Efficient Neural Network, SONG HAN, Stanford University, JEFF POOL, and JOHN TRAN, NVIDIA, BILL DALLY, Stanford University	1135
Optimal Rates for Random Fourier Features, BHARATH SRIPERUMBUDUR, The Pennsylvania State University, ZOLTAN SZABO, Gatsby Unit UCL	1144
The Population Posterior and Bayesian Modeling on Streams, JAMES MCINERNEY, Columbia, RAJESH RANGANATH, Princeton University, and DAVID BLEI, Columbia University	1153

Frank-Wolfe Bayesian Quadrature: Probabilistic Integration with Theoretical Guarantees , FRANÇOIS-XAVIER BRIOL, University of Warwick, CHRIS. OATES, University of Tech. Sydney, and MARK GIROLAMI, University of Warwick, MIKE OSBORNE, U Oxford	1162
Scheduled Sampling for Sequence Prediction with Recurrent Neural Networks , SAMY BENGIO, Google Research, ORIOL VINYALS, and NAVDEEP JAITLEY, NOAM SHAZEER, Google	1171
Unified View of Matrix Completion under General Structural Constraints , SURIYA GUNASEKAR, UT Austin, ARINDAM BANERJEE, University of Minnesota, and JOYDEEP GHOSH, UT Austin	1180
Efficient Output Kernel Learning for Multiple Tasks , PRATIK JAWANPURIA, Saarlanduniversity, MAKSIM LAPIN, Max Planck Institute for Informatics, and MATTHIAS HEIN, Saarland University, BERNT SCHIELE, Max Planck Institute for Informatics	1189
Scalable Adaptation of State Complexity for Nonparametric Hidden Markov Models , MICHAEL HUGHES, WILLIAM STEPHENSON, and ERIK SUDDERTH, Brown University	1198
Variational Consensus Monte Carlo , MAXIM RABINOVICH, UC Berkeley, ELAINE ANGELINO, Harvard, and MICHAEL JORDAN, UC Berkeley .	1207
Newton-Stein Method: A Second Order Method for GLMs via Stein’s Lemma , MURAT A. ERDOGDU, Stanford University	1216
Practical and Optimal LSH for Angular Distance , ALEXANDR ANDONI, Columbia, PIOTR INDYK, Massachusetts Institute of Technology, and THUIS LAARHOVEN, TU/e, ILYA RAZENSHTEYN, LUDWIG SCHMIDT, Massachusetts Institute of Technology	1225
Learning to Linearize Under Uncertainty , ROSS GOROSHIN, MICHAEL MATHIEU, and YANN LECUN, New York University	1234
Finite-Time Analysis of Projected Langevin Monte Carlo , SEBASTIEN BUBECK, Microsoft Research, RONEN ELKAN, and JOSEPH LEHEC	1243
Deep Visual Analogy-Making , SCOTT REED, YI ZHANG, and YUTING ZHANG, University of Michigan, HONGLAK LEE, U. Michigan	1252
Matrix Completion from Fewer Entries: Spectral Detectability and Rank Estimation , ALAA SAADE, École normale supérieure, FLORENT KRZAKALA, Ecole Normale Supérieure CNRS, and LENKA ZDEBOROVÁ, CEA	1261
Online Learning with Adversarial Delays , KENT QUANRUD, DANIEL KHASHABI, University of Illinois at Urbana-Champaign	1270
Multi-Layer Feature Reduction for Tree Structured Group Lasso via Hierarchical Projection , JIE WANG, University of Michigan-Arbor, JIEPING YE, University of Michigan	1279
Minimum Weight Perfect Matching via Blossom Belief Propagation , SUNGSOO AHN, SEJUN PARK, KAIST, and MICHAEL CHERTKOV, JINWOO SHIN, KAIST	1288
Efficient Thompson Sampling for Online Matrix-Factorization Recommendation , JAYA KAWALE, HUNG BUI, and BRANISLAV KVETON, Adobe Research, LONG TRAN-THANH, University of Southampton, SANJAY CHAWLA, Qatar Computing Research Institute HBKU and University of Sydney	1297

Improved Iteration Complexity Bounds of Cyclic Block Coordinate Descent for Convex Problems, RUOYU SUN, Stanford university, MINGYI HONG	1306
Lifted Symmetry Detection and Breaking for MAP Inference, TIMOTHY KOPP, University of Rochester, PARAG SINGLA, Indian Institute of Technology, and HENRY KAUTZ, University of Rochester	1315
Evaluating the statistical significance of biclusters, JASON LEE, YUEKAI SUN, and JONATHAN TAYLOR, Stanford University	1324
Discriminative Robust Transformation Learning, JIAJI HUANG, QIANG QIU, and GUILLERMO SAPIRO, ROBERT CALDERBANK, Duke University	1333
Bandits with Unobserved Confounders: A Causal Approach, ELIAS BAREINBOIM, ANDREW FORNEY, and JUDEA PEARL, University of California Los Angeles	1342
Scalable Semi-Supervised Aggregation of Classifiers, AKSHAY BALSUBRAMANI, University of California San Diego, YOAV FREUND, UC San Diego	1351
Online Learning with Gaussian Payoffs and Side Observations, YIFAN WU, University of Alberta, ANDRAS GYORGY, Imperial College London, and CSABA SZEPESVARI, University of Alberta	1360
Private Graphon Estimation for Sparse Graphs, CHRISTIAN BORGS, JENNIFER CHAYES, Microsoft Research, and ADAM SMITH, Pennsylvania State University	1369
SubmodBoxes: Near-Optimal Search for a Set of Diverse Object Proposals, QING SUN, DHRUV BATRA, Virginia Tech	1378
Fast Second Order Stochastic Backpropagation for Variational Inference, KAI FAN, Duke University, ZITENG WANG, and JEFF BECK, Duke University, JAMES KWOK, Hong Kong University of Science and Technology, KATHERINE HELLER, Duke	1387
Randomized Block Krylov Methods for Stronger and Faster Approximate Singular Value Decomposition, CAMERON MUSCO, Massachusetts Institute of Technology, CHRISTOPHER MUSCO, Mass. Institute of Technology	1396
Cross-Domain Matching for Bag-of-Words Data via Kernel Embeddings of Latent Distributions, YUYA YOSHIKAWA, NAIST, TOMOHARU IWATA, Nippon Telegraph and Telephone Corporation, and HIROSHI SAWADA, NTT Service Evolution Labs., TAKESHI YAMADA, NTT Communication Science Labs.	1405
Scalable Inference for Gaussian Process Models with Black-Box Likelihoods, AMIR DEZFOULI, The University of New South Wales, EDWIN BONILLA, University of New South Wales	1414
Fast Bidirectional Probability Estimation in Markov Models, SIDDHARTHA BANERJEE, Cornell University, PETER LOFGREN, Stanford University	1423
Probabilistic Variational Bounds for Graphical Models, QIANG LIU, JOHN FISHER III, Massachusetts Institute of Technology, and ALEX IHLER, UC Irvine	1432

Linear Response Methods for Accurate Covariance Estimates from Mean Field Variational Bayes, RYAN GIORDANO, UC Berkeley, TAMARA BRODERICK, Massachusetts Institute of Technology, and MICHAEL JORDAN, UC Berkeley	1441
Combinatorial Cascading Bandits, BRANISLAV KVETON, Adobe Research, ZHENG WEN, Yahoo Labs, and AZIN ASHKAN, Technicolor Research, CSABA SZEPESVARI, University of Alberta	1450
Mixing Time Estimation in Reversible Markov Chains from a Single Sample Path, DANIEL HSU, Columbia University, ARYEH KONTOROVITCH, Ben Gurion University, and CSABA SZEPESVARI, University of Alberta	1459
Policy Gradient for Coherent Risk Measures, AVIV TAMAR, Technion, YINLAM CHOW, Stanford University, and MOHAMMAD GHAVAMZADEH, Adobe Research & INRIA, SHIE MANNOR, Technion	1468
Fast Rates for Exp-concave Empirical Risk Minimization, TOMER KOREN, Kfir LEVY, Technion	1477
Deep Generative Image Models using a Laplacian Pyramid of Adversarial Networks, EMILY DENTON, New York University, SOUMITH CHINTALA, Facebook AI Research, and ARTHUR SZLAM, Facebook, ROB FERGUS, Facebook AI Research	1486
Decoupled Deep Neural Network for Semi-supervised Semantic Segmentation, SEUNGHOON HONG, HYEONWOO NOH, and BOHYUNG HAN, POSTECH	1495
Equilibrated adaptive learning rates for non-convex optimization, YANN DAUPHIN, Facebook AI Research, HARM VRIES, and YOSHUA BENGIO, U. Montreal	1504
BACKSHIFT: Learning causal cyclic graphs from unknown shift interventions, DOMINIK ROTHENHÄUSLER, CHRISTINA HEINZE, Eidgenössische Technische Hochschule Zürich, and JONAS PETERS, MPI Tübingen, NICOLAI MEINSHAUSEN, Eidgenössische Technische Hochschule Zürich	1513
Risk-Sensitive and Robust Decision-Making: a CVaR Optimization Approach, YINLAM CHOW, Stanford University, AVIV TAMAR, and SHIE MANNOR, Technion, MARCO PAVONE, Stanford University	1522
Asynchronous stochastic convex optimization: the noise is in the noise and SGD don't care, SORATHAN CHATURAPRUEK, JOHN DUCHI, and CHRISTOPHER RE, Stanford University	1531
Lifelong Learning with Non-i.i.d. Tasks, ANASTASIA PENTINA, CHRISTOPH LAMPERT, IST Austria	1540
Optimal Linear Estimation under Unknown Nonlinear Transform, XINYANG YI, Utaustin, ZHAORAN WANG, Princeton University, and CONSTANTINE CARAMANIS, UT Austin, HAN LIU, Princeton University	1549
Learning with Group Invariant Features: A Kernel Perspective., YOUSSEF MROUEH, International Business Machines (IBM), STEPHEN VOINEA, and TOMASO POGGIO, Massachusetts Institute of Technology	1558
Regularized EM Algorithms: A Unified Framework and Statistical Guarantees, XINYANG YI, Utaustin, CONSTANTINE CARAMANIS, UT Austin	1567

Distributionally Robust Logistic Regression, SOROOSH ABADEH, PEYMAN ESFAHANI, and DANIEL KUHN, École polytechnique fédérale de Lausanne	1576
Adaptive Stochastic Optimization: From Sets to Paths, ZHAN LIM, DAVID HSU, and WEE SUN LEE, National University of Singapore	1585
Beyond Convexity: Stochastic Quasi-Convex Optimization, ELAD HAZAN, Princeton University, KFIR LEVY, Technion, and SHAI SHALEV-SHWARTZ, Hebrew University	1594
A Tractable Approximation to Optimal Point Process Filtering: Application to Neural Encoding, YUVAL HAREL, RON MEIR, Technion, and MANFRED OPPER, TU Berlin	1603
Sum-of-Squares Lower Bounds for Sparse PCA, TENG-YU MA, Princeton University, AVI WIGDERSON, Institute for Advanced Study	1612
Max-Margin Majority Voting for Learning from Crowds, TIAN TIAN, JUNE ZHU, Tsinghua University	1621
Learning with Incremental Iterative Regularization, LORENZO ROSASCO, University of Genova, SILVIA VILLA, IIT-MIT	1630
Halting in Random Walk Kernels, MAHITO SUGIYAMA, Osaka University, KARSTEN BORGFWARDT, Eidgenössische Technische Hochschule Zürich	1639
MCMC for Variationally Sparse Gaussian Processes, JAMES HENSMAN, The University of Sheffield, ALEXANDER MATTHEWS, University of Cambridge, and MAURIZIO FILIPPONE, University of Glasgow, ZOUBIN GHAHRAMANI, University of Cambridge	1648
Less is More: Nyström Computational Regularization, ALESSANDRO RUDI, Massachusetts Institute of Technology, RAFFAELLO CAMORIANO, IIT - UNIGE, and LORENZO ROSASCO, University of Genova ...	1657
Infinite Factorial Dynamical Model, ISABEL VALERA, MPI-SWS, FRANCISCO RUIZ, University Carlos III Madrid, and LENNART SVENSSON, Chalmers University of Technology Göteborg, FERNANDO PEREZ-CRUZ	1666
Regularization Path of Cross-Validation Error Lower Bounds, ATSUSHI SHIBAGAKI, YOSHIKI SUZUKI, and MASAYUKI KARASUYAMA, ICHIRO TAKEUCHI, Nagoya Institute of Technology	1675
Attractor Network Dynamics Enable Preplay and Rapid Path Planning in Maze-like Environments, DANE CORNEIL, WULFRAM GERSTNER, École polytechnique fédérale de Lausanne	1684
Teaching Machines to Read and Comprehend, KARL HERMANN, Google DeepMind, TOMAS KOCISKY, Oxford University, and EDWARD GREFENSTETTE, LASSE ESPEHOLT, WILL KAY, MUSTAFA SULEYMAN, PHIL BLUNSOM, Google DeepMind	1693
Principal Differences Analysis: Interpretable Characterization of Differences between Distributions, JONAS MUELLER, TOMMI JAAKKOLA, Massachusetts Institute of Technology	1702
When are Kalman-Filter Restless Bandits Indexable?, CHRISTOPHER DANCE, TOMI SILANDER, Xerox Research Centre Europe	1711

Segregated Graphs and Marginals of Chain Graph Models , ILYA SHPITSER, Johns Hopkins University	1720
Efficient Non-greedy Optimization of Decision Trees , MOHAMMAD NOROUZI, University of Toronto, MAXWELL COLLINS, UW-Madison, and MATHEW JOHNSON, Microsoft Research, DAVID FLEET, University of Toronto, PUSHMEET KOHLI, Microsoft Research	1729
Probabilistic Curve Learning: Coulomb Repulsion and the Electrostatic Gaussian Process , YE WANG, Duke University, DAVID DUNSON, Duke University	1738
Inverse Reinforcement Learning with Locally Consistent Reward Functions , QUOC NGUYEN, BRYAN LOW, National University of Singapore, and PATRICK JAILLET, Massachusetts Institute of Technology	1747
Communication Complexity of Distributed Convex Learning and Optimization , YOSSI ARJEVANI, The Weizmann Institute, OHAD SHAMIR, Weizmann Institute of Science	1756
End-to-end Learning of LDA by Mirror-Descent Back Propagation over a Deep Architecture , JIANSHU CHEN, Microsoft Research Redmond WA, JI HE, University Washington, and YELONG SHEN, Microsoft Research Redmond WA, LIN XIAO, Microsoft, XIAODONG HE, JIANFENG GAO, XINYING SONG, Microsoft Research Redmond WA, DENG LI, Microsoft Research	1765
Subset Selection by Pareto Optimization , CHAO QIAN, YANG YU, and ZHI-HUA ZHOU, Nanjing University	1774
On the Accuracy of Self-Normalized Log-Linear Models , JACOB ANDREAS, MAXIM RABINOVICH, and MICHAEL JORDAN, DAN KLEIN, UC Berkeley	1783
Regret Lower Bound and Optimal Algorithm in Finite Stochastic Partial Monitoring , JUNPEI KOMIYAMA, JUNYA HONDA, and HIROSHI NAKAGAWA, The University of Tokyo	1792
Is Approval Voting Optimal Given Approval Votes? , ARIEL PROCACCIA, NISARG SHAH, Carnegie Mellon University	1801
Regressive Virtual Metric Learning , MICHAËL PERROT, AMAURY HABRARD, University of Saint-Etienne	1810
Analysis of Robust PCA via Local Incoherence , HUI SHUAI ZHANG, YI ZHOU, Syracuse University, and YINGBIN LIANG, Syracuse University	1819
Learning to Transduce with Unbounded Memory , EDWARD GREFFENSTETTE, KARL HERMANN, and MUSTAFA SULEYMAN, PHIL BLUNSOM, Google DeepMind	1828
Max-Margin Deep Generative Models , CHONGXUAN LI, JUNE ZHU, and TIANLIN SHI, BO ZHANG, Tsinghua University	1837
Spherical Random Features for Polynomial Kernels , JEFFREY PENNINGTON, Google, FELIX YU, Google Research, and SANJIV KUMAR, Google	1846
Rectified Factor Networks , DJORK-ARNÉ CLEVERT, Johannes Kepler University, ANDREAS MAYR, and THOMAS UNTERTHINER, SEPP HOCHREITER, Johannes Kepler University Linz	1855

Learning Bayesian Networks with Thousands of Variables, MAURO SCANAGATTA, IDSIA, CASSIO CAMPOS, Queen’s University Belfast, and GIORGIO CORANI, MARCO ZAFFALON, IDSIA	1864
Matrix Completion Under Monotonic Single Index Models, RAVI GANTI, UW Madison, LAURA BALZANO, University of Michigan-Ann Arbor, and REBECCA WILLETT, University of Wisconsin	1873
Visalogy: Answering Visual Analogy Questions, FERESHTEH SADEGHI, University of Washington, LARRY ZITNICK, Microsoft Research, and ALI FARHADI, University of Washington	1882
Tree-Guided MCMC Inference for Normalized Random Measure Mixture Models, JUHO LEE, SEUNGJIN CHOI, POSTECH	1891
Streaming Min-max Hypergraph Partitioning, DAN ALISTARH, Microsoft Research, JENNIFER IGLESIAS, Carnegie Mellon University, and MILAN VOJNOVIC, Microsoft Research	1900
Collaboratively Learning Preferences from Ordinal Data, SEWOONG OH, KIRAN THEKUMPARAMPIL, University of Illinois at Urbana-Champaign, and JIAMING XU	1909
Biologically Inspired Dynamic Textures for Probing Motion Perception, JONATHAN VACHER, Université Paris Dauphine, ANDREW MESO, and LAURANT PERRINET, Institut des neurosciences de la Timone, GABRIEL PEYRÉ, CNRS and Ceremade Université Paris-Dauphine	1918
Generative Image Modeling Using Spatial LSTMs, LUCAS THEIS, U.Tuebingen, MATTHIAS BETHGE, CIN University Tübingen	1927
Robust PCA with compressed data, WOOSEOK HA, The University of Chicago, RINA BARBER, University of Chicago	1936
Sampling from Probabilistic Submodular Models, ALKIS GOTOVOS, HAMED HASSANI, and ANDREAS KRAUSE, Eidgenössische Technische Hochschule Zürich	1945
COEVOLVE: A Joint Point Process Model for Information Diffusion and Network Co-evolution, MEHRDAD FARAJTABAR, Georgia Tech, YICHEN WANG, Georgia Institute of Technology, and MANUEL RODRIGUEZ, MPI SWS, SHUANG LI, Georgia Institute of Technology, HONGYUAN ZHA, Georgia Tech, LE SONG, Georgia Institute of Technology ...	1954
Supervised Learning for Dynamical System Learning, AHMED HEFNY, Carnegie Mellon University, CARLTON DOWNEY, Carnegie Mellon University, and GEOFFREY GORDON, Carnegie Mellon University	1963
Regret-Based Pruning in Extensive-Form Games, NOAM BROWN, TUOMAS SANDHOLM, Carnegie Mellon University	1972
Fast Two-Sample Testing with Analytic Representations of Probability Measures, KACPER CHWIALKOWSKI, University College London, AADITYA RAMDAS, Carnegie Mellon University, and DINO SEJDINOVIC, University of Oxford, ARTHUR GRETTON, University Collage London	1981
Learning to Segment Object Candidates, PEDRO PINHEIRO, École polytechnique fédérale de Lausanne, RONAN COLLOBERT, Facebook, and PIOTR DOLLAR, Facebook AI Research	1990

GP Kernels for Cross-Spectrum Analysis , KYLE ULRICH, Duke, DAVID CARLSON, and KAFUI DZIRASA, LAWRENCE CARIN, Duke University	1999
Secure Multi-party Differential Privacy , PETER KAIROUZ, SEWOONG OH, and PRAMOD VISWANATH, University of Illinois at Urbana-Champaign	2008
Spatial Transformer Networks , MAX JADERBERG, KAREN SIMONYAN, and ANDREW ZISSERMAN, KORAY KAVUKCUOGLU, Google DeepMind	2017
Anytime Influence Bounds and the Explosive Behavior of Continuous-Time Diffusion Networks , KEVIN SCAMAN, RÉMI LEMONNIER, and NICOLAS VAYATIS, ENS Cachan - CMLA	2026
Multi-class SVMs: From Tighter Data-Dependent Generalization Bounds to Novel Algorithms , YUNWEN LEI, City University of Hong Kong, URUN DOGAN, Microsoft, and ALEXANDER BINDER, Technical University of Berlin and Singapore University of Technology and Design, MARIUS KLOFT, Humboldt University Berlin	2035
High-dimensional neural spike train analysis with generalized count linear dynamical systems , YUANJUN GAO, LARS BUSING, Columbia University, and KRISHNA SHENOY, Stanford University, JOHN CUNNINGHAM, University of Columbia	2044
Learning with a Wasserstein Loss , CHARLIE FROGNER, CHIYUAN ZHANG, and HOSSEIN MOBAHI, Massachusetts Institute of Technology, MAURICIO ARAYA, Shell Intl. E&P Inc., TOMASO POGGIO, Massachusetts Institute of Technology	2053
b-bit Marginal Regression , MARTIN SLAWSKI, Rutgers University, PING LI, Rutgers University	2062
Natural Neural Networks , GUILLAUME DESJARDINS, KAREN SIMONYAN, and RAZVAN PASCANU, KORAY KAVUKCUOGLU, Google DeepMind	2071
Optimization Monte Carlo: Efficient and Embarrassingly Parallel Likelihood-Free Inference , TED MEEDS, U. Amsterdam, MAX WELLING, University of Amsterdam	2080
Adaptive Primal-Dual Splitting Methods for Statistical Learning and Image Processing , TOM GOLDSTEIN, University of Maryland, MIN LI, Southeast University, and XIAOMING YUAN, Hong Kong Baptist University	2089
On some provably correct cases of variational inference for topic models , PRANJAL AWASTHI, ANDREJ RISTESKI, Princeton	2098
Collaborative Filtering with Graph Information: Consistency and Scalable Methods , NIKHIL RAO, University of Texas at Austin, HSIANG-FU YU, University of Texas, and PRADEEP RAVIKUMAR, INDERJIT DHILLON, University of Texas at Austin	2107
Combinatorial Bandits Revisited , RICHARD COMBES, Supelec, M. TALEBI, KTH Royal Inst. of Technology, and ALEXANDRE PROUTIERE, MARC LELARGE, INRIA - ENS	2116
Variational Information Maximisation for Intrinsically Motivated Reinforcement Learning , SHAKIR MOHAMED, DANILO REZENDE, Google DeepMind	2125

A Structural Smoothing Framework For Robust Graph Comparison, PINAR YANARDAG, Purdue University, S.V.N. VISHWANATHAN, University of California Santa Cruz	2134
Competitive Distribution Estimation: Why is Good-Turing Good, ALON ORLITSKY, ANANDA SURESH, University of California San Diego	2143
Efficient Learning by Directed Acyclic Graph For Resource Constrained Prediction, JOSEPH WANG, Boston University, KIRILL TRAPEZNIKOV, STR, and VENKATESH SALIGRAMA, Boston University	2152
A hybrid sampler for Poisson-Kingman mixture models, MARIA LOMELI, Gatsby Unit UCL, STEFANO FAVARO, University of Torino and Collegio Carlo Alberto, and YEE WHYIE TEH, University of Oxford	2161
An Active Learning Framework using Sparse-Graph Codes for Sparse Polynomials and Graph Sketching, XIAO LI, KANNAN RAMCHANDRAN, UC Berkeley	2170
Local Smoothness in Variance Reduced Optimization, DANIEL VAINSENER, HAN LIU, Princeton University, and TONG ZHANG, Rutgers ..	2179
Saliency, Scale and Information: Towards a Unifying Theory, SHAFIN RAHMAN, NEIL BRUCE, University of Manitoba	2188
Fighting Bandits with a New Kind of Smoothness, JACOB ABERNETHY, University of Michigan, CHANSOO LEE, University of Michigan Ann Arb, and AMBUJ TEWARI, University of Michigan	2197
Beyond Sub-Gaussian Measurements: High-Dimensional Structured Estimation with Sub-Exponential Designs, VIDYASHANKAR SIVAKUMAR, University of Minnesota TC, ARINDAM BANERJEE, University of Minnesota, and PRADEEP RAVIKUMAR, University of Texas at Austin	2206
Spectral Norm Regularization of Orthonormal Representations for Graph Transduction, RAKESH SHIVANNA, Google Inc., BIBASWAN CHATTERJEE, and RAMAN SANKARAN, CHIRANJIB BHATTACHARYYA, Indian Institute of Science, FRANCIS BACH, INRIA - ENS	2215
Convolutional Networks on Graphs for Learning Molecular Fingerprints, DAVID DUVENAUD, DOUGAL MACLAURIN, and JORGE IPARRAGUIRRE, RAFAEL BOMBARELL, TIMOTHY HIRZEL, ALAN ASPURU-GUZYK, Harvard University, RYAN ADAMS, Harvard	2224
Mixed Robust/Average Submodular Partitioning: Fast Algorithms, Guarantees, and Applications, KAI WEI, University of Washington, RISHABH IYER, University of Washington Seattle, and SHENGJIE WANG, WENRUO BAI, University of Washington, JEFF BILMES, University of Washington Seattle	2233
Tractable Learning for Complex Probability Queries, JESSA BEKKER, JESSE DAVIS, KU Leuven, and ARTHUR CHOI, ADNAN DARWICHE, GUY VAN DEN BROECK, University of California Los Angeles	2242
StopWasting My Gradients: Practical SVRG, REZA HARIKANDEH, UBC, MOHAMED AHMED, and ALIM VIRANI, MARK SCHMIDT, University of British Columbia, JAKUB KONEČNÝ, SCOTT SALLINEN, UBC	2251
Mind the Gap: A Generative Approach to Interpretable Feature Selection and Extraction, BEEN KIM, JULIE SHAH, Massachusetts Institute of Technology, and FINALE DOSHI-VELEZ, Harvard	2260

A Normative Theory of Adaptive Dimensionality Reduction in Neural Networks, CENGIZ PEHLEVAN, DMITRI CHKLOVSKII, Simons Foundation	2269
On the Convergence of Stochastic Gradient MCMC Algorithms with High-Order Integrators, CHANGYOU CHEN, Duke University, NAN DING, Google, and LAWRENCE CARIN, Duke University	2278
Learning structured densities via infinite dimensional exponential families, SIQI SUN, TTIC, MLADEN KOLAR, University of Chicago Booth School of Business, and JINBO XU, Toyota Technological Institute at Chicago ..	2287
Are You Talking to a Machine? Dataset and Methods for Multilingual Image Question, HAOYUAN GAO, Baidu, JUNHUA MAO, University of California Los Angeles, and JIE ZHOU, ZHIHENG HUANG, LEI WANG, WEI XU, Baidu	2296
Variance Reduced Stochastic Gradient Descent with Neighbors, THOMAS HOFMANN, AURELIEN LUCCHI, Eidgenössische Technische Hochschule Zürich, and SIMON LACOSTE-JULIEN, INRIA, BRIAN MCWILLIAMS, Eidgenössische Technische Hochschule Zürich	2305
Sample Efficient Path Integral Control under Uncertainty, YUNPENG PAN, Georgia Institute of Technology, EVANGELOS THEODOROU, and MICHAEL KONTITSIS, Georgia Tech	2314
Stochastic Expectation Propagation, YINGZHEN LI, University of Cambridge, JOSE HERNANDEZ-LOBATO, Harvard, and RICHARD TURNER, University of Cambridge	2323
Exactness of Approximate MAP Inference in Continuous MRFs, NICHOLAS RUOZZI, UTDallas	2332
Scale Up Nonlinear Component Analysis with Doubly Stochastic Gradients, BO XIE, Georgia Tech, YINGYU LIANG, Princeton University, and LE SONG, Georgia Institute of Technology	2341
Generalization in Adaptive Data Analysis and Holdout Reuse, CYNTHIA DWORK, Microsoft Research, VITALY FELDMAN, IBM Research - Almaden, and MORITZ HARDT, Google, TONI PITASSI, University of Toronto, OMER REINGOLD, Samsung Research, AARON ROTH, University of Pennsylvania	2350
Market Scoring Rules Act As Opinion Pools For Risk-Averse Agents, MITHUN CHAKRABORTY, Washington Univ. in St. Louis, SANMAY DAS, Washington University in St. Louis	2359
Sparse Linear Programming via Primal and Dual Augmented Coordinate Descent, IAN YEN, University of Texas at Austin, KAI ZHONG, UT Austin, and CHO-JUI HSIEH, UC Davis, PRADEEP RAVIKUMAR, INDERJIT DHILLON, University of Texas at Austin	2368
Training Very Deep Networks, RUPESH SRIVASTAVA, KLAUS GREFF, IDSIA, and JÜRGEN SCHMIDHUBER	2377
Bayesian Active Model Selection with an Application to Automated Audiometry, JACOB GARDNER, Cornell University, GUSTAVO MALKOMES, Washington University in St. Louis, and ROMAN GARNETT, Washington University in STL, KILIAN WEINBERGER, Cornell University, DENNIS BARBOUR, Washington University in St. Louis, JOHN CUNNINGHAM, University of Columbia	2386

Particle Gibbs for Infinite Hidden Markov Models, NILESH TRIPURANENI, Cambridge University, SHIXIANG GU, University of Cambridge and Max Planck Institute for Intelligent Systems, and HONG GE, ZOUBIN GHAHRAMANI, University of Cambridge	2395
Learning spatiotemporal trajectories from manifold-valued longitudinal data, JEAN-BAPTISTE SCHIRATTI, STÉPHANIE ALLASSONNIERE, Ecole Polytechnique, and OLIVIER COLLIOT, Université Pierre et Marie Curie, STANLEY DURRLEMAN, INRIA	2404
A Bayesian Framework for Modeling Confidence in Perceptual Decision Making, KOOSHA KHALVATI, RAJESH RAO, University of Washington	2413
Path-SGD: Path-Normalized Optimization in Deep Neural Networks, BEHNAM NEYSHABUR, TTI Chicago, RUSLAN SALAKHUTDINOV, University of Toronto, and NATI SREBRO, Toyota Technological Institute at Chicago	2422
On the consistency theory of high dimensional variable screening, XIANGYU WANG, Duke University, CHENLEI LENG, and DAVID DUNSON, Duke University	2431
End-To-End Memory Networks, SAINBAYAR SUKHBAATAR, New York University, ARTHUR SZLAM, Facebook, and JASON WESTON, ROB FERGUS, Facebook AI Research	2440
Spectral Representations for Convolutional Neural Networks, OREN RIPPEL, Massachusetts Institute of Technology, JASPER SNOEK, and RYAN ADAMS, Harvard	2449
Online Gradient Boosting, ALINA BEYGELZIMER, Yahoo Labs, ELAD HAZAN, Princeton University, and SATYEN KALE, Yahoo Labs, HAIPENG LUO, Princeton University	2458
Deep Temporal Sigmoid Belief Networks for Sequence Modeling, ZHE GAN, CHUNYUAN LI, and RICARDO HENAO, DAVID CARLSON, LAWRENCE CARIN, Duke University	2467
Recognizing retinal ganglion cells in the dark, EMILE RICHARD, GEORGES GOETZ, and EJ CHICHILNISKY, Stanford University	2476
A Theory of Decision Making Under Dynamic Context, MICHAEL SHVARTSMAN, Princeton Neuroscience Inst., VAIBHAV SRIVASTAVA, Princeton Neuroscience Institute, and JONATHAN COHEN, Princeton University	2485
A Gaussian Process Model of Quasar Spectral Energy Distributions, ANDREW MILLER, ALBERT WU, Harvard, and JEFFREY REGIER, JON MCAULIFFE, Berkeley, DUSTIN LANG, Carnegie Mellon University, MR. PRABHAT, LBL/NERSC, DAVID SCHLEGEL, RYAN ADAMS, Harvard	2494
Hidden Technical Debt in Machine Learning Systems, D SCULLEY, Google Research, GARY HOLT, and DANIEL GOLOVIN, EUGENE DAVYDOV, TODD PHILLIPS, Google Inc., DIETMAR EBNER, VINAY CHAUDHARY, MICHAEL YOUNG, JEAN-FRANCOIS CRESPO, DAN DENNISON, Google Inc.	2503
Local Causal Discovery of Direct Causes and Effects, TIAN GAO, QIANG JI, Rensselaer Polytechnic Institute	2512

High Dimensional EM Algorithm: Statistical Optimization and Asymptotic Normality , ZHAORAN WANG, Princeton University, QUANQUAN GU, University of Virginia, and YANG NING, HAN LIU, Princeton University	2521
Revenue Optimization against Strategic Buyers , MEHRYAR MOHRI, Courant Institute and Google, ANDRES MEDINA, Courant Institute of Mathematical Sciences	2530
Deep Convolutional Inverse Graphics Network , TEJAS KULKARNI, WILL WHITNEY, Massachusetts Institute of Technology, and PUSHMEET KOHLI, Microsoft Research, JOSH TENENBAUM, Massachusetts Institute of Technology	2539
Sparse and Low-Rank Tensor Decomposition , PARIKSHIT SHAH, Yahoo Labs, NIKHIL RAO, University of Texas at Austin, and GONGGUO TANG, Colorado School of Mines	2548
Minimax Time Series Prediction , WOUTER KOOLEN, Queensland University of Technology, ALAN MALEK, and PETER BARTLETT, UC Berkeley, YASIN ABBASI, Queensland University of Technology	2557
Differentially Private Learning of Structured Discrete Distributions , ILIAS DIAKONIKOLAS, University of Edinburgh, MORITZ HARDT, Google, and LUDWIG SCHMIDT, Massachusetts Institute of Technology	2566
Variational Dropout and the Local Reparameterization Trick , DIEDERIK KINGMA, U. Amsterdam, TIM SALIMANS, Algoritmica, and MAX WELLING, University of Amsterdam	2575
Sample Complexity of Learning Mahalanobis Distance Metrics , NAKUL VERMA, KRISTIN BRANSON, Janelia Research Campus HHMI	2584
Learning Wake-Sleep Recurrent Attention Models , JIMMY BA, RUSLAN SALAKHUTDINOV, and ROGER GROSSE, University of Toronto, BRENDAN FREY, U. Toronto	2593
Robust Gaussian Graphical Modeling with the Trimmed Graphical Lasso , EUNHO YANG, IBM Thomas J. Watson Research Center, AURELIE LOZANO, IBM Research	2602
Testing Closeness With Unequal Sized Samples , BHASWAR BHATTACHARYA, GREG VALIANT, Stanford University	2611
Estimating Jaccard Index with Missing Observations: A Matrix Calibration Approach , WENYE LI, Macao Polytechnic Institute	2620
Neural Adaptive Sequential Monte Carlo , SHIXIANG GU, University of Cambridge and Max Planck Institute for Intelligent Systems, ZOUBIN GHAHRAMANI, and RICHARD TURNER, University of Cambridge	2629
Local Expectation Gradients for Black Box Variational Inference , MICHALIS TITSIAS, Athens University of Economics and Business, MIGUEL GREDILLA, Vicarious	2638
On Variance Reduction in Stochastic Gradient Descent and its Asynchronous Variants , SASHANK REDDI, AHMED HEFNY, Carnegie Mellon University, and SUVRIT SRA, Massachusetts Institute of Technology, BARNABAS POZOS, ALEX SMOLA, Carnegie Mellon University	2647
NEXT: A System for Real-World Development, Evaluation, and Application of Active Learning , KEVIN JAMIESON, LALIT JAIN, and	

CHRIS FERNANDEZ, NICHOLAS GLATTARD, University of Wisconsin, ROB NOWAK, Wisconsin	2656
Super-Resolution Off the Grid , QINGQING HUANG, Massachusetts Institute of Technology, SHAM KAKADE, University of Washington	2665
Taming the Wild: A Unified Analysis of Hogwild-Style Algorithms , CHRISTOPHER SA, Stanford University, CE ZHANG, Wisconsin, and KUNLE OLUKOTUN, CHRISTOPHER RE, Stanford University	2674
The Return of the Gating Network: Combining Generative Models and Discriminative Training in Natural Image Priors , DAN ROSENBAUM, The Hebrew University, YAIR WEISS, Hebrew University ..	2683
Pointer Networks , ORIOL VINYALS, Google, MEIRE FORTUNATO, UC Berkeley, and NAVDEEP JAITLY, Google	2692
Associative Memory via a Sparse Recovery Model , ARYA MAZUMDAR, University of Minnesota – Twin Cities, ANKIT RAWAT, Carnegie Mellon University	2701
Robust Spectral Inference for Joint Stochastic Matrix Factorization , MOONTAE LEE, DAVID BINDEL, and DAVID MIMNO, Cornell University	2710
Fast, Provable Algorithms for Isotonic Regression in all L_p-norms , RASMUS KYNG, Yale University, ANUP RAO, School of Computer Science Georgia Tech, and SUSHANT SACHDEVA, Yale University	2719
Adversarial Prediction Games for Multivariate Losses , HONG WANG, University of Illinois at Chic, WEI XING, and KAISER ASIF, BRIAN ZIEBART, University of Illinois at Chicago	2728
Asynchronous Parallel Stochastic Gradient for Nonconvex Optimization , XIANGRU LIAN, YIJUN HUANG, and YUNCHENG LI, JI LIU, University of Rochester	2737
Embed to Control: A Locally Linear Latent Dynamics Model for Control from Raw Images , MANUEL WATTER, JOST SPRINGENBERG, and JOSCHKA BOEDECKER, University of Freiburg, MARTIN RIEDMILLER, Google DeepMind	2746
Efficient and Parsimonious Agnostic Active Learning , TZU-KUO HUANG, Microsoft, ALEKH AGARWAL, Microsoft Research, and DANIEL HSU, Columbia University, JOHN LANGFORD, Microsoft Research New York, ROBERT SCHAPIRE, Microsoft Research	2755
Softstar: Heuristic-Guided Probabilistic Inference , MATHEW MONFORT, University of Illinois at Chicago, BRENDEN LAKE, New York University, and BRIAN ZIEBART, University of Illinois at Chicago, PATRICK LUCEY, Disney Research Pittsburgh, JOSH TENENBAUM, Massachusetts Institute of Technology	2764
Grammar as a Foreign Language , ORIOL VINYALS, LUKASZ KAISER, and TERRY KOO, SLAV PETROV, ILYA SUTSKEVER, GEOFFREY HINTON, Google	2773
Regularization-Free Estimation in Trace Regression with Symmetric Positive Semidefinite Matrices , MARTIN SLAWSKI, Rutgers University, PING LI, Rutgers University, and MATTHIAS HEIN, Saarland University	2782

Winner-Take-All Autoencoders , ALIREZA MAKHZANI, University of Toronto, BRENDAN FREY, U. Toronto	2791
Deep Poisson Factor Modeling , RICARDO HENAO, ZHE GAN, and JAMES LU, LAWRENCE CARIN, Duke University	2800
Bayesian Optimization with Exponential Convergence , KENJI KAWAGUCHI, LESLIE KAEHLING, and TOMÁS LOZANO-PÉREZ, Massachusetts Institute of Technology	2809
Sample Complexity of Episodic Fixed-Horizon Reinforcement Learning , CHRISTOPH DANN, EMMA BRUNSKILL, Carnegie Mellon University	2818
Learning with Relaxed Supervision , JACOB STEINHARDT, PERCY LIANG, Stanford University	2827
Subsampled Power Iteration: a Unified Algorithm for Block Models and Planted CSP's , VITALY FELDMAN, IBM Research - Almaden, WILL PERKINS, University of Birmingham, and SANTOSH VEMPALA, Georgia Tech	2836
Accelerated Mirror Descent in Continuous and Discrete Time , WALID KRICHENE, ALEXANDRE BAYEN, and PETER BARTLETT, UC Berkeley	2845
The Human Kernel , ANDREW WILSON, CHRISTOPH DANN, Carnegie Mellon University, and CHRIS LUCAS, University of Edinburgh, ERIC XING, Carnegie Mellon University	2854
Action-Conditional Video Prediction using Deep Networks in Atari Games , JUNHYUK OH, University of Michigan, XIAOXIAO GUO, Uni, and HONGLAK LEE, U. Michigan, RICHARD LEWIS, SATINDER SINGH, University of Michigan	2863
A Pseudo-Euclidean Iteration for Optimal Recovery in Noisy ICA , JAMES VOSS, MIKHAIL BELKIN, Ohio State University, and LUIS RADEMACHER, The Ohio State University	2872
Distributed Submodular Cover: Succinctly Summarizing Massive Data , BAHARAN MIRZASOLEIMAN, Eidgenössische Technische Hochschule Zürich, AMIN KARBASI, Yale, and ASHWINKUMAR BADANIDIYURU, Google Research, ANDREAS KRAUSE, Eidgenössische Technische Hochschule Zürich	2881
Community Detection via Measure Space Embedding , MARK KOZDOBA, SHIE MANNOR, Technion	2890
Basis refinement strategies for linear value function approximation in MDPs , GHEORGHE COMANICI, McGill University Montreal, DOINA PRECUP, University of McGill, and PRAKASH PANANGADEN, McGill University Montreal	2899
Structured Estimation with Atomic Norms: General Bounds and Applications , SHENG CHEN, ARINDAM BANERJEE, University of Minnesota	2908
A Complete Recipe for Stochastic Gradient MCMC , YI-AN MA, TIANQI CHEN, University of Washington, and EMILY FOX, Washington	2917
Bandit Smooth Convex Optimization: Improving the Bias-Variance Tradeoff , OFER DEKEL, Microsoft Research, RONEN ELDAN, and TOMER KOREN, Technion	2926

Online Prediction at the Limit of Zero Temperature, MARK HERBSTER, STEPHEN PASTERIS, University College London, and SHAONA GHOSH, University of Southampton	2935
Learning Continuous Control Policies by Stochastic Value Gradients, NICOLAS HEESS, GREG WAYNE, Google DeepMind, and DAVID SILVER, DeepMind, TIMOTHY LILICRAP, TOM EREZ, YUVAL TASSA, Google DeepMind	2944
Exploring Models and Data for Image Question Answering, MENGYE REN, University of Toronto, RYAN KIROS, U. Toronto, and RICHARD ZEMEL, University of Toronto	2953
Efficient and Robust Automated Machine Learning, MATTHIAS FEURER, AARON KLEIN, and KATHARINA EGGENSBERGER, JOST SPRINGENBERG, MANUEL BLUM, University of Freiburg, FRANK HUTTER, U Freiburg	2962
Preconditioned Spectral Descent for Deep Learning, DAVID CARLSON, Duke University, EDO COLLINS, and YA-PING HSIEH, École polytechnique fédérale de Lausanne, LAWRENCE CARIN, Duke University, VOLKAN CEVHER, École polytechnique fédérale de Lausanne	2971
A Recurrent Latent Variable Model for Sequential Data, JUNYOUNG CHUNG, University of Montreal, KYLE KASTNER, Université de Montreal, and VIET DINH, KRATARTH GOEL, University of Montreal, AARON COURVILLE, YOSHUA BENGIO, U. Montreal	2980
Fast Convergence of Regularized Learning in Games, VASILIS SYRGKANIS, ALEKH AGARWAL, Microsoft Research, and HAIPENG LUO, Princeton University, ROBERT SCHAPIRE, Microsoft Research	2989
Parallel Multi-Dimensional LSTM, With Application to Fast Biomedical Volumetric Image Segmentation, MARIJN STOLLENGA, WONMIN BYEON, IDSIA, and MARCUS LIWICKI, TU Kaiserslautern, JÜRGEN SCHMIDHUBER	2998
Reflection, Refraction, and Hamiltonian Monte Carlo, HADI AFSHAR, Australian National University, JUSTIN DOMKE, NICTA	3007
The Consistency of Common Neighbors for Link Prediction in Stochastic Blockmodels, PURNAMRITA SARKAR, DEEPAYAN CHAKRABARTI, UT Austin, and PETER BICKEL, U C Berkeley	3016
Nearly Optimal Private LASSO, KUNAL TALWAR, Google, ABHRADEEP THAKURTA, and LI ZHANG, Google	3025
Convergence Analysis of Prediction Markets via Randomized Subspace Descent, RAFAEL FRONGILLO, CU Boulder, MARK REID, Australia National University	3034
The Poisson Gamma Belief Network, MINGYUAN ZHOU, University of Texas at Austin, YULAI CONG, and BO CHEN, Xidian University	3043
Convergence rates of sub-sampled Newton methods, MURAT A. ERDOGDU, ANDREA MONTANARI, Stanford University	3052
No-Regret Learning in Bayesian Games, JASON HARTLINE, Northwestern University, VASILIS SYRGKANIS, Microsoft Research, and EVA TARDOS, Cornell University	3061

Statistical Topological Data Analysis - A Kernel Perspective, ROLAND KWITT, University of Salzburg, STEFAN HUBER, IST Austria, and MARC NIETHAMMER, WEILI LIN, UNC Chapel Hill, ULRICH BAUER, TU Munich	3070
Semi-supervised Sequence Learning, ANDREW DAI, Google Inc, QUOC LE, Google	3079
Structured Transforms for Small-Footprint Deep Learning, VIKAS SINDHWANI, TARA SAINATH, and SANJIV KUMAR, Google	3088
Rapidly Mixing Gibbs Sampling for a Class of Factor Graphs Using Hierarchy Width, CHRISTOPHER SA, Stanford University, CE ZHANG, Wisconsin, and KUNLE OLUKOTUN, CHRISTOPHER RE, Stanford University	3097
Interpolating Convex and Non-Convex Tensor Decompositions via the Subspace Norm, QINQING ZHENG, University of Chicago, RYOTA TOMIOKA, Toyota Technological Institute at Chicago	3106
Sample Complexity Bounds for Iterative Stochastic Policy Optimization, MARIN KOBILAROV, Johns Hopkins University	3114
BinaryConnect: Training Deep Neural Networks with binary weights during propagations, MATTHIEU COURBARIAUX, École Polytechnique de Montréal, YOSHUA BENGIO, U. Montreal, and JEAN-PIERRE DAVID, Polytechnique Montréal	3123
Interactive Control of Diverse Complex Characters with Neural Networks, IGOR MORDATCH, KENDALL LOWREY, University of Washington, and GALEN ANDREW, University of Washington Seattle, ZORAN POPOVIĆ, EMANUEL TODOROV, University of Washington	3132
Submodular Hamming Metrics, JENNIFER GILLENWATER, University of Washington, RISHABH IYER, University of Washington Seattle, and BETHANY LUSCH, RAHUL KIDAMBI, University of Washington, JEFF BILMES, University of Washington Seattle	3141
A Universal Primal-Dual Convex Optimization Framework, ALP YURTSEVER, LIONS EPFL Lausanne, QUOC DINH, Department of Statistics and Operations Research UNC North Carolina, and VOLKAN CEVHER, École polytechnique fédérale de Lausanne	3150
Learning From Small Samples: An Analysis of Simple Decision Heuristics, OZGUR SIMSEK, Max Planck Institute Berlin, MARCUS BUCKMANN, Max Planck Institute	3159
Explore no more: Improved high-probability regret bounds for non-stochastic bandits, GERGELY NEU, INRIA	3168
Fast and Memory Optimal Low-Rank Matrix Approximation, SEYOUNG YUN, MSR-INRIA, MARC LELARGE, INRIA - ENS, and ALEXANDRE PROUTIERE	3177
Learnability of Influence in Networks, HARIKRISHNA NARASIMHAN, DAVID PARKES, and YARON SINGER, Harvard University	3186
Learning Causal Graphs with Small Interventions, KARTHIKEYAN SHANMUGAM, MURAT KOCAOGLU, UT Austin, and ALEX DIMAKIS, Utaustin, SRIRAM VISHWANATH, UT Austin	3195

Information-theoretic lower bounds for convex optimization with erroneous oracles, YARON SINGER, Harvard University, JAN VONDRAK, IBM Research	3204
Fixed-Length Poisson MRF: Adding Dependencies to the Multinomial, DAVID INOUE, PRADEEP RAVIKUMAR, and INDERJIT DHILLON, University of Texas at Austin	3213
Large-Scale Bayesian Multi-Label Learning via Topic-Based Label Embeddings , PIYUSH RAI, CHANGWEI HU, and RICARDO HENAO, LAWRENCE CARIN, Duke University	3222
The Self-Normalized Estimator for Counterfactual Learning, ADITH SWAMINATHAN, Cornell University, THORSTEN JOACHIMS, Cornell ...	3231
Fast Lifted MAP Inference via Partitioning, SOMDEB SARKHEL, University of Texas at Dallas, PARAG SINGLA, Indian Institute of Technology, and VIBHAV GOGATE, UT Dallas	3240
Data Generation as Sequential Decision Making, PHILIP BACHMAN, McGill University, DOINA PRECUP, University of McGill	3249
On Elicitation Complexity, RAFAEL FRONGILLO, CU Boulder, IAN KASH, Microsoft	3258
Decomposition Bounds for Marginal MAP, WEI PING, UC Irvine, QIANG LIU, Massachusetts Institute of Technology, and ALEX IHLER, UC Irvine	3267
Discrete R’enyi Classifiers, MEISAM RAZAVIYAYN, FARZAN FARNIA, and DAVID TSE, Stanford University	3276
A class of network models recoverable by spectral clustering, YALI WAN, MARINA MEILA, University of Washington	3285
Skip-Thought Vectors, RYAN KIROS, U. Toronto, YUKUN ZHU, and RUSLAN SALAKHUTDINOV, RICHARD ZEMEL, RAQUEL URTASUN, University of Toronto, ANTONIO TORRALBA, Massachusetts Institute of Technology, SANJA FIDLER, University of Toronto	3294
Rate-Agnostic (Causal) Structure Learning, SERGEY PLIS, The Mind Research Network, DAVID DANKS, Carnegie Mellon University, and CYNTHIA FREEMAN, The Mind Research Network, VINCE CALHOUN, MRN	3303
Principal Geodesic Analysis for Probability Measures under the Optimal Transport Metric, VIVIEN SEGUY, MARCO CUTURI, Kyoto University	3312
Consistent Multilabel Classification, OLUWASANMI KOYEJO, Stanford University, NAGARAJAN NATARAJAN, UT Austin, and PRADEEP RAVIKUMAR, INDERJIT DHILLON, University of Texas at Austin	3321
Parallel Predictive Entropy Search for Batch Global Optimization of Expensive Objective Functions, AMAR SHAH, ZOUBIN GHAHRAMANI, University of Cambridge	3330
Cornering Stationary and Restless Mixing Bandits with Remix-UCB, JULIEN AUDIFFREN, CMLA ENS Cachan, LIVA RALAIVOLA, University of Marseille	3339
Semi-Supervised Factored Logistic Regression for High-Dimensional Neuroimaging Data, DANILO BZDOK, INRIA, MICHAEL	

EICKENBERG, and OLIVIER GRISEL, BERTRAND THIRION, INRIA, GAEL VAROQUAUX, Parietal Team INRIA	3348
Gaussian Process Random Fields , DAVID MOORE, STUART RUSSELL, UC Berkeley	3357
M-Statistic for Kernel Change-Point Detection , SHUANG LI, Georgia Institute of Technology, YAO XIE, and HANJUN DAI, Georgia Tech, LE SONG, Georgia Institute of Technology	3366
Adaptive Online Learning , DYLAN FOSTER, Cornell University, ALEXANDER RAKHLIN, UPenn, and KARTHIK SRIDHARAN, Cornell	3375
A Universal Catalyst for First-Order Optimization , HONGZHOU LIN, Inria, JULIEN MAIRAL, INRIA, and ZAID HARCHAOUI, Inria	3384
Inference for determinantal point processes without spectral knowledge , REMI BARDENET, University of Lille, MICHALIS TITSIAS, Athens University of Economics and Business	3393
Kullback-Leibler Proximal Variational Inference , MOHAMMAD KHAN, École polytechnique fédérale de Lausanne, PIERRE BAQUÉ, and FRANÇOIS FLEURET, Idiap Research Institute, PASCAL FUA	3402
Semi-Proximal Mirror-Prox for Nonsmooth Composite Minimization , NIAO HE, Georgia Institute of Technology, ZAID HARCHAOUI, Inria	3411
LASSO with Non-linear Measurements is Equivalent to One With Linear Measurements , CHRISTOS THRAMOULIDIS, EHSAN ABBASI, and BABAK HASSIBI, Caltech	3420
From random walks to distances on unweighted graphs , TATSUNORI HASHIMOTO, MIT CSAIL, YI SUN, MIT Mathematics, and TOMMI JAAKKOLA, Massachusetts Institute of Technology	3429
Bayesian dark knowledge , ANOOP KORATTIKARA, VIVEK RATHOD, and KEVIN MURPHY, Google, MAX WELLING	3438
Matrix Completion with Noisy Side Information , KAI-YANG CHIANG, UT Austin, CHO-JUI HSIEH, UC Davis, and INDERJIT DHILLON, University of Texas at Austin	3447
Dependent Multinomial Models Made Easy: Stick-Breaking with the Polya-gamma Augmentation , SCOTT LINDERMAN, Harvard University, MATTHEW JOHNSON, Massachusetts Institute of Technology, and RYAN ADAMS, Harvard	3456
On-the-Job Learning with Bayesian Decision Theory , KEENON WERLING, ARUN CHAGANTY, and PERCY LIANG, CHRIS MANNING, Stanford University	3465
Calibrated Structured Prediction , VOLODYMYR KULESHOV, PERCY LIANG, Stanford University	3474
Learning Structured Output Representation using Deep Conditional Generative Models , KIHYUK SOHN, University of Michigan, HONGLAK LEE, U. Michigan, and XINCHEN YAN, UMich	3483
Time-Sensitive Recommendation From Recurrent User Activities , NAN DU, YICHEN WANG, Georgia Tech, and NIAO HE, LE SONG, Georgia Institute of Technology	3492

Learning Stationary Time Series using Gaussian Processes with Nonparametric Kernels, FELIPE TOBAR, Universidad de Chile, THANG BUI, and RICHARD TURNER, University of Cambridge	3501
A Market Framework for Eliciting Private Data, BO WAGGONER, Harvard, RAFAEL FRONGILLO, CU Boulder, and JACOB ABERNETHY, University of Michigan	3510
Lifted Inference Rules With Constraints, HAPPY MITTAL, IIT Delhi, ANUJ MAHAJAN, and VIBHAV GOGATE, UT Dallas, PARAG SINGLA, Indian Institute of Technology	3519
Gradient Estimation Using Stochastic Computation Graphs, JOHN SCHULMAN, UC Berkeley / Google, NICOLAS HEESS, and THEOPHANE WEBER, Google DeepMind, PIETER ABBEEL, UC Berkeley	3528
Model-Based Relative Entropy Stochastic Search, ABBAS ABDOLMALEKI, University of Aveiro, RUDOLF LIOUTIKOV, and JANUARY PETERS, TU Darmstadt, NUNO LAU, University of Aveiro, LUIS REIS, University of Minho, GERHARD NEUMANN	3537
Semi-supervised Learning with Ladder Networks, ANTTI RASMUS, The Curious AI Company, MATHIAS BERGLUND, Aalto University, and MIKKO HONKALA, Nokia Labs, HARRI VALPOLA, The Curious AI Company, TAPANI RAIKO, Aalto University The Curious AI Company	3546
Embedding Inference for Structured Multilabel Prediction, FARZANEH MIRZAZADEH, SIAMAK RAVANBAKHSI, University of Alberta, and NAN DING, Google, DALE SCHUURMANS, Alberta	3555
Copula variational inference, DUSTIN TRAN, Harvard University, DAVID BLEI, Columbia University, and EDOARDO AIROLDI, Harvard University	3564
Recursive Training of 2D-3D Convolutional Networks for Neuronal Boundary Prediction, KISUK LEE, ALEKSANDAR ZLATESKI, Massachusetts Institute of Technology, and ASHWIN VISHWANATHAN, H. SEUNG, Princeton University	3573
A Dual Augmented Block Minimization Framework for Learning with Limited Memory, IAN YEN, University of Texas at Austin, SHAN-WEI LIN, and SHOU-DE LIN, National Taiwan University	3582
Optimal Testing for Properties of Distributions, JAYADEV ACHARYA, CONSTANTINOS DASKALAKIS, and GAUTAM KAMATH, Massachusetts Institute of Technology	3591
Efficient Learning of Continuous-Time Hidden Markov Models for Disease Progression, YU-YING LIU, SHUANG LI, and FUXIN LI, Georgia Tech, LE SONG, Georgia Institute of Technology, JAMES REHG, Georgia Tech ..	3600
Expectation Particle Belief Propagation, THIBAUT LIENART, YEE WHYI TEH, and ARNAUD DOUCET, University of Oxford	3609
Latent Bayesian melding for integrating individual and population models, MINGJUN ZHONG, University of Edinburgh, NIGEL GODDARD, and CHARLES SUTTON, University of Edinburgh	3618
Author Index	