

Neural Information Processing Systems

Advances in Neural Information Processing Systems 26

27th Annual Conference on Neural Information
Processing Systems 2013

December 5-10, 2013
Lake Tahoe, Nevada, USA

Volume 1 of 4

Printed from e-media with permission by:

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

ISBN: 978-1-63266-024-4

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

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© (2013) by Neural Information Processing Systems
All rights reserved.

Printed by Curran Associates, Inc. (2014)

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

Phone: (858) 453-4100
Fax: (858) 453-8534

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

Contents

Contents	iii
Preface	xxvii
Donors	xxxix
NIPS foundation	xxxix
Committees	xxxix
Reviewers	xxxix
The Randomized Dependence Coefficient, DAVID LOPEZ-PAZ, MPI for Intelligent Systems & University of Cambridge, PHILIPP HENNIG, and BERNHARD SCHÖLKOPF, MPI Tübingen	1
Documents as multiple overlapping windows into grids of counts, ALESSANDRO PERINA, NEBOJSA JOJIC, Microsoft Research, MANUELE BICEGO, University of Verona, and ANDRZEJ TRUSKI, Microsoft Research	10
Reciprocally Coupled Local Estimators Implement Bayesian Information Integration Distributively, WENHAO ZHANG, and SI WU, Beijing Normal University	19
Latent Maximum Margin Clustering, GUANG-TONG ZHOU, TIAN LAN, ARASH VAHDAT, and GREG MORI, Simon Fraser University	28
Data-driven Distributionally Robust Polynomial Optimization, MARTIN MEVISSSEN, EMANUELE RAGNOLI, and JIA YUAN YU, IBM Research ..	37
Transfer Learning in a Transductive Setting, MARCUS ROHRBACH, SANDRA EBERT, and BERNT SCHIELE, Max Planck Institute for Informatics	46
Bayesian optimization explains human active search, ALI BORJI, and LAURENT ITTI, University of Southern California (USC)	55
Provable Subspace Clustering: When LRR meets SSC, YU-XIANG WANG, National University of Singapore, HUAN XU, NUS, and CHENLEI LENG, University of Warwick	64
Generalized Random Utility Models with Multiple Types, HOSSEIN AZARI SOUFIANI, HANSHENG DIAO, ZHENYU LAI, and DAVID PARKES, Harvard University	73
Polar Operators for Structured Sparse Estimation, XINHUA ZHANG, NICTA, YAO-LIANG YU, and DALE SCHUURMANS, University of Alberta	82
On Decomposing the Proximal Map, YAO-LIANG YU, University of Alberta	91

Point Based Value Iteration with Optimal Belief Compression for Dec-POMDPs, LIAM MACDERMED, and CHARLES ISBELL, Georgia Tech	100
PAC-Bayes-Empirical-Bernstein Inequality, ILYA TOLSTIKHIN, Russian Academy of Sciences, and YEVGENY SELDIN, Queensland University of Technology & UC Berkeley	109
Modeling Clutter Perception using Parametric Proto-object Partitioning, CHEN-PING YU, Stony Brook University, WEN-YU HUA, Penn State University, DIMITRIS SAMARAS, and GREG ZELINSKY, Stony Brook University	118
Robust Multimodal Graph Matching: Sparse Coding Meets Graph Matching, MARCELO FIORI, Universidad de la República, Uruguay, PABLO SPRECHMANN, JOSHUA VOGELSTEIN, Duke University, PABLO MUSE, Universidad de la República, Uruguay, and GUILLERMO SAPIRO, Duke University	127
Transportability from Multiple Environments with Limited Experiments, ELIAS BAREINBOIM, University of California, Los Angeles, SANGHACK LEE, VASANT HONAVAR, Penn State University, and JUDEA PEARL, University of California, Los Angeles	136
More data speeds up training time in learning halfspaces over sparse vectors, AMIT DANIELY, Hebrew University, NATI LINIAL, and SHAI SHALEV-SHWARTZ, The Hebrew University	145
Causal Inference on Time Series using Restricted Structural Equation Models, JONAS PETERS, ETH Zurich, DOMINIK JANZING, and BERNHARD SCHÖLKOPF, MPI Tübingen	154
Deep Fisher Networks for Large-Scale Image Classification, KAREN SIMONYAN, ANDREA VEDALDI, and ANDREW ZISSERMAN, University of Oxford	163
Sparse Additive Text Models with Low Rank Background, LEI SHI, Baidu	172
Variance Reduction for Stochastic Gradient Optimization, CHONG WANG, XI CHEN, ALEX SMOLA, and ERIC XING, Carnegie Mellon University	181
Training and Analysing Deep Recurrent Neural Networks, MICHIEL HERMANS, and BENJAMIN SCHRAUWEN, Ghent University	190
A simple example of Dirichlet process mixture inconsistency for the number of components, JEFFREY MILLER, and MATTHEW HARRISON, Brown University	199
Variational Policy Search via Trajectory Optimization, SERGEY LEVINE, Stanford University, and VLADLEN KOLTUN, Adobe Research	207
Scalable kernels for graphs with continuous attributes, AASA FERAGEN, NIKLAS KASENBURG, MPI Tübingen & University of Copenhagen, JENS PETERSEN, University of Copenhagen, MARLEEN DE BRUIJNE, Erasmus MC, and KARSTEN BORGWARDT, MPI Tübingen & University of Tübingen	216
Density estimation from unweighted k-nearest neighbor graphs: a roadmap, ULRIKE VON LUXBURG, and MORTEZA ALAMGIR, University of Hamburg	225

Decision Jungles: Compact and Rich Models for Classification, JAMIE SHOTTON, TOBY SHARP, PUSHMEET KOHLI, SEBASTIAN NOWOZIN, JOHN WINN, and ANTONIO CRIMINISI, Microsoft Research	234
What Are the Invariant Occlusive Components of Image Patches? A Probabilistic Generative Approach, ZHENWEN DAI, Goethe-University Frankfurt, GEORGIOS EXARCHAKIS, UC Berkeley, and JÖRG LÜCKE, TU Berlin	243
Actor-Critic Algorithms for Risk-Sensitive MDPs, PRASHANTH L.A., INRIA, and MOHAMMAD GHAVAMZADEH, INRIA & Adobe Research	252
Summary Statistics for Partitionings and Feature Allocations, ISIK FIDANER, and TAYLAN CEMGIL, Boğaziçi University	261
One-shot learning and big data with n=2, LEE DICKER, Rutgers University, and DEAN FOSTER, University of Pennsylvania	270
Variational Inference for Mahalanobis Distance Metrics in Gaussian Process Regression, MICHALIS TITSIAS, Athens University of Economics and Business, and MIGUEL LAZARO-GREDILLA, Universidad Carlos III de Madrid	279
Correlations strike back (again): the case of associative memory retrieval, CRISTINA SAVIN, University of Cambridge, PETER DAYAN, Gatsby Unit, UCL, and MATE LENGYEL, University of Cambridge	288
Fisher-Optimal Neural Population Codes for High-Dimensional Diffeomorphic Stimulus Representations, ZHUO WANG, ALAN STOCKER, and DANIEL LEE, University of Pennsylvania	297
Online Variational Approximations to non-Exponential Family Change Point Models: With Application to Radar Tracking, RYAN TURNER, STEVEN BOTTONE, and CLAY STANEK, Northrop Grumman	306
Accelerating Stochastic Gradient Descent using Predictive Variance Reduction, RIE JOHNSON, RJ Research Consulting, and TONG ZHANG, Baidu & Rutgers	315
Using multiple samples to learn mixture models, JASON LEE, Stanford University, RAN GILAD-BACHRACH, and RICH CARUANA, Microsoft Research	324
Learning Hidden Markov Models from Non-sequence Data via Tensor Decomposition, TZU-KUO HUANG, and JEFF SCHNEIDER, Carnegie Mellon University	333
On model selection consistency of M-estimators with geometrically decomposable penalties, JASON LEE, YUEKAI SUN, and JONATHAN TAYLOR, Stanford University	342
Dropout Training as Adaptive Regularization, STEFAN WAGER, SIDA WANG, and PERCY LIANG, Stanford University	351
New Subsampling Algorithms for Fast Least Squares Regression, PARAMVEER DHILLON, YICHAO LU, DEAN FOSTER, and LYLE UNGAR, University of Pennsylvania	360
Faster Ridge Regression via the Subsampled Randomized Hadamard Transform, YICHAO LU, PARAMVEER DHILLON, DEAN FOSTER, and LYLE UNGAR, University of Pennsylvania	369

Accelerated Mini-Batch Stochastic Dual Coordinate Ascent , SHAI SHALEV-SHWARTZ, The Hebrew University, and TONG ZHANG, Baidu & Rutgers	378
Improved and Generalized Upper Bounds on the Complexity of Policy Iteration , BRUNO SCHERRER, INRIA	386
Online Learning of Nonparametric Mixture Models via Sequential Variational Approximation , DAHUA LIN, TTI Chicago	395
Online Robust PCA via Stochastic Optimization , JIASHI FENG, HUAN XU, NUS, and SHUICHENG YAN, National University of Singapore	404
Least Informative Dimensions , FABIAN SINZ, Universität Tübingen, ANNA STOCKL, Lund University, Sweden, JANUARY GREWE, and JANUARY BENDA, Universität Tübingen	413
A Scalable Approach to Probabilistic Latent Space Inference of Large-Scale Networks , JUNMING YIN, QIRONG HO, and ERIC XING, Carnegie Mellon University	422
Understanding variable importances in forests of randomized trees , GILLES LOUPPE, LOUIS WEHENKEL, ANTONIO SUTERA, and PIERRE GEURTS, Université de Liège	431
Correlated random features for fast semi-supervised learning , BRIAN MCWILLIAMS, DAVID BALDUZZI, and JOACHIM BUHMANN, ETH Zurich	440
Dynamic Clustering via Asymptotics of the Dependent Dirichlet Process Mixture , TREVOR CAMPBELL, MIT, MIAO LIU, Duke University, BRIAN KULIS, Ohio State University, JONATHAN HOW, MIT, and LAWRENCE CARIN, Duke University	449
Better Approximation and Faster Algorithm Using the Proximal Average , YAO-LIANG YU, University of Alberta	458
Rapid Distance-Based Outlier Detection via Sampling , MAHITO SUGIYAMA, MPI Tübingen, and KARSTEN BORWARDT, MPI Tübingen & University of Tübingen	467
Regularized M-estimators with nonconvexity: Statistical and algorithmic theory for local optima , PO-LING LOH, and MARTIN WAINWRIGHT, UC Berkeley	476
Non-Linear Domain Adaptation with Boosting , CARLOS BECKER, CHRISTOS CHRISTOUDIAS, and PASCAL FUA, EPFL	485
Mid-level Visual Element Discovery as Discriminative Mode Seeking , CARL DOERSCH, ABHINAV GUPTA, Carnegie Mellon University, and ALEXEI EFROS, UC Berkeley	494
q-OCSVM: A q-Quantile Estimator for High-Dimensional Distributions , ASSAF GLAZER, MICHAEL LINDENBAUM, and SHAUL MARKOVITCH, Technion	503
Auditing: Active Learning with Outcome-Dependent Query Costs , SIVAN SABATO, Microsoft Research, ANAND SARWATE, and NATI SREBRO, TTI Chicago	512
A message-passing algorithm for multi-agent trajectory planning , JOSE BENTO, JOSE BENTO, NATE DERBINSKY, JAVIER ALONSO-MORA, and JONATHAN YEDIDIA, Disney Research	521

Learning Stochastic Feedforward Neural Networks, YICHUAN TANG, and RUSLAN SALAKHUTDINOV, University of Toronto	530
Inferring neural population dynamics from multiple partial recordings of the same neural circuit, SRINI TURAGA, LARS BUESING, Gatsby Unit, UCL, ADAM PACKER, HENRY DALGLEISH, NOAH PETTIT, MICHAEL HAUSSER, UCL, and JAKOB MACKE, MPI for Biological Cybernetics	539
Multi-Prediction Deep Boltzmann Machines, IAN GOODFELLOW, MEHDI MIRZA, AARON COURVILLE, and YOSHUA BENGIO, University of Montreal	548
Higher Order Priors for Joint Intrinsic Image, Objects, and Attributes Estimation, VIBHAV VINEET, Oxford Brookes University, CARSTEN ROTHER, TU Dresden, and PHILIP TORR, University of Oxford	557
Blind Calibration in Compressed Sensing using Message Passing Algorithms, CHRISTOPHE SCHULKE, ESPCI ParisTech, FRANCESCO CALTAGIRONE, IPhT, CEA Saclay, FLORENT KRZAKALA, École Normale Supérieure, and LENKA ZDEBOROVA, CEA Saclay and CNRS URA 2306	566
Learning Trajectory Preferences for Manipulators via Iterative Improvement, ASHESH JAIN, BRIAN WOJCIK, THORSTEN JOACHIMS, and ASHUTOSH SAXENA, Cornell University	575
Large Scale Distributed Sparse Precision Estimation, HUAHUA WANG, ARINDAM BANERJEE, University of Minnesota, Twin Cities, CHO-JUI HSIEH, PRADEEP RAVIKUMAR, UT Austin, and INDERJIT DHILLON, University of Texas	584
Neural representation of action sequences: how far can a simple snippet-matching model take us?, CHESTON TAN, Institute for Infocomm Research, Singapore, JEDEDIAH SINGER, Boston Children's Hospital, THOMAS SERRE, DAVID SHEINBERG, Brown University, and TOMASO POGGIO, MIT	593
On Algorithms for Sparse Multi-factor NMF, SIWEI LYU, and XIN WANG, SUNY at Albany	602
Dirty Statistical Models, EUNHO YANG, and PRADEEP RAVIKUMAR, UT Austin	611
Parallel Sampling of DP Mixture Models using Sub-Cluster Splits, JASON CHANG, and JOHN FISHER III, MIT	620
Trading Computation for Communication: Distributed Stochastic Dual Coordinate Ascent, TIANBAO YANG, NEC Labs America	629
Prior-free and prior-dependent regret bounds for Thompson Sampling, SEBASTIEN BUBECK, and CHE-YU LIU, Princeton University	638
Structured Learning via Logistic Regression, JUSTIN DOMKE, NICTA	647
Which Space Partitioning Tree to Use for Search?, PARIKSHIT RAM, and ALEXANDER GRAY, Georgia Tech	656
Projecting Ising Model Parameters for Fast Mixing, JUSTIN DOMKE, NICTA, and XIANGHANG LIU, NICTA/UNSW	665
Mixed Optimization for Smooth Functions, MEHRDAD MAHDAVI, LIJUN ZHANG, and RONG JIN, Michigan State University (MSU)	674

Conditional Random Fields via Univariate Exponential Families, EUNHO YANG, PRADEEP RAVIKUMAR, UT Austin, GENEVERA ALLEN, Rice University, and ZHANDONG LIU, Baylor College of Medicine	683
Stochastic blockmodel approximation of a graphon: Theory and consistent estimation, EDOARDO AIROLDI, THIAGO COSTA, and STANLEY CHAN, Harvard University	692
Reinforcement Learning in Robust Markov Decision Processes, SHIAU HONG LIM, National University of Singapore, HUAN XU, NUS, and SHIE MANNOR, Technion	701
On the Linear Convergence of the Proximal Gradient Method for Trace Norm Regularization, KE HOU, ZIRUI ZHOU, ANTHONY MAN-CHO SO, CUHK, and ZHI-QUAN LUO, University of Minnesota, Twin Cites	710
Recurrent networks of coupled Winner-Take-All oscillators for solving constraint satisfaction problems, HESHAM MOSTAFA, LORENZ. MUELLER, and GIACOMO INDIVERI, ETH Zurich	719
Latent Structured Active Learning, WENJIE LUO, TTI Chicago, ALEX SCHWING, ETH Zurich, and RAQUEL URTASUN, TTI Chicago	728
A Gang of Bandits, NICOLÒ CESA-BIANCHI, University of Milan, CLAUDIO GENTILE, University of Insubria, and GIOVANNI ZAPPELLA, University of Milan	737
Learning Feature Selection Dependencies in Multi-task Learning, DANIEL HERNÁNDEZ-LOBATO, Universidad Autónoma de Madrid, and JOSÉ MIGUEL HERNÁNDEZ-LOBATO, University of Cambridge	746
B-tests: Low Variance Kernel Two-Sample Tests, WOJCIECH ZAREMBA, École Centrale Paris, ARTHUR GRETTON, UCL, and MATTHEW BLASCHKO, École Centrale Paris	755
Online PCA for Contaminated Data, JIASHI FENG, HUAN XU, NUS, SHIE MANNOR, Technion, and SHUICHENG YAN, National University of Singapore	764
Non-strongly-convex smooth stochastic approximation with convergence rate $O(1/n)$, FRANCIS BACH, INRIA & ENS, and ERIC MOULINES, Telecom ParisTech	773
Efficient Algorithm for Privately Releasing Smooth Queries, ZITENG WANG, KAI FAN, JIAQI ZHANG, and LIWEI WANG, Peking University	782
Beyond Pairwise: Provably Fast Algorithms for Approximate k-Way Similarity Search, ANSHUMALI SHRIVASTAVA, and PING LI, Cornell University	791
Unsupervised Spectral Learning of Finite State Transducers, RAPHAEL BAILLY, XAVIER CARRERAS, and ARIADNA QUATTONI, Universitat Politècnica de Catalunya	800
Learning a Deep Compact Image Representation for Visual Tracking, NAIYAN WANG, and DIT-YAN YEUNG, Hong Kong University of Science and Technology	809
Learning Multi-level Sparse Representations, FERRAN DIEGO ANDILLA, and FRED HAMPRECHT, University of Heidelberg	818

Robust Data-Driven Dynamic Programming, GRANI ADIWENA HANASUSANTO, Imperial College London, and DANIEL KUHN, EPFL	827
Low-Rank Matrix and Tensor Completion via Adaptive Sampling, AKSHAY KRISHNAMURTHY, and AARTI SINGH, Carnegie Mellon University	836
Probabilistic Low-Rank Matrix Completion with Adaptive Spectral Regularization Algorithms, ADRIEN TODESCHINI, INRIA, FRANÇOIS CARON, University of Oxford, and MARIE CHAVENT, Université de Bordeaux II & INRIA	845
Distributed Exploration in Multi-Armed Bandits, ESHCAR HILLEL, ZOHAR KARNIN, Yahoo! Labs, TOMER KOREN, Technion, RONNY LEMPEL, and OREN SOMEKH, Yahoo! Labs	854
The Pareto Regret Frontier, WOUTER KOOLEN, Queensland University of Technology	863
Direct 0-1 Loss Minimization and Margin Maximization with Boosting, SHAODAN ZHAI, TIAN XIA, MING TAN, and SHAOJUN WANG, Wright State University	872
Regret based Robust Solutions for Uncertain Markov Decision Processes, ASRAR AHMED, PRADEEP VARAKANTHAM, Singapore Management University, YOSSIRI ADULYASAK, Singapore-MIT Alliance for Research and Technology, and PATRICK JAILLET, MIT	881
Speeding up Permutation Testing in Neuroimaging, CHRIS HINRICHS, VAMSI ITHAPU, QINYUAN SUN, STERLING JOHNSON, and VIKAS SINGH, UW-Madison	890
Generalized Denoising Auto-Encoders as Generative Models, YOSHUA BENGIO, LI YAO, GUILLAUME ALAIN, and PASCAL VINCENT, University of Montreal	899
Supervised Sparse Analysis and Synthesis Operators, PABLO SPRECHMANN, Duke University, ROEE LITMAN, TAL BEN YAKAR, ALEXANDER BRONSTEIN, Tel Aviv University, and GUILLERMO SAPIRO, Duke University	908
Low-rank matrix reconstruction and clustering via approximate message passing, RYOSUKE MATSUSHITA, NTT DATA Mathematical Systems Inc., and TOSHIYUKI TANAKA, Kyoto University	917
Reasoning With Neural Tensor Networks for Knowledge Base Completion, RICHARD SOCHER, Stanford University, DANQI CHEN, Saarland University, CHRISTOPHER MANNING, and ANDREW NG, Stanford University	926
Zero-Shot Learning Through Cross-Modal Transfer, RICHARD SOCHER, MILIND GANJOO, CHRISTOPHER MANNING, and ANDREW NG, Stanford University	936
Estimating LASSO Risk and Noise Level, MOHSEN BAYATI, MURAT A. ERDOGDU, and ANDREA MONTANARI, Stanford University	946
Learning Adaptive Value of Information for Structured Prediction, DAVID WEISS, University of Pennsylvania, and BEN TASKAR, University of Washington	955

Efficient Online Inference for Bayesian Nonparametric Relational Models, DAE IL KIM, Brown University, PREM GOPALAN, DAVID BLEI, Princeton University, and ERIK SUDDERTH, Brown University	964
Approximate inference in latent Gaussian-Markov models from continuous time observations, BOTOND CSEKE, University of Edinburgh, MANFRED OPPER, TU Berlin, and GUIDO SANGUINETTI, University of Edinburgh	973
Linear Convergence with Condition Number Independent Access of Full Gradients, LIJUN ZHANG, MEHRDAD MAHDAVI, and RONG JIN, Michigan State University (MSU)	982
When in Doubt, SWAP: High-Dimensional Sparse Recovery from Correlated Measurements, DIVYANSHU VATS, and RICHARD BARANIUK, Rice University	991
Wavelets on Graphs via Deep Learning, RAIF RUSTAMOV, and LEONIDAS GUIBAS, Stanford University	1000
Robust Spatial Filtering with Beta Divergence, WOJCIECH SAMEK, DUNCAN BLYTHE, KLAUS-ROBERT MÜLLER, TU Berlin, and MOTOAKI KAWANABE, ATR	1009
Convex Relaxations for Permutation Problems, FAJWEL FOGEL, École Polytechnique, RODOLPHE JENATTON, CMAP, FRANCIS BACH, INRIA & ENS, and ALEXANDRE D’ASPREMONT, CNRS – ENS	1018
High-Dimensional Gaussian Process Bandits, JOSIP DJOLONGA, ANDREAS KRAUSE, ETH Zurich, and VOLKAN CEVHER, EPFL	1027
A memory frontier for complex synapses, SUBHANEIL LAHIRI, and SURYA GANGULI, Stanford University	1036
Marginals-to-Models Reducibility, TIM ROUGHGARDEN, Stanford University, and MICHAEL KEARNS, University of Pennsylvania	1045
First-Order Decomposition Trees, NIMA TAGHIPOUR, JESSE DAVIS, and HENDRIK BLOCKEEL, KU Leuven	1054
A Comparative Framework for Preconditioned Lasso Algorithms, FABIAN WAUTHIER, UC Berkeley, NEBOJSA JOJIC, Microsoft Research, and MICHAEL JORDAN, UC Berkeley	1063
Lasso Screening Rules via Dual Polytope Projection, JIE WANG, JIAYU ZHOU, PETER WONKA, and JIEPING YE, Arizona State University	1072
Binary to Bushy: Bayesian Hierarchical Clustering with the Beta Coalescent, YUENING HU, JORDAN BOYD-GRABER, HAL DAUME III, University of Maryland, and Z. IRENE YING, US Department of Agriculture ...	1081
A Latent Source Model for Nonparametric Time Series Classification, GEORGE CHEN, MIT, STANISLAV NIKOLOV, Twitter, and DEVAVRAT SHAH, MIT	1090
Efficient Optimization for Sparse Gaussian Process Regression, YANSHUAI CAO, University of Toronto, MARCUS BRUBAKER, TTI Chicago, DAVID FLEET, University of Toronto, and AARON HERTZMANN, Adobe Research	1099
Lexical and Hierarchical Topic Regression, VIET-AN NGUYEN, JORDAN BOYD-GRABER, and PHILIP RESNIK, University of Maryland	1108

Stochastic Convex Optimization with Multiple Objectives, MEHRDAD MAHDAVI, Michigan State University (MSU), TIANBAO YANG, NEC Labs America, and RONG JIN, Michigan State University (MSU)	1117
A Kernel Test for Three-Variable Interactions, DINO SEJDINOVIC, Gatsby Unit, UCL, ARTHUR GREYTON, UCL, and WICHER BERGSMA, LSE .	1126
Memoized Online Variational Inference for Dirichlet Process Mixture Models, MICHAEL HUGHES, and ERIK SUDDERTH, Brown University	1135
Designed Measurements for Vector Count Data, LIMING WANG, DAVID CARLSON, Duke University, MIGUEL RODRIGUES, UCL, DAVID WILCOX, Purdue University, ROBERT CALDERBANK, and LAWRENCE CARIN, Duke University	1144
Robust Transfer Principal Component Analysis with Rank Constraints, YUHONG GUO, Temple University	1153
Online Learning with Switching Costs and Other Adaptive Adversaries, NICOLÒ CESA-BIANCHI, University of Milan, OFER DEKEL, Microsoft Research, and OHAD SHAMIR, The Weizmann Institute	1162
Learning Prices for Repeated Auctions with Strategic Buyers, KAREEM AMIN, University of Pennsylvania, AFSHIN ROSTAMIZADEH, and UMAR SYED, Google Research	1171
Probabilistic Principal Geodesic Analysis, MIAOMIAO ZHANG, and P.T. FLETCHER, University of Utah	1180
Confidence Intervals and Hypothesis Testing for High-Dimensional Statistical Models, ADEL JAVANMARD, and ANDREA MONTANARI, Stanford University	1189
Learning with Noisy Labels, NAGARAJAN NATARAJAN, UT Austin, INDERJIT DHILLON, University of Texas, PRADEEP RAVIKUMAR, UT Austin, and AMBUJ TEWARI, University of Michigan	1198
Tracking Time-varying Graphical Structure, ERICH KUMMERFELD, and DAVID DANKS, Carnegie Mellon University	1207
Factorized Asymptotic Bayesian Inference for Latent Feature Models, KOHEI HAYASHI, NII, and RYOHEI FUJIMAKI, NEC Labs America	1216
More Effective Distributed ML via a Stale Synchronous Parallel Parameter Server, QIRONG HO, JAMES CIPAR, HENGGANG CUI, JIN KYU KIM, SEUNGHAK LEE, Carnegie Mellon University, PHILLIP B. GIBBONS, Intel Labs, GARTH GIBSON, GREG GANGER, and ERIC XING, Carnegie Mellon University	1225
Bayesian Estimation of Latently-grouped Parameters in Undirected Graphical Models, JIE LIU, and DAVID PAGE, UW-Madison	1234
Online Learning with Costly Features and Labels, NAVID ZOLGHADR, University of Alberta, GABOR BARTOK, ETH Zurich, RUSSELL GREINER, ANDRÁS GYÖRGY, and CSABA SZEPESVARI, University of Alberta	1243
Sparse nonnegative deconvolution for compressive calcium imaging: algorithms and phase transitions, EFTYCHIOS PNEVMATIKAKIS, and LIAM PANINSKI, Columbia University	1252

A Novel Two-Step Method for Cross Language Representation Learning, MIN XIAO, and YUHONG GUO, Temple University	1261
On Sampling from the Gibbs Distribution with Random Maximum A-Posteriori Perturbations, TAMIR HAZAN, University of Haifa, SUBHRANSU MAJI, TTI Chicago, and TOMMI JAAKKOLA, MIT	1270
Graphical Models for Inference with Missing Data, KARTHIKA MOHAN, JUDEA PEARL, University of California, Los Angeles, and JIN TIAN, Iowa State University	1279
Reshaping Visual Datasets for Domain Adaptation, BOQING GONG, University of Southern California (USC), KRISTEN GRAUMAN, UT Austin, and FEI SHA, University of Southern California (USC)	1288
Statistical Active Learning Algorithms, MARIA-FLORINA BALCAN, Georgia Tech, and VITALY FELDMAN, IBM Research	1297
Bayesian Inference and Online Experimental Design for Mapping Neural Microcircuits, BEN SHABABO, Columbia University, BROOKS PAIGE, University of Oxford, ARI PAKMAN, and LIAM PANINSKI, Columbia University	1306
Reflection methods for user-friendly submodular optimization, STEFANIE JEGELKA, UC Berkeley, FRANCIS BACH, INRIA & ENS, and SUVRIT SRA, MPI for Intelligent Systems & Carnegie Mellon University	1315
Unsupervised Structure Learning of Stochastic And-Or Grammars, KEWEI TU, MARIA PAVLOVSKAIA, and SONG-CHUN ZHU, University of California, Los Angeles	1324
Convex Tensor Decomposition via Structured Schatten Norm Regularization, RYOTA TOMIOKA, TTI Chicago, and TAIJI SUZUKI, Tokyo Institute of Technology	1333
Stochastic Ratio Matching of RBMs for Sparse High-Dimensional Inputs, YANN DAUPHIN, and YOSHUA BENGIO, University of Montreal	1342
Learning Chordal Markov Networks by Constraint Satisfaction, JUKKA CORANDER, University of Helsinki, TOMI JANHUNEN, JUSSI RINTANEN, Aalto University, HENRIK NYMAN, and JOHAN PENSAR, Åbo Akademi	1351
Parametric Task Learning, ICHIRO TAKEUCHI, TATSUYA HONGO, Nagoya Institute of Technology, MASASHI SUGIYAMA, Tokyo Institute of Technology, and SHINICHI NAKAJIMA, Nikon	1360
A Deep Architecture for Matching Short Texts, ZHENG DONG LU, and HANG LI, Noah's Ark Lab, Huawei Technologies	1369
Computing the Stationary Distribution Locally, CHRISTINA LEE, ASUMAN OZDAGLAR, and DEVAVRAT SHAH, MIT	1378
Nonparametric Multi-group Membership Model for Dynamic Networks, MYUNGHWAN KIM, and JURE LESKOVEC, Stanford University .	1387
Adaptive Step-Size for Policy Gradient Methods, MATTEO PIROTTA, MARCELLO RESTELLI, and LUCA BASCETTA, Politecnico di Milano	1396
Optimistic Concurrency Control for Distributed Unsupervised Learning, XINGHAO PAN, JOSEPH GONZALEZ, STEFANIE JEGELKA, TAMARA BRODERICK, and MICHAEL JORDAN, UC Berkeley	1405

Reservoir Boosting : Between Online and Offline Ensemble Learning, LEONIDAS LEFAKIS, and FRANÇOIS FLEURET, Idiap Research Institute	1414
Multiclass Total Variation Clustering, XAVIER BRESSON, City University of Hong Kong, THOMAS LAURENT, Loyola Marymount University, DAVID UMINSKY, University of San Francisco, and JAMES VON BRECHT, University of California, Los Angeles	1423
Approximate Inference in Continuous Determinantal Processes, RAJA HAFIZ AFFANDI, University of Pennsylvania, EMILY FOX, and BEN TASKAR, University of Washington	1432
Global Solver and Its Efficient Approximation for Variational Bayesian Low-rank Subspace Clustering, SHINICHI NAKAJIMA, Nikon, AKIKO TAKEDA, University of Tokyo, S. DERIN BABACAN, Google Research, MASASHI SUGIYAMA, Tokyo Institute of Technology, and ICHIRO TAKEUCHI, Nagoya Institute of Technology	1441
Thompson Sampling for 1-Dimensional Exponential Family Bandits, NATHANIEL KORDA, INRIA, EMILIE KAUFMANN, Telecom ParisTech, and REMI MUNOS, INRIA	1450
Active Learning for Probabilistic Hypotheses Using the Maximum Gibbs Error Criterion, NGUYEN VIET CUONG, WEE SUN LEE, NAN YE, National University of Singapore, KIAN MING CHAI, and HAI LEONG CHIEU, DSO National Laboratories	1459
It is all in the noise: Efficient multi-task Gaussian process inference with structured residuals, BARBARA RAKITSCH, MPI Tübingen, CHRISTOPH LIPPERT, Microsoft Research, KARSTEN BORGWARDT, MPI Tübingen & University of Tübingen, and OLIVER STEGLE, EMBL-EBI ...	1468
Convex Calibrated Surrogates for Low-Rank Loss Matrices with Applications to Subset Ranking Losses, HARISH RAMASWAMY, SHIVANI AGARWAL, Indian Institute of Science, and AMBUJ TEWARI, University of Michigan	1477
Inverse Density as an Inverse Problem: the Fredholm Equation Approach, QICHAO QUE, and MIKHAIL BELKIN, Ohio State University	1486
Adaptive Multi-Column Deep Neural Networks with Application to Robust Image Denoising, FOREST AGOSTINELLI, MICHAEL ANDERSON, and HONGLAK LEE, University of Michigan	1495
EDML for Learning Parameters in Directed and Undirected Graphical Models, KHALED REFAAT, ARTHUR CHOI, and ADNAN DARWICHE, University of California, Los Angeles	1504
Similarity Component Analysis, SORAVIT CHANGPINYO, KUAN LIU, and FEI SHA, University of Southern California (USC)	1513
Approximate Bayesian Image Interpretation using Generative Probabilistic Graphics Programs, VIKASH MANSINGHKA, TEJAS KULKARNI, YURA PEROV, and JOSH TENENBAUM, MIT	1522
Local Privacy and Minimax Bounds: Sharp Rates for Probability Estimation, JOHN DUCHI, MICHAEL JORDAN, and MARTIN WAINWRIGHT, UC Berkeley	1531

Firing rate predictions in optimal balanced networks, DAVID BARRETT, SOPHIE DENÈVE, École Normale Supérieure, and CHRISTIAN MACHENS, Champalimaud Centre for the Unknown	1540
Manifold-based Similarity Adaptation for Label Propagation, MASAYUKI KARASUYAMA, and HIROSHI MAMITSUKA, Kyoto University	1549
Non-Uniform Camera Shake Removal Using a Spatially-Adaptive Sparse Penalty, HAICHAO ZHANG, Duke University, and DAVID WIPF, Microsoft Research	1558
Near-Optimal Entrywise Sampling for Data Matrices, DIMITRIS ACHLIOPTAS, UC Santa Cruz, ZOHAR KARNIN, Yahoo! Labs, and EDO LIBERTY, Yahoo! Research	1567
Learning to Prune in Metric and Non-Metric Spaces, LEONID BOYTISOV, Carnegie Mellon University, and BILEGSAIKHAN NAIDAN, Norwegian University of Science and Technology (NTNU)	1576
Online Learning in Episodic Markovian Decision Processes by Relative Entropy Policy Search, ALEXANDER ZIMIN, Institute of Science and Technology Austria, and GERGELY NEU, INRIA	1585
Optimistic policy iteration and natural actor-critic: A unifying view and a non-optimality result, PAUL WAGNER, Aalto University ...	1594
Bayesian Hierarchical Community Discovery, CHARLES BLUNDELL, Gatsby Unit, UCL, and YEE WHYIE TEH, University of Oxford	1603
From Bandits to Experts: A Tale of Domination and Independence, NOGA ALON, Tel Aviv University, NICOLÒ CESA-BIANCHI, University of Milan, CLAUDIO GENTILE, University of Insubria, and YISHAY MANSOUR, Tel Aviv University	1612
Predictive PAC Learning and Process Decompositions, COSMA SHALIZI, Carnegie Mellon University, and ARYEH KONTOROVITCH, Ben Gurion University	1621
Pass-efficient unsupervised feature selection, CRYSTAL MAUNG, and HAIM SCHWEITZER, UT Dallas	1630
Simultaneous Rectification and Alignment via Robust Recovery of Low-rank Tensors, XIAOQIN ZHANG, DI WANG, Wenzhou University, ZHENGYUAN ZHOU, Stanford University, and YI MA, Microsoft Research	1639
Bayesian Mixture Modeling and Inference based Thompson Sampling in Monte-Carlo Tree Search, AIJUN BAI, University of Science and Technology of China (USTC), FENG WU, University of Southampton, and XIAOPING CHEN, University of Science and Technology of China (USTC)	1648
Solving inverse problem of Markov chain with partial observations, TETSURO MORIMURA, TAKAYUKI OSOGAMI, and TSUYOSHI IDE, IBM Research	1657
Locally Adaptive Bayesian Multivariate Time Series, DANIELE DURANTE, BRUNO SCARPA, University of Padua, and DAVID DUNSON, Duke University	1666
Mapping paradigm ontologies to and from the brain, YANNICK SCHWARTZ, BERTRAND THIRION, and GAEL VAROQUAUX, INRIA	1675

Noise-Enhanced Associative Memories, AMIN KARBASI, ETH Zurich, AMIR HESAM SALAVATI, AMIN SHOKROLLAHI, EPFL, and LAV VARSHNEY, IBM Watson Research Center	1684
Exact and Stable Recovery of Pairwise Interaction Tensors, SHOUYUAN CHEN, MICHAEL LYU, CUHK, IRWIN KING, Chinese University of Hong Kong, and ZENGLIN XU, University of Purdue	1693
Bayesian entropy estimation for binary spike train data using parametric prior knowledge, EVAN ARCHER, IL PARK, and JONATHAN PILLOW, UT Austin	1702
Perfect Associative Learning with Spike-Timing-Dependent Plasticity, CHRISTIAN ALBERS, MAREN WESTKOTT, University of Bremen, and KLAUS PAWELZIK, Universität Bremen	1711
On Poisson Graphical Models, EUNHO YANG, PRADEEP RAVIKUMAR, UT Austin, GENEVERA ALLEN, Rice University, and ZHANDONG LIU, Baylor College of Medicine	1720
Streaming Variational Bayes, TAMARA BRODERICK, NICHOLAS BOYD, ANDRE WIBISONO, ASHIA WILSON, and MICHAEL JORDAN, UC Berkeley ..	1729
Gaussian Process Conditional Copulas with Applications to Financial Time Series, JOSÉ MIGUEL HERNÁNDEZ-LOBATO, JAMES LLOYD, University of Cambridge, and DANIEL HERNÁNDEZ-LOBATO, Universidad Autónoma de Madrid	1738
Extracting regions of interest from biological images with convolutional sparse block coding, MARIUS PACHITARIU, Gatsby Unit, UCL, ADAM PACKER, NOAH PETTIT, HENRY DALGLEISH, MICHAEL HAUSSER, UCL, and MANEESH SAHANI, Gatsby Unit, UCL	1747
Approximate Dynamic Programming Finally Performs Well in the Game of Tetris, VICTOR GABILLON, INRIA, MOHAMMAD GHAVAMZADEH, INRIA & Adobe Research, and BRUNO SCHERRER, INRIA ..	1756
Third-Order Edge Statistics: Contour Continuation, Curvature, and Cortical Connections, MATTHEW LAWLOR, and STEVEN ZUCKER, Yale University	1765
DESPOT: Online POMDP Planning with Regularization, ADHIRAJ SOMANI, NUS, NAN YE, DAVID HSU, and WEE SUN LEE, National University of Singapore	1774
Matrix Completion From any Given Set of Observations, TROY LEE, Centre for Quantum Technologies, and ADI SHRAIBMAN, Weizmann Institute of Science	1783
Regression-tree Tuning in a Streaming Setting, SAMORY KPOTUFE, and FRANCESCO ORABONA, TTI Chicago	1790
Multiscale Dictionary Learning for Estimating Conditional Distributions, FRANCESCA PETRALIA, Mt Sinai School of Medicine, JOSHUA VOGELSTEIN, and DAVID DUNSON, Duke University	1799
Dimension-Free Exponentiated Gradient, FRANCESCO ORABONA, TTI Chicago	1808
Stochastic Optimization of PCA with Capped MSG, RAMAN ARORA, University of Washington, ANDY COTTER, and NATI SREBRO, TTI Chicago	1817

On Flat versus Hierarchical Classification in Large-Scale Taxonomies, ROHIT BABBAR, Université Joseph Fourier, Grenoble, IOANNIS PARTALAS, UJF/LIG, ERIC GAUSSIER, and MASSIH-REZA AMINI, Université Joseph Fourier, Grenoble	1826
Learning Gaussian Graphical Models with Observed or Latent FVSs, YING LIU, and ALAN WILLSKY, MIT	1835
Visual Concept Learning: Combining Machine Vision and Bayesian Generalization on Concept Hierarchies, YANGQING JIA, JOSHUA ABBOTT, UC Berkeley, JOSEPH AUSTERWEIL, Brown University, THOMAS GRIFFITHS, and TREVOR DARRELL, UC Berkeley	1844
Robust Bloom Filters for Large MultiLabel Classification Tasks, MOUSTAPHA CISSE, LIP6/UPMC, NICOLAS USUNIER, Université de Technologie de Compiègne (UTC), THIERRY ARTIÈRES, and PATRICK GALLINARI, LIP6/UPMC	1853
Solving the multi-way matching problem by permutation synchronization, DEEPTI PACHAURI, UW-Madison, RISI KONDOR, University of Chicago, and VIKAS SINGH, UW-Madison	1862
Generalizing Analytic Shrinkage for Arbitrary Covariance Structures, DANIEL BARTZ, and KLAUS-ROBERT MÜLLER, TU Berlin	1871
Top-Down Regularization of Deep Belief Networks, HANLIN GOH, LIP6/UPMC, NICOLAS THOME, MATTHIEU CORD, University Pierre & Marie Curie and CNRS (UMR 7606), and JOO-HWEE LIM, Institute for Infocomm Research, Singapore	1880
Learning Efficient Random Maximum A-Posteriori Predictors with Non-Decomposable Loss Functions, TAMIR HAZAN, University of Haifa, SUBHRANSU MAJI, TTI Chicago, JOSEPH KESHET, Bar-Ilan University, and TOMMI JAAKKOLA, MIT	1889
Heterogeneous-Neighborhood-based Multi-Task Local Learning Algorithms, YU ZHANG, Hong Kong Baptist University	1898
Machine Teaching for Bayesian Learners in the Exponential Family, XIAOJIN ZHU, UW-Madison	1907
Scoring Workers in Crowdsourcing: How Many Control Questions are Enough?, QIANG LIU, MARK STEYVERS, and ALEX IHLER, UC Irvine	1916
Action from Still Image Dataset and Inverse Optimal Control to Learn Task Specific Visual Scanpaths, STEFAN MATHE, University of Toronto, and CRISTIAN SMINCHISESCU, LTH	1925
A Determinantal Point Process Latent Variable Model for Inhibition in Neural Spiking Data, JASPER SNOEK, RICHARD ZEMEL, University of Toronto, and RYAN ADAMS, Harvard University	1934
Robust Sparse Principal Component Regression under the High Dimensional Elliptical Model, FANG HAN, Johns Hopkins University, and HAN LIU, Princeton University	1943
Global MAP-Optimality by Shrinking the Combinatorial Search Area with Convex Relaxation, BOGDAN SAVCHYNSKY, JÖRG HENDRIK KAPPES, PAUL SWOBODA, and CHRISTOPH SCHNÖRR, University of Heidelberg	1952

Near-optimal Anomaly Detection in Graphs using Lovasz Extended Scan Statistic, JAMES SHARPNACK, AKSHAY KRISHNAMURTHY, and AARTI SINGH, Carnegie Mellon University	1961
Demixing odors - fast inference in olfaction, AGNIESZKA GRABSKA-BARWINSKA, JEFF BECK, Gatsby Unit, UCL, ALEXANDRE POUGET, University of Geneva, and PETER LATHAM, Gatsby Unit, UCL	1970
Learning Multiple Models via Regularized Weighting, DANIEL VAINSENER, SHIE MANNOR, Technion, and HUAN XU, NUS	1979
When Are Overcomplete Topic Models Identifiable? Uniqueness of Tensor Tucker Decompositions with Structured Sparsity, ANIMA ANANDKUMAR, UC Irvine, DANIEL HSU, Columbia University, MAJID JANZAMIN, UC Irvine, and SHAM KAKADE, Microsoft Research	1988
Distributed k-Means and k-Median Clustering on General Topologies, MARIA-FLORINA BALCAN, STEVEN EHRLICH, and YINGYU LIANG, Georgia Tech	1997
Multi-Task Bayesian Optimization, KEVIN SWERSKY, JASPER SNOEK, University of Toronto, and RYAN ADAMS, Harvard University	2006
Online Learning of Dynamic Parameters in Social Networks, SHAHIN SHAHRAMPOUR, SASHA RAKHLIN, and ALI JADBABAIE, University of Pennsylvania	2015
A Graphical Transformation for Belief Propagation: Maximum Weight Matchings and Odd-Sized Cycles, JINWOO SHIN, KAIST, ANDREW GELFAND, UC Irvine, and MISHA CHERTKOV, Los Alamos National Laboratory	2024
Learning with Invariance via Linear Functionals on Reproducing Kernel Hilbert Space, XINHUA ZHANG, NICTA, WEE SUN LEE, National University of Singapore, and YEE WHYIE TEH, University of Oxford	2033
Approximate Gaussian process inference for the drift function in stochastic differential equations, ANDREAS RUTTOR, PHILIPP BATZ, and MANFRED OPPER, TU Berlin	2042
Distributed Submodular Maximization: Identifying Representa- tive Elements in Massive Data, BAHARAN MIRZASOLEIMAN, AMIN KARBASI, ETH Zurich, RIK SARKAR, University of Edinburgh, and ANDREAS KRAUSE, ETH Zurich	2051
Adaptive Market Making via Online Learning, JACOB ABERNETHY, University of Pennsylvania, and SATYEN KALE, IBM Research	2060
On the Sample Complexity of Subspace Learning, ALESSANDRO RUDI, Istituto Italiano di Tecnologia, GUILLERMO CANAS, and LORENZO ROSASCO, MIT	2069
Spike train entropy-rate estimation using hierarchical Dirichlet process priors, KARIN KNUDSON, and JONATHAN PILLOW, UT Austin ...	2078
Embed and Project: Discrete Sampling with Universal Hashing, STEFANO ERMON, CARLA GOMES, Cornell University, ASHISH SABHARWAL, IBM Watson Research Center, and BART SELMAN, Cornell University	2087
Discriminative Transfer Learning with Tree-based Priors, NITISH SRIVASTAVA, and RUSLAN SALAKHUTDINOV, University of Toronto	2096

Small-Variance Asymptotics for Hidden Markov Models, ANIRBAN ROYCHOWDHURY, KE JIANG, and BRIAN KULIS, Ohio State University	2105
Convergence of Monte Carlo Tree Search in Simultaneous Move Games, VILIAM LISY, VOJTA KOVARIK, CTU in Prague, MARC LANCTOT, Maastricht University, and BRANISLAV BOSANSKY, CTU in Prague	2114
DeViSE: A Deep Visual-Semantic Embedding Model, ANDREA FROME, GREG CORRADO, JON SHLENS, SAMY BENGIO, JEFF DEAN, MARC'AURELIO RANZATO, and TOMAS MIKOLOV, Google Research	2123
Reward Mapping for Transfer in Long-Lived Agents, XIAOXIAO GUO, SATINDER SINGH, and RICHARD LEWIS, University of Michigan	2132
Minimax Theory for High-dimensional Gaussian Mixtures with Sparse Mean Separation, MARTIN AZIZYAN, AARTI SINGH, and LARRY WASSERMAN, Carnegie Mellon University	2141
Predicting Parameters in Deep Learning, MISHA DENIL, BABAK SHAKIBI, University of British Columbia, LAURENT DINH, École Centrale Paris, MARC'AURELIO RANZATO, Google Research, and NANDO DE FREITAS, University of British Columbia	2150
Estimating the Unseen: Improved Estimators for Entropy and other Properties, PAUL VALIANT, Brown University, and GREGORY VALIANT, Stanford University	2159
What do row and column marginals reveal about your dataset?, BEHZAD GOLSHAN, JOHN BYERS, and EVIMARIA TERZI, Boston University .	2168
RNADE: The real-valued neural autoregressive density-estimator, BENIGNO URIA, IAIN MURRAY, University of Edinburgh, and HUGO LAROCHELLE, Université de Sherbrooke (Quebec)	2177
Two-Target Algorithms for Infinite-Armed Bandits with Bernoulli Rewards, THOMAS BONALD, Telecom ParisTech, and ALEXANDRE PROUTIERE, KTH	2186
Reconciling "priors" & "priors" without prejudice?, REMI GRIBONVAL, and PIERRE MACHART, INRIA	2195
Sparse Overlapping Sets Lasso for Multitask Learning and its Application to fMRI Analysis, NIKHIL RAO, CHRISTOPHER COX, ROB NOWAK, and TIMOTHY ROGERS, UW-Madison	2204
Sensor Selection in High-Dimensional Gaussian Trees with Nuisances, DANIEL LEVINE, and JONATHAN HOW, MIT	2213
Sequential Transfer in Multi-armed Bandit with Finite Set of Models, MOHAMMAD GHESHLAGHI AZAR, Carnegie Mellon University, ALESSANDRO LAZARIC, INRIA, and EMMA BRUNSKILL, Carnegie Mellon University	2222
Buy-in-Bulk Active Learning, LIU YANG, and JAIME CARBONELL, Carnegie Mellon University	2231
Contrastive Learning Using Spectral Methods, JAMES ZOU, Harvard University, DANIEL HSU, Columbia University, DAVID PARKES, and RYAN ADAMS, Harvard University	2240
Message Passing Inference with Chemical Reaction Networks, NILS NAPP, and RYAN ADAMS, Harvard University	2249

Eluder Dimension and the Sample Complexity of Optimistic Exploration, DAN RUSSO, and BENJAMIN VAN ROY, Stanford University .	2258
Learning word embeddings efficiently with noise-contrastive estimation, ANDRIY MNIH, Gatsby Unit, UCL, and KORAY KAVUKCUOGLU, NEC Labs .	2267
Sparse Inverse Covariance Estimation with Calibration, TUO ZHAO, Johns Hopkins University, and HAN LIU, Princeton University .	2276
Stochastic Majorization-Minimization Algorithms for Large-Scale Optimization, JULIEN MAIRAL, INRIA .	2285
Sinkhorn Distances: Lightspeed Computation of Optimal Transportation, MARCO CUTURI, Kyoto University .	2294
Speedup Matrix Completion with Side Information: Application to Multi-Label Learning, MIAO XU, Nanjing University, RONG JIN, Michigan State University (MSU), and ZHI-HUA ZHOU, Nanjing University .	2303
Compete to Compute, RUPESH SRIVASTAVA, JONATHAN MASCI, SOHROB KAZEROUNIAN, FAUSTINO GOMEZ, and JÜRGEN SCHMIDHUBER, IDSIA .	2312
Fast Determinantal Point Process Sampling with Application to Clustering, BYUNGKON KANG, Samsung Electronics .	2321
Information-theoretic lower bounds for distributed statistical estimation with communication constraints, YUCHEN ZHANG, JOHN DUCHI, MICHAEL JORDAN, and MARTIN WAINWRIGHT, UC Berkeley .	2330
Projected Natural Actor-Critic, PHILIP THOMAS, WILLIAM DABNEY, SRIDHAR MAHADEVAN, and STEPHEN GIGUERE, University of Massachusetts Amherst .	2339
How to Hedge an Option Against an Adversary: Black-Scholes Pricing is Minimax Optimal, JACOB ABERNETHY, University of Pennsylvania, PETER BARTLETT, UC Berkeley, RAFAEL FRONGILLO, Microsoft Research, and ANDRE WIBISONO, UC Berkeley .	2348
Discovering Hidden Variables in Noisy-Or Networks using Quartet Tests, YACINE JERNITE, Courant Institute, NYU, YONATAN HALPERN, and DAVID SONTAG, NYU .	2357
Error-Minimizing Estimates and Universal Entry-Wise Error Bounds for Low-Rank Matrix Completion, FRANZ KIRALY, TU Berlin, and LOUIS THERAN, Freie Universität Berlin .	2366
Learning the Local Statistics of Optical Flow, DAN ROSENBAUM, DANIEL ZORAN, and YAIR WEISS, Hebrew University .	2375
Aggregating Optimistic Planning Trees for Solving Markov Decision Processes, GUNNAR KEDENBURG, INRIA, RAPHAEL FONTENEAU, Université de Liège, and REMI MUNOS, INRIA .	2384
Robust learning of low-dimensional dynamics from large neural ensembles, DAVID PFAU, EFTYCHIOS PNEVMATIKAKIS, and LIAM PANINSKI, Columbia University .	2393
Estimation Bias in Multi-Armed Bandit Algorithms for Search Advertising, MIN XU, Carnegie Mellon University, TAO QIN, and TIE-YAN LIU, Microsoft Research .	2402

Action is in the Eye of the Beholder: Eye-gaze Driven Model for Spatio-Temporal Action Localization, NATALIYA SHAPOVALOVA, Simon Fraser University, MICHALIS RAPTIS, LEONID SIGAL, Disney Research, and GREG MORI, Simon Fraser University	2411
A* Lasso for Learning a Sparse Bayesian Network Structure for Continuous Variables, JING XIANG, and SEYOUNG KIM, Carnegie Mellon University	2420
The Total Variation on Hypergraphs - Learning on Hypergraphs Revisited, MATTHIAS HEIN, SIMON SETZER, LEONARDO JOST, and SYAMA SUNDAR RANGAPURAM, Saarland University	2429
Submodular Optimization with Submodular Cover and Submodular Knapsack Constraints, RISHABH IYER, and JEFF BILMES, University of Washington	2438
Scalable Inference for Logistic-Normal Topic Models, JIANFEI CHEN, JUNE ZHU, ZI WANG, XUN ZHENG, and BO ZHANG, Tsinghua University	2447
Spectral methods for neural characterization using generalized quadratic models, IL PARK, EVAN ARCHER, NICHOLAS PRIEBE, and JONATHAN PILLOW, UT Austin	2456
Universal models for binary spike patterns using centered Dirichlet processes, IL PARK, EVAN ARCHER, KENNETH LATIMER, and JONATHAN PILLOW, UT Austin	2465
Synthesizing Robust Plans under Incomplete Domain Models, TUAN NGUYEN, SUBBARAO KAMBHAMPATI, Arizona State University, and MINH DO, NASA	2474
Integrated Non-Factorized Variational Inference, SHAOBO HAN, XUEJUN LIAO, and LAWRENCE CARIN, Duke University	2483
Auxiliary-variable Exact Hamiltonian Monte Carlo Samplers for Binary Distributions, ARI PAKMAN, and LIAM PANINSKI, Columbia University	2492
Symbolic Opportunistic Policy Iteration for Factored-Action MDPs, ASWIN RAGHAVAN, Oregon State University, RONI KHARDON, Tufts University, ALAN FERN, and PRASAD TADEPALLI, Oregon State University ...	2501
Online Learning in Markov Decision Processes with Adversarially Chosen Transition Probability Distributions, YASIN ABBASI, Queensland University of Technology, PETER BARTLETT, VARUN KANADE, UC Berkeley, YEVGENY SELDIN, Queensland University of Technology & UC Berkeley, and CSABA SZEPESVARI, University of Alberta	2510
Flexible sampling of discrete data correlations without the marginal distributions, ALFREDO KALAITZIS, and RICARDO SILVA, UCL	2519
One-shot learning by inverting a compositional causal process, BRENDEN LAKE, MIT, RUSLAN SALAKHUTDINOV, University of Toronto, and JOSH TENENBAUM, MIT	2528
Statistical analysis of coupled time series with Kernel Cross-Spectral Density operators., MICHEL BESSERVE, MPI for Intelligent Systems, NIKOS LOGOTHETIS, MPI for Biological Cybernetics, and BERNHARD SCHÖLKOPF, MPI Tübingen	2537

Fast Algorithms for Gaussian Noise Invariant Independent Component Analysis, JAMES VOSS, LUIS RADEMACHER, and MIKHAIL BELKIN, Ohio State University	2546
Deep Neural Networks for Object Detection, CHRISTIAN SZEGEDY, ALEXANDER TOSHEV, and DUMITRU ERHAN, Google Research	2555
Geometric optimisation on positive definite matrices with application to elliptically contoured distributions, SUVRIT SRA, MPI for Intelligent Systems & Carnegie Mellon University, and RESHAD HOSSEINI, MPI Tübingen	2564
Sign Cauchy Projections and Chi-Square Kernel, PING LI, GENNADY SAMORODNITSK, and JOHN HOPCROFT, Cornell University	2573
Relevance Topic Model for Unstructured Social Group Activity Recognition, FANG ZHAO, YONGZHEN HUANG, LIANG WANG, and TIENIU TAN, Chinese Academy of Sciences	2582
k-Prototype Learning for 3D Rigid Structures, HU DING, SUNY at Buffalo, RONALD BEREZNEY, University of Buffalo, and JINHUI XU, SUNY at Buffalo	2591
Restricting exchangeable nonparametric distributions, SINEAD WILLIAMSON, UT Austin, STEVE MACEACHERN, Ohio State University, and ERIC XING, Carnegie Mellon University	2600
Forgetful Bayes and myopic planning: Human learning and decision-making in a bandit setting, SHUNAN ZHANG, and ANGELA YU, UC San Diego	2609
Probabilistic Movement Primitives, ALEXANDROS PARASCHOS, CHRISTIAN DANIEL, JANUARY PETERS, and GERHARD NEUMANN, TU Darmstadt	2618
Policy Shaping: Integrating Human Feedback with Reinforcement Learning, SHANE GRIFFITH, KAUSHIK SUBRAMANIAN, JONATHAN SCHOLZ, CHARLES ISBELL, and ANDREA THOMAZ, Georgia Tech	2627
Multilinear Dynamical Systems for Tensor Time Series, MARK ROGERS, LEI LI, and STUART RUSSELL, UC Berkeley	2636
Deep content-based music recommendation, AARON VAN DEN OORD, SANDER DIELEMAN, and BENJAMIN SCHRAUWEN, Ghent University .	2645
A Stability-based Validation Procedure for Differentially Private Machine Learning, KAMALIKA CHAUDHURI, and STAAL VINTERBO, UC San Diego	2654
Capacity of strong attractor patterns to model behavioural and cognitive prototypes, ABBAS EDALAT, Imperial College London	2673
Fantope Projection and Selection: A near-optimal convex relaxation of sparse PCA, VINCENT VU, Ohio State University, JUHEE CHO, UW-Madison, JING LEI, Carnegie Mellon University, and KARL ROHE, UW-Madison	2682
Cluster Trees on Manifolds, SIVARAMAN BALAKRISHNAN, SRIVATSAN NARAYANAN, ALESSANDRO RINALDO, AARTI SINGH, and LARRY WASSERMAN, Carnegie Mellon University	2691

Bayesian inference for low rank spatiotemporal neural receptive fields, MIJUNG PARK, University of Texas, and JONATHAN PILLOW, UT Austin	2700
Adaptive Submodular Maximization in Bandit Setting, VICTOR GABILLON, INRIA, BRANISLAV KVETON, Technicolor Labs, ZHENG WEN, Stanford University, BRIAN ERIKSSON, Technicolor Labs, and S. MUTHUKRISHNAN, Rutgers University	2709
Generalized Method-of-Moments for Rank Aggregation, HOSSEIN AZARI SOUFIANI, WILLIAM CHEN, DAVID PARKES, and LIRONG XIA, Harvard University	2718
Analyzing Hogwild Parallel Gaussian Gibbs Sampling, MATTHEW JOHNSON, JAMES SAUNDERSON, and ALAN WILLSKY, MIT	2727
Minimax Optimal Algorithms for Unconstrained Linear Optimization, BRENDAN MCMAHAN, Google Research, and JACOB ABERNETHY, University of Pennsylvania	2736
(Nearly) Optimal Algorithms for Private Online Learning in Full-information and Bandit Settings, ABHRADEEP GUHA THAKURTA, Stanford University & Microsoft, and ADAM SMITH, Penn State University	2745
Curvature and Optimal Algorithms for Learning and Minimizing Submodular Functions, RISHABH IYER, University of Washington, STEFANIE JEGELKA, UC Berkeley, and JEFF BILMES, University of Washington	2754
Optimality for Active Learning on Gaussian Random Fields, YIFEI MA, Carnegie Mellon University, ROMAN GARNETT, University of Bonn, and JEFF SCHNEIDER, Carnegie Mellon University	2763
Learning Kernels Using Local Rademacher Complexity, CORINNA CORTES, Google Research, MARIUS KLOFT, Courant Institute, NYU & Sloan-Kettering Institute (MSKCC), and MEHRYAR MOHRI, Courant Institute, NYU & Google	2772
Annealing between distributions by averaging moments, ROGER GROSSE, MIT, CHRIS MADDISON, and RUSLAN SALAKHUTDINOV, University of Toronto	2781
Optimizing Instructional Policies, ROBERT LINDSEY, MICHAEL MOZER, WILLIAM HUGGINS, University of Colorado, and HAROLD PASHLER, UC San Diego	2790
Translating Embeddings for Modeling Multi-relational Data, ANTOINE BORDES, NICOLAS USUNIER, ALBERTO GARCIA-DURAN, Université de Technologie de Compiègne (UTC), JASON WESTON, and OKSANA YAKHNENKO, Google Research	2799
Phase Retrieval using Alternating Minimization, PRANEETH NETRAPALLI, UT Austin, PRATEEK JAIN, Microsoft Research, and SUJAY SANGHAVI, UT Austin	2808
Real-Time Inference for a Gamma Process Model of Neural Spiking, DAVID CARLSON, Duke University, VINAYAK RAO, Gatsby Unit, UCL, JOSHUA VOGELSTEIN, and LAWRENCE CARIN, Duke University	2817
Understanding Dropout, PIERRE BALDI, and PETER SADOWSKI, UC Irvine	2826

The Power of Asymmetry in Binary Hashing, BEHNAM NEYSHABUR, PAYMAN YADOLLAHPOUR, YURY MAKARYCHEV, TTI Chicago, RUSLAN SALAKHUTDINOV, University of Toronto, and NATI SREBRO, TTI Chicago	2835
Estimation, Optimization, and Parallelism when Data is Sparse, JOHN DUCHI, MICHAEL JORDAN, UC Berkeley, and BRENDAN MCMAHAN, Google Research	2844
A multi-agent control framework for co-adaptation in brain- computer interfaces, JOSH MEREL, Columbia University, ROY FOX, Hebrew University, TONY JEBARA, and LIAM PANINSKI, Columbia University	2853
Modeling Overlapping Communities with Node Popularities, PREM GOPALAN, Princeton University, CHONG WANG, Carnegie Mellon University, and DAVID BLEI, Princeton University	2862
Learning from Limited Demonstrations, BEOMJOON KIM, AMIR MASSOUD FARAHMAND, JOELLE PINEAU, and DOINA PRECUP, McGill University	2871
On the Complexity and Approximation of Binary Evidence in Lifted Inference, GUY VAN DEN BROECK, and ADNAN DARWICHE, University of California, Los Angeles	2880
On the Representational Efficiency of Restricted Boltzmann Machines, JAMES MARTENS, University of Toronto, ARKADEV CHATTOPADHYA, Tata Institute of Fundamental Research, TONI PITASSI, and RICHARD ZEMEL, University of Toronto	2889
Memory Limited, Streaming PCA, IOANNIS MITLIAGKAS, CONSTANTINE CARAMANIS, UT Austin, and PRATEEK JAIN, Microsoft Research	2910
An Approximate, Efficient Solver for LP Rounding, SRIKRISHNA SRIDHAR, STEPHEN WRIGHT, CHRISTOPHER RE, JI LIU, VICTOR BITTORF, and CE ZHANG, UW-Madison	2919
Linear decision rule as aspiration for simple decision heuristics, OZGUR SIMSEK, Max Planck Institute Berlin	2928
On the Relationship Between Binary Classification, Bipartite Ranking, and Binary Class Probability Estimation, HARIKRISHNA NARASIMHAN, and SHIVANI AGARWAL, Indian Institute of Science	2937
Bayesian inference as iterated random functions with applications to sequential inference in graphical models, ARASH AMINI, and XUANLONG NGUYEN, University of Michigan	2946
Compressive Feature Learning, HRISTO PASKOV, ROBERT WEST, JOHN MITCHELL, and TREVOR HASTIE, Stanford University	2955
Moment-based Uniform Deviation Bounds for k-means and Friends, MATUS TELGARSKY, and SANJOY DASGUPTA, UC San Diego	2964
Fast Template Evaluation with Vector Quantization, MOHAMMAD AMIN SADEGHI, and DAVID FORSYTH, University of Illinois at Urbana-Champaign	2973
Context-sensitive active sensing in humans, SHEERAZ AHMAD, HE HUANG, and ANGELA YU, UC San Diego	2982

A New Convex Relaxation for Tensor Completion, BERNARDINO ROMERA-PAREDES, and MASSIMILIANO PONTIL, UCL	2991
Variational Planning for Graph-based MDPs, QIANG CHENG, Tsinghua University, QIANG LIU, UC Irvine, FENG CHEN, Tsinghua University, and ALEX IHLER, UC Irvine	3000
Convex Two-Layer Modeling, ÖZLEM ASLAN, HAO CHENG, DALE SCHUURMANS, University of Alberta, and XINHUA ZHANG, NICTA	3009
Sketching Structured Matrices for Faster Nonlinear Regression, HAIM AVRON, VIKAS SINDHWANI, and DAVID WOODRUFF, IBM Research ..	3018
(More) Efficient Reinforcement Learning via Posterior Sampling, IAN OSBAND, DAN RUSSO, and BENJAMIN VAN ROY, Stanford University ..	3027
Model Selection for High-Dimensional Regression under the Generalized Irrepresentability Condition, ADEL JAVANMARD, and ANDREA MONTANARI, Stanford University	3037
Efficient Exploration and Value Function Generalization in Deterministic Systems, ZHENG WEN, and BENJAMIN VAN ROY, Stanford University	3046
Bellman Error Based Feature Generation using Random Projections on Sparse Spaces, MAHDI MILANI FARD, YURI GRINBERG, AMIR MASSOUD FARAHMAND, JOELLE PINEAU, and DOINA PRECUP, McGill University	3055
Learning and using language via recursive pragmatic reasoning about other agents, NATHANIEL SMITH, University of Edinburgh, NOAH GOODMAN, and MICHAEL FRANK, Stanford University	3064
Learning Stochastic Inverses, ANDREAS STUHLMÜLLER, MIT, JACOB TAYLOR, and NOAH GOODMAN, Stanford University	3073
Learning invariant representations and applications to face verification , QIANLI LIAO, JOEL LEIBO, and TOMASO POGGIO, MIT ...	3082
Optimization, Learning, and Games with Predictable Sequences, SASHA RAKHLIN, and KARTHIK SRIDHARAN, University of Pennsylvania	3091
Adaptivity to Local Smoothness and Dimension in Kernel Regression, SAMORY KPOTUFE, and VIKAS GARG, TTI Chicago	3100
Adaptive dropout for training deep neural networks, JIMMY BA, and BRENDAN FREY, University of Toronto	3109
Hierarchical Modular Optimization of Convolutional Networks Achieves Representations Similar to Macaque IT and Human Ventral Stream, DANIEL YAMINS, HA HONG, CHARLES CADIEU, and JAMES DiCARLO, MIT	3118
Stochastic Gradient Riemannian Langevin Dynamics on the Probability Simplex, SAM PATTERSON, Gatsby Unit, UCL, and YEE WHYIE TEH, University of Oxford	3127
Distributed Representations of Words and Phrases and their Compositionality, TOMAS MIKOLOV, ILYA SUTSKEVER, KAI CHEN, GREG CORRADO, and JEFF DEAN, Google Research	3136
Regularized Spectral Clustering under the Degree-Corrected Stochastic Blockmodel, TAI QIN, and KARL ROHE, UW-Madison	3145

Analyzing the Harmonic Structure in Graph-Based Learning, XIAO-MING WU, ZHENGUO LI, and SHIH-FU CHANG, Columbia University	3154
Recurrent linear models of simultaneously-recorded neural populations, MARIUS PACHITARIU, Gatsby Unit, UCL, BILJANA PETRESKA, UCL, and MANEESH SAHANI, Gatsby Unit, UCL	3163
Scalable Influence Estimation in Continuous-Time Diffusion Networks, NAN DU, LE SONG, Georgia Tech, MANUEL GOMEZ-RODRIGUEZ, MPI for Intelligent Systems, and HONGYUAN ZHA, Georgia Tech	3172
Bayesian Inference and Learning in Gaussian Process State-Space Models with Particle MCMC, ROGER FRIGOLA, University of Cambridge, FREDRIK LINDSTEN, Linköping University, THOMAS SCHON, Uppsala University, and CARL RASMUSSEN, University of Cambridge	3181
BIG & QUIC: Sparse Inverse Covariance Estimation for a Million Variables, CHO-JUI HSIEH, UT Austin, MATYAS SUSTIK, INDERJIT DHILLON, University of Texas, PRADEEP RAVIKUMAR, UT Austin, and RUSSELL POLDRACK, University of Texas	3190
The Fast Convergence of Incremental PCA, AKSHAY BALSUBRAMANI, SANJOY DASGUPTA, and YOAV FREUND, UC San Diego	3199
Multisensory Encoding, Decoding, and Identification, AUREL LAZAR, and YEVGENIY SLUTSKIY, Columbia University	3208
Adaptive Anonymity via b-Matching, KRZYSZTOF CHOROMANSKI, Google Research, TONY JEBARA, and KUI TANG, Columbia University	3217
Optimal integration of visual speed across different spatiotemporal frequency channels, MATJAZ JOGAN, and ALAN STOCKER, University of Pennsylvania	3226
Matrix factorization with Binary Components, MARTIN SLAWSKI, MATTHIAS HEIN, and PAVLO LUTSIK, Saarland University	3235
Learning to Pass Expectation Propagation Messages, NICOLAS HEESS, Gatsby Unit, UCL, DANIEL TARLOW, and JOHN WINN, Microsoft Research	3244
Robust Low Rank Kernel Embeddings of Multivariate Distributions, LE SONG, and BO DAI, Georgia Tech	3253
Author Index	