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 xxxi

NIPS foundation xxxii

Committees xxxiii

Reviewers xxxvi

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 Mevissen, 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 Borgwardt, 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 Lindenbaoum, 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 Cites, 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 Gretton, 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, Zhengdong 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 Boytsov, 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 Whye 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 Savchynskyy, 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 Vainsencher, 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 Whye 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 Representative 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 Whye 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