## Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Preface</td>
<td>xxxi</td>
</tr>
<tr>
<td>Donors</td>
<td>xxxv</td>
</tr>
<tr>
<td>NIPS foundation</td>
<td>xxxvi</td>
</tr>
<tr>
<td>Committees</td>
<td>xxxvi</td>
</tr>
<tr>
<td>Kernel Mean Estimation via Spectral Filtering, Krikamol Muandet,</td>
<td>1</td>
</tr>
<tr>
<td>MPI for Intelligent Systems, Bharath Sriperumbudur, supervised</td>
<td></td>
</tr>
<tr>
<td>learning, and Bernhard Schoelkopf, Max Planck Institute for Biological</td>
<td></td>
</tr>
<tr>
<td>Cybernetics</td>
<td></td>
</tr>
<tr>
<td>Semi-Sepparable Hamiltonian Monte Carlo for Inference in Bayesian</td>
<td>10</td>
</tr>
<tr>
<td>Hierarchical Models, Yichuan Zhang, Charles Sutton, University of</td>
<td></td>
</tr>
<tr>
<td>Edinburgh</td>
<td></td>
</tr>
<tr>
<td>Communication Efficient Distributed Machine Learning with the Parameter</td>
<td>19</td>
</tr>
<tr>
<td>Server, Mu Li, David Andersen, Carnegie Mellon University, and Alex</td>
<td></td>
</tr>
<tr>
<td>Smola, Yahoo! Research, Kai Yu, Baidu IDL</td>
<td></td>
</tr>
<tr>
<td>The Infinite Mixture of Infinite Gaussian Mixtures, Halid Yerebakan,</td>
<td>28</td>
</tr>
<tr>
<td>IUPUI, Bartek Rajwa, Purdue University, and Murat Dundar, IUPUI</td>
<td></td>
</tr>
<tr>
<td>Robust Classification Under Sample Selection Bias, Anqi Liu, Univ of</td>
<td>37</td>
</tr>
<tr>
<td>Illinois at Chicago, Brian Ziebart, University of Illinois Chicago</td>
<td></td>
</tr>
<tr>
<td>Zeta Hull Pursuits: Learning Nonconvex Data Hulls, Yuanjun Xiong,</td>
<td>46</td>
</tr>
<tr>
<td>The Chinese University of HK, Wei Liu, IBM T. J. Watson Research</td>
<td></td>
</tr>
<tr>
<td>Center, and Deli Zhao, Advanced Algorithm Research Group HTC Beijing,</td>
<td></td>
</tr>
<tr>
<td>Xiaoou Tang, Chinese University of Hong Kong</td>
<td></td>
</tr>
<tr>
<td>Grouping-Based Low-Rank Trajectory Completion and 3D Reconstruction,</td>
<td>55</td>
</tr>
<tr>
<td>Aikaterini Fragkiadaki, University of California Berk, Marta Salas,</td>
<td></td>
</tr>
<tr>
<td>and Pablo Arbelaez, Jitendra Malik, UC Berkeley</td>
<td></td>
</tr>
<tr>
<td>Sparse Space-Time Deconvolution for Calcium Image Analysis, Ferran</td>
<td>64</td>
</tr>
<tr>
<td>Diego Andilla, HCI IWR Uni Heidelberg, Fred Hamprecht</td>
<td></td>
</tr>
<tr>
<td>Restricted Boltzmann machines modeling human choice, Takayuki Osogami,</td>
<td>73</td>
</tr>
<tr>
<td>Makoto Otsuka, IBM Research - Tokyo</td>
<td></td>
</tr>
<tr>
<td>Multiscale Fields of Patterns, Pedro Felzenszwalb, John Oberlin,</td>
<td>82</td>
</tr>
<tr>
<td>Brown University</td>
<td></td>
</tr>
</tbody>
</table>
large scale canonical correlation analysis with iterative least squares, **Yichao Lu, Dean Foster** .................................. 91

**Altitude Training: Strong Bounds for Single-Layer Dropout,**
**Stefan Wager, William Fithian,** and **Sida Wang, Percy Liang,**
Stanford University .......................................................... 100

**Rounding-based Moves for Metric Labeling,** **M. Kumar,** Ecole Centrale Paris ....................................................... 109

**Parallel Double Greedy Submodular Maximization,** **Xinghao Pan, Stefanie Jegelka,** and **Joseph Gonzalez,** UC Berkeley, Joseph Bradley, Databricks, Michael Jordan, UC Berkeley ......................... 118

**Multivariate Regression with Calibration,** **Han Liu,** supervised learning, **Lie Wang,** Massachusetts Institute of Technology, and **Tuò Zhao** ... 127

**Exact Post Model Selection Inference for Marginal Screening,**
**Jason Lee, Jonathan Taylor,** Stanford University ......................... 136

**On a Theory of Nonparametric Pairwise Similarity for Clustering:**
**Connecting Clustering to Classification,** **Yingzhen Yang,**
University of Illinois at Urbana-Champaign, **Feng Liang,** Univ. of Illinois Urbana-Champaign Statistics, and **Shuicheng Yan,** NUS Singapore, **Zhangyang Wang,** University of Illinois at Urbana-Champaign, **Thomas Huang** ......................................................... 145

**Just-In-Time Learning for Fast and Flexible Inference,** **S. Eslami,**
Daniel Tarlow, Microsoft Research, and **Pushmeet Kohli,** Microsoft Research Cambridge, **John Winn,** Microsoft Research .................. 154

**Fundamental Limits of Online and Distributed Algorithms for Statistical Learning and Estimation,** **Ohad Shamir,** The Weizmann Institute ...................................................... 163

**Quantized Kernel Learning for Feature Matching,** **Danfeng Qin,** Eidgenössische Technische Hochschule Zürich, **Xuanli Chen,** Computer Aided Medical Procedures and Augmented Reality TU Munich, and **Matthieu Guillaumin,** Eidgenössische Technische Hochschule Zürich, **Luc Gool,** Computer Vision Lab ETH Zurich ........................................... 172

**Parallel Direction Method of Multipliers,** **Huahua Wang, Arindam Banerjee,** and **Zhi-Quan Luo,** University of Minnesota .............. 181

**(Almost) No Label No Cry,** **Giorgio Patrini,** Australian National University, **Richard Nock,** and **Tiberio Caetano,** NICTA Canberra, **Paul Rivera** ................................................................. 190

**Stochastic Multi-Armed-Bandit Problem with Non-stationary Rewards,** **Yonatan Gur,** Stanford University, **Assaf Zeevi,** and **Omar Besbes,** Columbia University ................................. 199

**Object Localization based on Structural SVM using Privileged Information,** **Jan Feyereisl,** Samsung, **Suha Kwak,** INRIA, and **Jeany Son,** POSTECH, **Bohyung Han,** POSTCH .............................. 208

**Multi-View Perceptron: a Deep Model for Learning Face Identity and View Representations,** **Zhenyao Zhu,** **Ping Luo,** Chinese University of Hong Kong, and **Xiaogang Wang, Xiaoou Tang,** Chinese University of Hong Kong .............................. 217
Shape and Illumination from Shading using the Generic Viewpoint Assumption, Daniel Zoran, Massachusetts Institute of Technology, Dilip Krishnan, New York University, and Jose Bento, Boston College, William Freeman, Massachusetts Institute of Technology 226

Parallel Sampling of HDPs using Sub-Cluster Splits, Jason Chang, John Fisher III, Massachusetts Institute of Technology 235

From MAP to Marginals: Variational Inference in Bayesian Submodular Models, Josip Djolonga, Andreas Krause, Eidgenössische Technische Hochschule Zürich 244

Robust Logistic Regression and Classification, Jiashi Feng, Huan Xu, National University of Singapore, and Shie Mannor, Technion, Shuicheng Yan, NUS Singapore 253

Extracting Certainty from Uncertainty: Transductive Pairwise Classification from Pairwise Similarities, Tianbao Yang, NEC Labs, Rong Jin, unsupervised learning 262

A Unified Semantic Embedding: Relating Taxonomies and Attributes, Sung Hwang, Disney Research, Leonid Sigal 271

Transportability from Multiple Environments with Limited Experiments: Completeness Results, Elias Bareinboim, Judea Pearl, University of California Los Angeles 280

Augmentative Message Passing for Traveling Salesman Problem and Graph Partitioning, Siamak Ravanbakhsh, Reihaneh Rabbany, and Russell Greiner, University of Alberta 289

Causal Inference through a Witness Protection Program, Ricardo Silva, University College London, Robin Evans, University of Oxford 298

Incremental Clustering: The Case for Extra Clusters, Margareta Ackerman, University of Waterloo, Sanjoy Dasgupta, UC San Diego 307

Multi-scale Graphical Models for Spatio-Temporal Processes, Firdaus Janoos, Exxonmobil, Huseyin Denli, ExxonMobil, and Niranjan Subrahmanya, ExxonMobil Research 316

Iterative Neural Autoregressive Distribution Estimator NADE-k, Tapani Raiko, Aalto University, Li Yao, and Kyunghyun Cho, University of Montreal, Yoshua Bengio, U. Montreal 325

Sparse PCA via Covariance Thresholding, Yash Deshpande, Andrea Montanari, Stanford University 334

Low-dimensional models of neural population activity in sensory cortical circuits, Evan Archer, MPI Tübingen, Urs Köster, University of California at Berkeley, and Jonathan Pillow, Princeton University, Jakob Macke, Max Planck Institute for Biological Cybernetics 343

A Synaptical Story of Persistent Activity with Graded Lifetime in a Neural System, Yuanyuan Mi, Beijing Normal University, LuoZheng Li, Beijing Normal University, and Dahui Wang, Si Wu, Beijing Normal University 352

A Representation Theory for Ranking Functions, Harsh Pareek, UT Austin, Pradeep Ravikumar, University of Texas-Austin 361
Near-optimal sample compression for nearest neighbors, Lee-Ad Gottlieb, Aryeh Kontorovitch, Ben Gurion University, and Pinhas Nisnevitch, Ariel University ................................. 370

Combinatorial Pure Exploration of Multi-Armed Bandits, Shouyuan Chen, The Chinese University of Hong Kong, Tian Lin, Tsinghua University, and Irwin King, Michael Lyu, The Chinese University of Hong Kong, Wei Chen, Microsoft ........................................ 379

Log-Hilbert-Schmidt metric between positive definite operators on Hilbert spaces, Minh Quang, Istituto Italiano di Tecnologi, Marco Biagio, istituto italiano di tecnologia, and Vittorio Murino, Italian Institute of Technology ................................. 388

Consistency of Spectral Partitioning of Uniform Hypergraphs under Planted Partition Model, Debarghya Ghoshdastidar, Ambedkar Dukkipati, Indian Institute of Science ........................................... 397

Spectral Clustering of graphs with the Bethe Hessian, Alaa Saade, Florent Krzakala, Ecole Normale Superieure, and Lenka Zdeborova, Centre national de la recherche scientifique .................. 406

Fast and Robust Least Squares Estimation in Corrupted Linear Models, Brian McWilliams, Gabriel Krummenacher, and Mario Lucic, Eidgenössische Technische Hochschule Zürich, Joachim Buhmann .... 415

Local Decorrelation For Improved Pedestrian Detection, Woonhyun Nam, StradVision Inc., Piotr Dollar, Microsoft, and Joon Han, POSTECH ............................................................... 424

Robust Kernel Density Estimation by Scaling and Projection in Hilbert Space, Robert Vandermeulen, Clayton Scott, University of Michigan ................................................................. 433

Beyond Disagreement-Based Agnostic Active Learning, Chicheng Zhang, UC San Diego, Kamalika Chaudhuri, University of California San Diego ......................................................... 442

Bayes-Adaptive Simulation-based Search with Value Function Approximation, Arthur Guez, Gatsby Unit UCL, Nicolas Heess, Gatsby Unit, and David Silver, Peter Dayan, University College London ... 451

A State-Space Model for Decoding Auditory Attentional Modulation from MEG in a Competing-Speaker Environment, Sahar Akram, Jonathan Simon, and Shihab Shamma, Behtash Babadi, University of Maryland ........................................... 460

Active Regression by Stratification, Sivan Sabato, Microsoft Research, Remi Munos, INRIA Lille - Nord Europe ................................. 469

Sensory Integration and Density Estimation, Joseph Makin, Philip Sabes, University of California San Francisco ........................................... 478

Learning Deep Features for Scene Recognition using Places Database, Bolei Zhou, Massachusetts Institute of Technology, Agata Lapedriza, MIT Universitat Oberta de Catalunya, and Jianxiong Xiao, Princeton University, Antonio Torralba, Aude Oliva, Massachusetts Institute of Technology ................................. 487

A Complete Variational Tracker, Ryan Turner, Steven Bottone, and Bhargav Avasarala, Northrop Grumman Corp ................................. 496
Spike Frequency Adaptation Implements Anticipative Tracking in Continuous Attractor Neural Networks, Yuanyuan Mi, Beijing Normal University, C. Fung, and K. Wong, Si Wu, Beijing Normal University .......................... 505

Efficient Sampling for Learning Sparse Additive Models in High Dimensions, Hemant Tyagi, Bernd Gärtner, and Andreas Krause, Eidgenössische Technische Hochschule Zürich ......................... 514

Deep Joint Task Learning for Generic Object Extraction, Xiaolong Wang, Liliang Zhang, and Liang Lin, Zhuojun Liang, Sun Yat-sen University, Wangmeng Zuo .................................. 523

Robust Bayesian Max-Margin Clustering, Changyou Chen, Australian National University, June Zhu, Tsinghua University, and Xinhua Zhang, NICTA and ANU ........................................... 532

Permutation Diffusion Maps (PDM) with Application to the Image Association Problem in Computer Vision, Deepti Pachauri, University of Wisconsin-Madison, Risi Kondor, University of Chicago, and Gautam Sargur, University of Wisconsin Madison, Vikas Singh ......................................................... 541

Bounded Regret for Finite-Armed Structured Bandits, Tor Lattimore, University of Alberta, Remi Munos, INRIA Lille - Nord Europe .......................... 550

Coresets for k-Segmentation of Streaming Data, Guy Rosman, Mikhail Volkov, and Dan Feldman, John Fisher III, Daniela Rus, Massachusetts Institute of Technology ........................................ 559

Two-Stream Convolutional Networks for Action Recognition in Videos, Karen Simonyan, University of Oxford / Google DeepMind, Andrew Zisserman, computer vision ........................................... 568

Discovering Structure in High-Dimensional Data Through Correlation Explanation, Greg Steeg, Information Sciences Institute, Aram Galstyan, USC Information Sciences Institute .................................................. 577

Positive Curvature and Hamiltonian Monte Carlo, Christof Seiler, Simon Rubinstein-Salzedo, and Susan Holmes, Stanford University ......................................................... 586

Learning Mixed Multinomial Logit Model from Ordinal Data, Sewoong Oh, University of Illinois at Urbana-Champaign, Devavrat Shah, Massachusetts Institute of Technology ........................................ 595

Near-optimal Reinforcement Learning in Factored MDPs, Ian Osband, Benjamin Van Roy, Stanford University ............................................. 604

Efficient learning by implicit exploration in bandit problems with side observations, Tomás Kocák, Inria Lille - Nord Europe, Gergely Neu, INRIA Lille-Nord Europe, and Michael Valko, Remi Munos, INRIA Lille - Nord Europe ............................ 613

Repetitive Contextual Auctions with Strategic Buyers, Kareem Amin, University of Pennsylvania, Afshin Rostamizadeh, Google Research, and Umair Syed, Google .................................................. 622

Recursive Inversion Models for Permutations, Christopher Meek, Microsoft Research, Marina Meila, University of Washington ............... 631
On the Convergence Rate of Decomposable Submodular Function Minimization, Robert Nishihara, Stefanie Jegelka, and Michael Jordan, UC Berkeley ......................................................... 640
New Rules for Domain Independent Lifted MAP Inference, Happy Mittal, Prasoon Goyal, IIT Delhi, and Vibhav Gogate, UT Dallas, Parag Singla, IIT Delhi .............................................. 649
PAC-Bayesian AUC classification and scoring, James Ridgway, CREST-ENSAE, Pierre Alquier, National School of Statistics and Economic Administration, and Nicolas Chopin, CREST, Feng Liang, Univ. of Illinois Urbana-Champaign Statistics .......................................................... 658
On Prior Distributions and Approximate Inference for Structured Variables, Oluwasanmi Koyejo, Rajiv Khanna, University of Texas at Austin, and Joydeep Ghosh, UT Austin, Russell Poldrack, University of Texas at Austin ................................................................. 676
On Iterative Hard Thresholding Methods for High-dimensional M-Estimation, Prateek Jain, Microsoft Research, Ambuj Tewari, University of Michigan, and Purushottam Kar, Microsoft Research ............... 685
Online and Stochastic Gradient Methods for Non-decomposable Loss Functions, Purushottam Kar, Microsoft Research, Harikrishna Narasimhan, Indian Institute of Science Bangalore, and Prateek Jain, Microsoft Research ................................................................. 694
Analysis of Learning from Positive and Unlabeled Data, Marthinus Plessis, Tokyo Institute of Technology, Gang Niu, Baidu, and Masashi Sugiyama, Tokyo Institute of Technology ................................. 703
Dimensionality Reduction with Subspace Structure Preservation, Devansh Arpit, Ifeoma Nwogu, and Venu Govindaraju, SUNY Buffalo ................................................................. 712
Constrained convex minimization via model-based excessive gap, Quoc Tran-Dinh, LIONS EPFL Switzerland, Volkan Cevher, École polytechnique fédérale de Lausanne ......................................................... 721
Poisson Process Jumping between an Unknown Number of Rates: Application to Neural Spike Data, Florian Stimberg, Andreas Ruttor, TU Berlin, and Manfred Opper, ”Computer Science TU Berlin” .... 730
Probabilistic ODE Solvers with Runge-Kutta Means, Michael Schober, MPI for Intelligent Systems, David Duvenaud, University of Cambridge, and Philipp Hennig, MPI Tuebingen .................................... 739
Optimal decision-making with time-varying evidence reliability, Jan Drugowitsch, University of Geneva, Ruben Moreno-Bote, Foundation Sant Joan de Deu, and Alexandre Pouget ......................... 748
Learning Shuffle Ideals Under Restricted Distributions, Dongqu Chen, Yale University ................................................................. 757
Discriminative Unsupervised Feature Learning with Convolutional Neural Networks, Alexey Dosovitskiy, Jost Springenberg, and Martin Riedmiller, Thomas Brox, University of Freiburg ......................... 766
Distance-Based Network Recovery under Feature Correlation, 
DAVID ADAMETZ, VOLKER ROTH, University of Basel .......................... 775

Bandit Convex Optimization: Towards Tight Bounds, ELAD HAZAN, KPIR LEVY, Technion ................................................................. 784

Projective dictionary pair learning for pattern classification, 
SHUHANG GU, The Hong Kong Polytechnic University, LEI ZHANG, The Hong Kong Polytechnic Univ, and WANGMENG ZUO, XIANGCHU FENG, Xidian University .................................................. 793

Provable Submodular Minimization using Wolfe’s Algorithm, 
DEEPAK CHAKRABARTY, PRATEEK JAIN, Microsoft Research, and PRAVESH KOTHARI, UT Austin ............................................................... 802

Exploiting easy data in online optimization, AMIR SANI, INRIA, GERGY NEU, INRIA Lille-Nord Europe, and ALESSANDRO LAZARIC, INRIA Lille ................................................................. 810

Sparse Multi-Task Reinforcement Learning, DANIELE CALANDRIELLO, Politecnico di Milano, ALESSANDRO LAZARIC, INRIA Lille, and MARCELLO RESTELLI, Politecnico di Milano ....................... 819

Best-Arm Identification in Linear Bandits, MARTA SOARE, Inria, ALESSANDRO LAZARIC, INRIA Lille, and REMI MUNOS, INRIA Lille - Nord Europe ................................................................. 828

Mind the Nuisance: Gaussian Process Classification using Privileged Noise, DANIEL HERNÁNDEZ-LOBATO, Universidad Autonoma de Madrid, VIKTORIJA SHARMANSKA, IST Austria, and KRISTIAN KERSTING, TU Dortmund, CHRISTOPH LAMPERT, IST Austria, NOVI QUADRIANTO, SMiLe CLiNiC University of Sussex .................................................. 837

Tight Bounds for Influence in Diffusion Networks and Application to Bond Percolation and Epidemiology, REMI LEMONNIER, ENS Cachan - CMLA, KEVIN SCA MAN, CMLA - ENS Cachan, and NICOLAS VAYATIS, Ecole Normale Supérieure de Cachan ................................................................. 846

On the Computational Efficiency of Training Neural Networks, ROI LIVNI, Hebrew University, SHAI SHALEV-SHWARTZ, Hebrew University of Jerusalem, and OHAD SHAMIR, The Weizmann Institute ......................... 855

Self-Adaptable Templates for Feature Coding, XAVIER BOIX, Eidgenössische Technische Hochschule Zürich, GEMMA ROIG, and SALOMON DIETHER, Eidgenössische Technische Hochschule Zürich, LUC GOOL, Computer Vision Lab ETH Zurich ................................................................. 864

Multilabel Structured Output Learning with Random Spanning Trees of Max-Margin Markov Networks, MARIO MARCHAND, Université Laval, HONGLU SU, Aalto University, and EMILIE MORVANT, University of St-Etienne LaHC, JUHO ROUSU, Aalto University, JOHN SHAWE-TAYLOR, University College London .................................................. 873

Stochastic Network Design in Bidirected Trees, XIAOHIAN WU, DANIEL SHELDON, and SHLOMO ZILBERSTEIN, University of Massachusetts .............................. 882

Learning convolution filters for inverse covariance estimation of neural network connectivity, GEORGE MOHLER, Santa Clara University ................................................................. 900

SerialRank: Spectral Ranking using Seriation, FAJWEL FOGEL, ALEXANDRE D’ASPREMONT, CNRS - ENS, and MILAN VOJNOVIĆ, Microsoft Research ................................................................. 900
Clamping Variables and Approximate Inference, Adrian Weller, Tony Jebara, Columbia University ........................................ 909

Predictive Entropy Search for Efficient Global Optimization of Black-box Functions, José Miguel Hernández-Lobato, Matthew Hoffman, and Zoubin Ghahramani, University of Cambridge ............... 918

A Block-Coordinate Descent Approach for Large-scale Sparse Inverse Covariance Estimation, Eran Treister, Javier Turek, Technion ......................................................... 927

Efficient Inference of Continuous Markov Random Fields with Polynomial Potentials, Shenlong Wang, Alex Schwing, and Raquel Urtasun, University of Toronto ................................. 936

Scalable Methods for Nonnegative Matrix Factorizations of Near-separable Tall-and-skinny Matrices, Austin Benson, Jason Lee, Stanford University, and Bartek Rajwa, David Gleich, Purdue University ........................................... 945

Inferring synaptic conductances from spike trains with a biophysically inspired point process model, Kenneth Latimer, University of Texas at Austin, E. Chichilnisky, Stanford University, and Fred Rieke, University of Washington Seattle, Jonathan Pillow, Princeton University ......................................................... 954

Expectation Backpropagation: Parameter-Free Training of Multilayer Neural Networks with Continuous or Discrete Weights, Daniel Soudry, Statistics department, Itay Hubara, and Ron Meir, Technion ......................................................... 963

Incremental Local Gaussian Regression, Franziska Meier, University of Southern California, Philipp Hennig, MPI Tuebingen, and Stefan Schaal, University of Southern California ................. 972

General Table Completion using a Bayesian Nonparametric Model, Isabel Valera, Universidad Carlos III de Madrid, Zoubin Ghahramani, University of Cambridge ........................................ 981

Universal Option Models, Hengshuai Yao, University of Alberta, Csaba Szepesvari, U. Alberta, and Rich Sutton, University of Alberta, Joseph Modayil, Shalabh Bhatnagar ......................... 990

Approximating Hierarchical MV-sets for Hierarchical Clustering, Assaf Glazer, Omer Weissbrod, Technion, and Michael Lindenbaum, Shaul Markovitch, Technion ......................... 999

Constant Nullspace Strong Convexity and Fast Convergence of Proximal Methods under High-Dimensional Settings, Én-Hsu Yen, UT-Austin, Cho-Jui Hsieh, University of Texas at Austin, and Pradeep Ravikumar, University of Texas-Austin, Inderjit Dhillon, University of Texas ............................................. 1008

Stochastic Gradient Descent, Weighted Sampling, and the Randomized Kaczmarz algorithm, Deanna Needell, Claremont McKenna College, Rachel Ward, Univ. of Texas Austin, and Nati Srebro, Toyota Technological Institute at Chicago ............................................. 1017

A Framework for Testing Identifiability of Bayesian Models of Perception, Luigi Acerbi, University of Edinburgh, Wei Ma, New York University, and Sethu Vijayakumar, University of Edinburgh ............................................. 1026
Optimistic Planning in Markov Decision Processes Using a
Generative Model, BALÁZS SZŐRÉNYI, University of Szeged/INRIA Lille,
GUNNAR KEDENBURG, and REMI MUNOS, INRIA Lille - Nord Europe ........ 1035

Gaussian Process Volatility Model, YUE WU, Cambridge university,
JOSÉ MIGUEL HERNÁNDEZ-LOBATO, and ZOUBIN GHAHRAMANI, University
of Cambridge .................................................. 1044

A Safe Screening Rule for Sparse Logistic Regression, JIE WANG,
Arizona State University, JIAYU ZHOU, Samsung Research, and JUN LIU,
PETER WONKA, JIEPING YE, Arizona State University ....................... 1053

Hardness of parameter estimation in graphical models, GUY
BRESLER, DAVID GAMARNIK, and DEVAVRAT SHAH, Massachusetts Institute
of Technology .......................................................... 1062

Learning Neural Network Policies with Guided Policy Search
under Unknown Dynamics, SERGEY LEVINE, PIETER ABBEEL, UC
Berkeley .............................................................. 1071

Magnitude-sensitive preference formation, NISHEETH SRIVASTAVA,
ED VUL, University of California San Diego, and PAUL SCHRATER, University
of Minnesota ....................................................... 1080

Extreme bandits, ALEXANDRA CARPENTIER, StatsLab Cambridge,
MICHAL Valko, INRIA Lille - Nord Europe .............................. 1089

Distributed Estimation, Information Loss and Exponential
Families, QIANG LIU, University of California Irvine, ALEX IHLER,
probabilistic models ............................................. 1098

Non-convex Robust PCA, PRANEETH NETRAPALLI, University of Texas
at Austin, NIRANJAN N, UC Irvine, and SUJAY SANGHAVI, Univ. of Austin,
ANIMA ANANDKUMAR, UC Irvine, PRATEEK JAIN, Microsoft Research ....... 1107

Simultaneous Model Selection and Optimization through
Parameter-free Stochastic Learning, FRANCESCO ORABONA, Toyota
Technological Institute at Chicago ........................................ 1116

Learning From Weakly Supervised Data by The Expectation Loss
SVM (e-SVM) algorithm, JUN ZHU, JUNHUA MAO, and ALAN YUILLE,
University of California Los Angeles ...................................... 1125

Message Passing Inference for Large Scale Graphical Models
with High Order Potentials, JIAN ZHANG, Eidgenössische Technische
Hochschule Zürich, ALEX SCHWING, and RAQUEL URTASUN, University of
Toronto ............................................................. 1134

Encoding High Dimensional Local Features by Sparse Coding
Based Fisher Vectors, LINGQIAO LIU, University of Adelaide, CHUNHUA
SHEN, and LEI WANG, University of Wollongong, ANTON HENGEL, The
University of Adelaide, CHAO WANG, University of Wollongong ............. 1143

Dependent nonparametric trees for dynamic hierarchical
clustering, KUMAR DUBEY, Carnegie Mellon University, QIRONG HO,
Institute for Infocomm Research ASTAR, and SINEAD WILLIAMSON, UT
Austin, ERIC XING, Carnegie-Mellon University ........................ 1152

Causal Strategic Inference in Networked Microfinance Economies,
MOHAMMAD IRFAN, Bowdoin College, LUIS ORTIZ, Stony Brook University . . 1161
<table>
<thead>
<tr>
<th>Title</th>
<th>Authors</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Conditional Swap Regret and Conditional Correlated Equilibrium,</td>
<td>Mehryar Mohri, Courant Institute and Google, Scott Yang, Courant Institute</td>
<td>1314</td>
</tr>
<tr>
<td>Mode Estimation for High Dimensional Discrete Tree Graphical Models,</td>
<td>Chao Chen, Rutgers University, Han Liu, supervised learning, and Dimitris Metaxas, Rutgers University, Tianqi Zhao, Princeton University</td>
<td>1323</td>
</tr>
<tr>
<td>Large-scale L-BFGS using MapReduce</td>
<td>Weizhu Chen, Zhenghao Wang, and Jingren Zhou, Microsoft</td>
<td>1332</td>
</tr>
<tr>
<td>Submodular Attribute Selection for Action Recognition in Video</td>
<td>Jingjing Zheng, University of Maryland, Zhuolin Jiang, Noah's Ark Lab Huawei Technologies, and Rama Chellappa, University of Maryland College Park, Jonathan Phillips, National Institute of Standards and Technology</td>
<td>1341</td>
</tr>
<tr>
<td>Efficient Structured Matrix Rank Minimization</td>
<td>Adams Yu, Wanli Ma, and Yaoliang Yu, Jaime Carbonell, Carnegie Mellon University</td>
<td>1350</td>
</tr>
<tr>
<td>On Integrated Clustering and Outlier Detection</td>
<td>Lionel Ott, Linsey Pang, Sydney University, and Fabio Ramos, University of Sydney, Sanjay Chawla, Sydney University</td>
<td>1359</td>
</tr>
<tr>
<td>A Drifting-Games Analysis for Online Learning and Applications to</td>
<td>Haipeng Luo, Princeton University, Robert Schapire, Microsoft Research</td>
<td>1368</td>
</tr>
<tr>
<td>Boosting</td>
<td>Projecting Markov Random Field Parameters for Fast Mixing,</td>
<td>1377</td>
</tr>
<tr>
<td>Automatic Discovery of Cognitive Skills to Improve the Prediction</td>
<td>Robert Lindsey, Mohammad Khajah, University of Colorado, and Michael Mozer, &quot;U Colorado Boulder&quot;</td>
<td>1386</td>
</tr>
<tr>
<td>of Student Learning</td>
<td>Near-Optimal-Sample Estimators for Spherical Gaussian Mixtures,</td>
<td>1395</td>
</tr>
<tr>
<td>Automated Variational Inference for Gaussian Process Models,</td>
<td>Trung Nguyen, ANU &amp; NICTA, Edwin Bonilla, NICTA</td>
<td>1404</td>
</tr>
<tr>
<td>Learning Mixtures of Submodular Functions for Image Collection</td>
<td>Learning Mixtures of Submodular Functions for Image Collection Summarization, Sebastian Tschiatschek, Graz University of Technology, Rishabh Iyer, &quot;University of Washington Seattle&quot;, and Haochen Wei, Jeffrey Bilmes, University of Washington</td>
<td>1413</td>
</tr>
<tr>
<td>Robust Tensor Decomposition with Gross Corruption</td>
<td>Quanquan Gu, CS UIUC, Huan Gu, University of Illinois at Urbana, and Jiawei Han, University of Illinois at Urbana-Champaign</td>
<td>1422</td>
</tr>
<tr>
<td>Provable Tensor Factorization with Missing Data,</td>
<td>Prateek Jain, Microsoft Research, Sewoong Oh, University of Illinois at Urbana-Champaign</td>
<td>1431</td>
</tr>
<tr>
<td>Parallel Successive Convex Approximation for Nonsmooth Nonconvex</td>
<td>Meisam Razaviyayn, Mingyi Hong, and Zhi-Quan Luo, University of Minnesota, Jong-Shi Pang, University of Southern California</td>
<td>1440</td>
</tr>
</tbody>
</table>
Convex Optimization Procedure for Clustering: Theoretical Revisit, Changbo Zhu, Huan Xu, National University of Singapore, and Chenlei Leng, Shuicheng Yan, NUS Singapore 1619

Sparse Bayesian structure learning with dependent relevance determination priors, Anqi Wu, Princeton University, Mijung Park, University College London, and Oluwasanmi Koyejo, University of Texas at Austin, Jonathan Pillow, Princeton University 1628

Weakly-supervised Discovery of Visual Pattern Configurations, Hyun Song, Yong Lee, and Stefanie Jegelka, Trevor Darrell, UC Berkeley 1637

SAGA: A Fast Incremental Gradient Method With Support for Non-Strongly Convex Composite Objectives, Aaron Defazio, Ambiata, Francis Bach, INRIA - ENS, and Simon Lacoste-Julien, INRIA 1646

Exclusive Feature Learning on Arbitrary Structures via $\ell_{1,2}$-norm, Deguang Kong, UT Arlington, Ryohei Fujimaki, NECLA, and Ji Liu, Univ. of Wisconsin-Madison, Feiping Nie, Chris Ding, University of Texas at Arlington 1655

Time–Data Tradeoffs by Aggressive Smoothing, John Bruer, Joel Tropp, Caltech, and Volkan Cevher, École polytechnique fédérale de Lausanne, Stephen Becker, Paris 6 1664

Distributed Power-law Graph Computing: Theoretical and Empirical Analysis, Cong Xie, Ling Yan, Shanghai Jiao Tong University, and Wu-Jun Li, Nanjing University China, Zhihua Zhang, Shanghai Jiao Tong University 1673

A Multi-World Approach to Question Answering about Real-World Scenes based on Uncertain Input, Mateusz Malinowski, MPI for Informatics, Mario Fritz, MPI Informatics 1682

Efficient Partial Monitoring with Prior Information, Hastagiri Vanchinathan, Gabor Bartok, and Andreas Krause, Eidgenössische Technische Hochschule Zürich 1691

Distributed Parameter Estimation in Probabilistic Graphical Models, Yariv Mizrahi, University of British Columbia, Misha Denil, and Nando Freitas, University of Oxford 1700

Unsupervised Deep Haar Scattering on Graphs, Xu Chen, Princeton University, XiuYuan Cheng, and Stephane Mallat, Ecole Polytechnique Paris 1709

Online Optimization for Max-Norm Regularization, Jie Shen, Rutgers University, Huan Xu, National University of Singapore, and Ping Li, Rutgers University 1718

Probabilistic low-rank matrix completion on finite alphabets, Jean Lafond, Télécom ParisTech, Olga Kollopp, Université Paris Ouest, and Eric Moulines, Joseph Salmon, Telecom Paristech 1727

Articulated Pose Estimation by a Graphical Model with Image Dependent Pairwise Relations, Xianjie Chen, Alan Yuille, University of California Los Angeles 1736
Bayesian Inference for Structured Spike and Slab Priors, Michael Andersen, Technical University of DK, Ole Winther, and Lars Hansen, Technical University of Denmark ........................................ 1745

Bayesian Nonlinear Support Vector Machines and Discriminative Factor Modeling, Ricardo Henao, Xin Yuan, and Lawrence Carin, Duke University ...................................................... 1754

Generalized Higher-Order Orthogonal Iteration for Tensor Decomposition and Completion, Yuanyuan Liu, Fanhua Shang, The Chinese University of Hong Kong, and Wei Fan, Huawei Noah’s Ark Lab Hong Kong, James Cheng, Hong Cheng, The Chinese University of Hong Kong .................................................. 1763

Making Pairwise Binary Graphical Models Attractive, Nicholas Ruozzi, Tony Jebara, Columbia University ........................................... 1772

Low Rank Approximation Lower Bounds in Row-Update Streams, David Woodruff, IBM Almaden ...................................................... 1781

Deep Convolutional Neural Network for Image Deconvolution, Li Xu, Lenovo IVCL, JimMY Ren, Lenovo Research, and Ce Liu, Microsoft Research, JiAYA Jia, The Chinese University of Hong Kong ........................................... 1790

Joint Training of a Convolutional Network and a Graphical Model for Human Pose Estimation, Jonathan Tompson, Arjun Jain, New York University, and YANN LeCun, NYU Facebook, Christoph Bregler, New York University .................................................. 1799

Learning Generative Models with Visual Attention, Yichuan Tang, Salakhutdinov, Nitish Srivastava, and Ruslan Salakhutdinov, University of Toronto .................................................. 1808

Metric Learning for Temporal Sequence Alignment, Rémi Lajugie, Damien Garreau, École normale supérieure, and Francis Bach, INRIA - ENS, Sylvain Arlot, Centre national de la recherche scientifique .... 1817

Learning Optimal Commitment to Overcome Insecurity, Avrim Blum, Nika Haghtalab, and Ariel Procaccia, Carnegie Mellon University .................................................. 1826

"How hard is my MDP?" The distribution-norm to the rescue, Odalric-Ambrym Maillard, Leoben, Timothy Mann, The Technion, and Shie Mannor, Technion .................................................. 1835

Near-Optimal Density Estimation in Near-Linear Time Using Variable-Width Histograms, Siu Chan, Microsoft Research, Ilias Diakonikolas, University of Edinburgh, and Rocco Servedio, Xiaorui Sun, Columbia University ........................................... 1844

An Autoencoder Approach to Learning Bilingual Word Representations, Sarath Chandar, IIT Madras, Stanislas Lauly, Université de Sherbrooke, and Hugo Larochelle, University de Sherbrooke, MİTEŞH Khapra, IBM India Research Lab, Balaraman Ravindran, IIT Madras, Vikas Raykar, Amrita Saha, IBM India Research Lab .................................................. 1853

Sequential Monte Carlo for Graphical Models, Christian Naesseth, Linköping University, Fredrik Lindsten, University of Cambridge, and Thomas Schon, Uppsala University ........................................... 1862
Optimal Regret Minimization in Posted-Price Auctions with Strategic Buyers, Mehryar Mohri, Courant Institute and Google, Andres Medina, Courant Institute of Mathematical Sciences 1871

Optimal prior-dependent neural population codes under shared input noise, Agnieszka Grabska-Barwinska, University College London, Jonathan Pillow, Princeton University 1880

Deep Fragment Embeddings for Bidirectional Image Sentence Mapping, Andrej Karpathy, Armand Joulin, Stanford University, and Fei Li, Stanford U 1889

Flexible Transfer Learning under Support and Model Shift, Xuezhi Wang, Carnegie Mellon Univ, Jeff Schneider, Carnegie Mellon University 1898

Probabilistic Differential Dynamic Programming, Yunpeng Pan, Georgia Tech, Evangelos Theodorou, Georgia Institute of Technology 1907

Predicting Useful Neighborhoods for Lazy Local Learning, Aron Yu, University of Texas at Austin, Kristen Grauman, U. Texas 1916

Modeling Deep Temporal Dependencies with Recurrent "Grammar Cells", Vincent Michalski, Goethe-University Frankfurt, Roland Memisevic, University of Montreal, and Kishore Konda, Goethe University Frankfurt 1925

Generalized Dantzig Selector: Application to the k-support norm, Soumyadeep Chatterjee, Sheng Chen, and Arindam Banerjee, University of Minnesota 1934

Neurons as Monte Carlo Samplers: Bayesian Inference and Learning in Spiking Networks, Yanping Huang, University Of Washington, Rajesh Rao, University of Washington 1943

The Bayesian Case Model: A Generative Approach for Case-Based Reasoning and Prototype Classification, Been Kim, Massachusetts Institute of Technology, Cynthia Rudin, supervised learning, and Julie Shah, Massachusetts Institute of Technology 1952

Latent Support Measure Machines for Bag-of-Words Data Classification, Yuya Yoshikawa, NAIST, Tomoharu Iwata, NTT, and Hiroshi Sawada, NTT Service Evolution Laboratories 1961

Local Linear Convergence of Forward–Backward under Partial Smoothness, Jingwei Liang, GREYC-CNRS-ENSICAEN, Jalal Fadili, and Gabriel Peyré, Université Paris Dauphine 1970

RAAM: The Benefits of Robustness in Approximating Aggregated MDPs in Reinforcement Learning, Marek Petrik, International Business Machines (IBM), Dharmashankar Subramanian 1979

Deep Learning Face Representation by Joint Identification-Verification, Yi Sun, The Chinese University of Hong Kong, Yuheng Chen, SenseTime Group, and Xiaogang Wang, Xiaoou Tang, Chinese University of Hong Kong 1988

A provable SVD-based algorithm for learning topics in dominant admixture corpus, Trapit Bansal, Chiranjib Bhattacharyya, Indian Institute of Science, and Ravindran Kannan, Indian Institute of Science Bangalore 1997

xvii
Learning the Learning Rate for Prediction with Expert Advice, 
Wouter Koolen, Queensland University of Technology, Tim Erven, and 
Peter Grunwald, CWI Amsterdam ............................................. 2294

On the Information Theoretic Limits of Learning Ising Models, 
Rashish Tandon, The University of Texas at Austin, Karthikeyan 
Shanmugam, The University of Texas at Austin, and Pradeep Ravikumar, 
University of Texas-Austin, Alexandros Dimakis, The University of Texas at 
Austin ................................................................. 2303

Efficient Optimization for Average Precision SVM, Prithish 
Mohapatra, IIIT Hyderabad, C.V. Jawahar, International Institute of 
Information Technology Hyderabad, and M. Kumar, Ecole Centrale Paris ...... 2312

Asymmetric LSH (ALSH) for Sublinear Time Maximum Inner 
Product Search (MIPS), Anshumali Shrivastava, Cornell University, 
Ping Li, Rutgers University ...................................................... 2321

A framework for studying synaptic plasticity with neural spike 
train data, Scott Linderman, Harvard University, Christopher 
Stock, Harvard College, and Ryan Adams, Harvard University ............ 2330

Randomized Experimental Design for Causal Graph Discovery, 
Huining Hu, Mcgill, Zhentao Li, École normale supérieure, and Adrian 
Vetta, Mcgill University ......................................................... 2339

A Multiplicative Model for Learning Distributed Text-Based 
Attribute Representations, Ryan Kiros, Ryan Kiros, Richard Zemel, 
and Ruslan Salakhutdinov, University of Toronto ....................... 2348

Learning Chordal Markov Networks by Dynamic Programming, 
Kustaa Kangas, University of Helsinki, Mikko Koivisto, Helsinki Institute 
for Information Technology, and Teppo Ninnimäki, University of Helsinki ... 2357

Depth Map Prediction from a Single Image using a Multi-Scale 
Deep Network, David Eigen, Christian Puhrsch, and Rob Fergus, 
New York University ........................................................... 2366

Proximal Quasi-Newton for Computationally Intensive L1-
regularized M-estimators, Kai Zhong, UT Austin, En-Hsu Yen, 
UT-Austin, and Inderjit Dhillon, University of Texas, Pradeep 
Ravikumar, University of Texas-Austin .................................. 2375

A Probabilistic Framework for Multimodal Retrieval using 
Integrative Indian Buffet Process, Bahadir Ozdemir, Larry Davis, 
University of Maryland ......................................................... 2384

Searching for Higgs Boson Decay Modes with Deep Learning, 
Peter Sadowski, UC Irvine, Daniel Whiteson, University of California 
Irvine, and Pierre Baldis, probabilistic models ....................... 2393

Structure Regularization for Structured Prediction, Xu Sun, 
Peking University ................................................................. 2402

On Multiplicative Multitask Feature Learning, Xin Wang, Jinbo 
Bi, University of Connecticut, and Shipeng Yu, Siemens, Jianguan Sun, 
University of Connecticut .................................................. 2411

Multivariate f-divergence Estimation With Confidence, Kevin 
Moon, Alfred Hero, University of Michigan ......................... 2420
<table>
<thead>
<tr>
<th>Title</th>
<th>Authors</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Generalized Unsupervised Manifold Alignment</td>
<td>Zhen Cui, ICT, Hong Chang, Institute of Computing Technology, Shiguang Shan, Xilin Chen</td>
<td>2429</td>
</tr>
<tr>
<td>Smoothed Gradients for Stochastic Variational Inference</td>
<td>Stephan Mandt, Princeton University, David Blei, Princeton</td>
<td>2438</td>
</tr>
<tr>
<td>Recursive Context Propagation Network for Semantic Scene Labeling</td>
<td>Abhishek Sharma, University of Maryland, Oncel Tuzel, Ming-Yu Liu, Mitsubishi Electric Research Laboratories</td>
<td>2447</td>
</tr>
<tr>
<td>Sparse Random Feature Algorithm as Coordinate Descent in Hilbert Space</td>
<td>En-Hsu Yen, UT-Austin, Ting-Wei Lin, Shou-De Lin, National Taiwan University, Pradeep Ravi Kumar, University of Texas-Austin, Inderjit Dhillon, University of Texas</td>
<td>2456</td>
</tr>
<tr>
<td>Optimal Teaching for Limited-Capacity Human Learners</td>
<td>Kaustubh Patil, Xiaojin Zhu, University of Wisconsin-Madison, Lukasz Kopec</td>
<td>2465</td>
</tr>
<tr>
<td>Shaping Social Activity by Incentivizing Users</td>
<td>Mehrdad Farajtabar, Georgia Institute of Technology, Nan Du, Georgia Institute of Technology, Manuel Rodriguez, Max Plank Institute for Intelligent Systems, Isabel Valera, Universidad Carlos III de Madrid, Hongyuan Zha, Le Song, Georgia Tech</td>
<td>2474</td>
</tr>
<tr>
<td>Analysis of Brain States from Multi-Region LFP Time-Series</td>
<td>Kyle Ulrich, David Carlson, Duke University, Wenzhao Lian, Duke university, Jana Borg, Kafui Dzirasa, Lawrence Carin, Duke University</td>
<td>2483</td>
</tr>
<tr>
<td>Reputation-based Worker Filtering in Crowdsourcing</td>
<td>Srikanth Jagabathula, Lakshminarayanan Subramanian, Ashwin Venkataraman, New York University</td>
<td>2492</td>
</tr>
<tr>
<td>Multi-Class Deep Boosting</td>
<td>Vitaly Kuznetsov, Courant Institute, Mehryar Mohri, Courant Institute and Google, Umar Syed, Google</td>
<td>2501</td>
</tr>
<tr>
<td>A Differential Equation for Modeling Nesterov’s Accelerated Gradient Method: Theory and Insights</td>
<td>Weijie Su, Stephen Boyd, Emmanuel Candes, Stanford University</td>
<td>2510</td>
</tr>
<tr>
<td>Difference of Convex Functions Programming for Reinforcement Learning</td>
<td>Bilal Piot, Université Lille 3, Matthieu Geist, SUPELEC, Olivier Pietquin, Université de Lille 1</td>
<td>2519</td>
</tr>
<tr>
<td>Design Principles of the Hippocampal Cognitive Map</td>
<td>Kimberly Stachenfeld, Matthew Botvinick, Princeton University, Samuel Gershman, Massachusetts Institute of Technology</td>
<td>2528</td>
</tr>
<tr>
<td>Nonparametric Bayesian inference on multivariate exponential families</td>
<td>William Vega-Brown, Marek Doniec, Nicholas Roy, Massachusetts Institute of Technology</td>
<td>2546</td>
</tr>
<tr>
<td>Optimal rates for k-NN density and mode estimation</td>
<td>Sanjoy Dasgupta, UC San Diego, Samory Kpotufe, Princeton University ORFE</td>
<td>2555</td>
</tr>
<tr>
<td>Feedforward Learning of Mixture Models</td>
<td>Matthew Lawlor, Steven Zucker, Yale University</td>
<td>2564</td>
</tr>
</tbody>
</table>
Cone-Constrained Principal Component Analysis, Yash Deshpande, Andrea Montanari, Stanford University, and Emile Richard, Institut Curie ......................................................... 2717

On Communication Cost of Distributed Statistical Estimation and Dimensionality, Ankit Garg, Tengyu Ma, Princeton, and Huy Nguyen, University of California Berkeley ........................................... 2726

Computing Nash Equilibria in Generalized Interdependent Security Games, Hau Chan, Luis Ortiz, Stony Brook University ....... 2735

Consistent Binary Classification with Generalized Performance Metrics, Oluwasanmi Koyejo, University of Texas at Austin, Nagarajan Natrajan, UT Austin, and Pradeep Ravikumar, University of Texas-Austin, Inderjit Dhillon, University of Texas .................. 2744

Greedy Subspace Clustering, Dohyung Park, The University of Texas at Austin, Constantine Caramanis, UT Austin, and Sujay Sanghavi, University of Texas ................................................................. 2753


Multi-Step Stochastic ADMM in High Dimensions: Applications to Sparse Optimization and Matrix Decomposition, Hanie Sedghi, University of Southern California, Anima Anandkumar, UC Irvine, and Edmond Jonckheere, University of Southern California .................. 2771

Online combinatorial optimization with stochastic decision sets and adversarial losses, Gergely Neu, INRIA Lille-Nord Europe, Michal Valko, INRIA Lille - Nord Europe .................................................. 2780

Sampling for Inference in Probabilistic Models with Fast Bayesian Quadrature, Tom Gunter, Mike Osborne, University of Oxford, and Roman Garnett, University of Bonn, Philipp Hennig, MPI Tuebingen, Stephen Roberts ................................................................. 2789

Multi-Scale Spectral Decomposition of Massive Graphs, Si Si, University of Texas, Donghyuk Shin, University of Texas at Austin, and Inderjit Dhillon, University of Texas, Beresford Parlett, University of California Berkeley .................................................. 2798

The limits of squared Euclidean distance regularization, Michal Derezenski, University of California Santa Cruz, Manfred Warmuth, UC Santa Cruz ................................................................. 2807

Bregman Alternating Direction Method of Multipliers, Huahua Wang, Arindam Banerjee, University of Minnesota .................................................. 2816

Multitask learning meets tensor factorization: task imputation via convex optimization, Kishan Wimalawarne, Masashi Sugiyama, Tokyo Institute of Technology, and Ryota Tomioka, TTI Chicago ............... 2825

Scalable Inference for Neuronal Connectivity from Calcium Imaging, Alyson Fletcher, Sundeep Rangan 2843

Structure learning of antiferromagnetic Ising models, Guy Bresler, David Gamarnik, andDevavrat Shah, Massachusetts Institute of Technology 2852

The Noisy Power Method: A Meta Algorithm with Applications, Moritz Hardt, IBM Research Almaden, Eric Price, IBM Research 2861

Algorithm selection by rational metareasoning as a model of human strategy selection, Falk Lieder, Dillon Plunkett, and Jessica Hamrick, Stuart Russell, Nicholas Hay, UC Berkeley, Thomas Griffiths, cognitive science 2870

Extremal Mechanisms for Local Differential Privacy, Peter Kairouz, Sewoong Oh, and Pramod Viswanath, University of Illinois at Urbana-Champaign 2879

Global Belief Recursive Neural Networks, Romain Paulus, ISEP, Richard Socher, and Chris Manning, Stanford University 2888

A statistical model for tensor PCA, Emile Richard, Institut Curie, Andrea Montanari, Stanford University 2897

Real-Time Decoding of an Integrate and Fire Encoder, Shreya Saxena, Massachusetts Institute of Technology, Munther Dahleh, Massachusetts Institute of Technology 2906

Delay-Tolerant Algorithms for Asynchronous Distributed Online Learning, Brendan McMahan, Google, Matthew Streeter, Duolingo 2915

On the Number of Linear Regions of Deep Neural Networks, Guido Montufar, Max Planck Institute, Razvan Pascanu, University of Montreal, Yoshua Bengio, U. Montreal 2924

Identifying and attacking the saddle point problem in high-dimensional non-convex optimization, Yann Dauphin, University of Montreal, Razvan Pascanu, Université de Montréal, and Caglar Gulcehre, Kyunghyun Cho, University of Montreal, Surya Ganguli, Stanford University, Yoshua Bengio, U. Montreal 2933

Extracting Latent Structure From Multiple Interacting Neural Populations, Joao Semedo, Carnegie Mellon University, Amin Zandvakili, and Adam Kohn, Albert Einstein College of Medicine, Christian Machens, Champalimaud Centre for the Unknown, Byron Yu, Carnegie Mellon University 2942

Learning with Fredholm Kernels, Qichao Que, Ohio State University, Mikhail Belkin, The Ohio State University, and Yusu Wang, Ohio State University 2951

Hamming Ball Auxiliary Sampling for Factorial Hidden Markov Models, Michalis Titsias, Athens University of Economics and Business, Christopher Yau, University of Oxford 2960

Optimizing Energy Production Using Policy Search and Predictive State Representations, Yuri Grinberg, Doina Precup, McGill Montreal, and Michel Gendreau, CIRRELT 2969
<table>
<thead>
<tr>
<th>Title</th>
<th>Authors</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scaling-up Importance Sampling for Markov Logic Networks,</td>
<td>Deepak Venugopal, University of Texas at Dallas, Vibhav Gogate, UT</td>
<td>2978</td>
</tr>
<tr>
<td>Optimal Neural Codes for Control and Estimation,</td>
<td>Alex Susemihl, Berlin Institute of Technology, Ron Meir, Technion, and</td>
<td>2987</td>
</tr>
<tr>
<td>MANfred Opper, &quot;Computer Science TU Berlin&quot;</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Graph Clustering With Missing Data: Convex Algorithms and Analysis,</td>
<td>Ramya Vinayak, Samet Oymak, and Babak Hassibi, Caltech</td>
<td>2996</td>
</tr>
<tr>
<td>Scale Adaptive Blind Deblurring,</td>
<td>Haichao Zhang, Jianchao Yang, Adobe Research</td>
<td>3005</td>
</tr>
<tr>
<td>Weighted importance sampling for off-policy learning with linear</td>
<td>Ashique Mahmood, University of Alberta, Hado Hasselt, and Richard Sutton, University of</td>
<td></td>
</tr>
<tr>
<td>function approximation,</td>
<td>Alberta</td>
<td></td>
</tr>
<tr>
<td>Information-based learning by agents in unbounded state spaces,</td>
<td>Shariq Mobin, James Arnemann, and Fritz Sommer, UC Berkeley</td>
<td>3023</td>
</tr>
<tr>
<td>Exponential Concentration of a Density Functional Estimator,</td>
<td>Shashank Singh, Barnabas Poczos, Carnegie Mellon University</td>
<td>3032</td>
</tr>
<tr>
<td>Scalable Kernel Methods via Doubly Stochastic Gradients,</td>
<td>Bo Dai, Bo Xie, and Xiao He, Georgia Institute of Technology, Yingyu Liang, Georgia Tech,</td>
<td></td>
</tr>
<tr>
<td>Anant Raj, IIT Kanpur, Maria-Florina Balcan, Carnegie Mellon</td>
<td>University, LE Song, Georgia Tech</td>
<td></td>
</tr>
<tr>
<td>Fast Training of Pose Detectors in the Fourier Domain,</td>
<td>Joao Henriques, Pedro Martins, and Rui Caseiro, Jorge Batista, University of Coimbra</td>
<td>3050</td>
</tr>
<tr>
<td>An Accelerated Proximal Coordinate Gradient Method,</td>
<td>QiHang Lin, University of Iowa, Zhaosong Lu, Simon Fraser University, and Lin</td>
<td></td>
</tr>
<tr>
<td>Communication-Efficient Distributed Dual Coordinate Ascent,</td>
<td>Martin Jaggi, Virginia Smith, UC Berkeley, and Martin Takac, Lehigh University, Jonathan</td>
<td></td>
</tr>
<tr>
<td>Simple MAP Inference via Low-Rank Relaxations,</td>
<td>Terhorst, Sanjay Krishnan, UC Berkeley, Thomas Hofmann, Michael Jordan, UC Berkeley</td>
<td></td>
</tr>
<tr>
<td>Improved Distributed Principal Component Analysis,</td>
<td>Yingyu Liang, Georgia Tech, Maria-Florina Balcan, Carnegie Mellon University, and Vandana</td>
<td></td>
</tr>
<tr>
<td>Improved Sequence to Sequence Learning with Neural Networks,</td>
<td>Kanchanapally, Georgia Institute of Technology, David Woodruff, IBM Almaden</td>
<td></td>
</tr>
</tbody>
</table>

xxv
<table>
<thead>
<tr>
<th>Title</th>
<th>Authors</th>
<th>Institution(s)</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sparse Polynomial Learning and Graph Sketching</td>
<td>Murat Kocaoglu, UT Austin, Karthikeyan Shanmugam, and Alexandros Dimakis</td>
<td>The University of Texas at Austin, Adam Klivans, UT Austin</td>
<td>3122</td>
</tr>
<tr>
<td>Tight Continuous Relaxation of the Balanced k-Cut Problem</td>
<td>Syama Sundar Rangapuram, Pramod Mudrakarta, and Matthias Hein</td>
<td>Saarland University</td>
<td>3131</td>
</tr>
<tr>
<td>Mondrian Forests: Efficient Online Random Forests</td>
<td>Balaji Lakshminarayanan, Gatsby Unit UCL, Daniel Roy, University of Cambridge, and Yee Whye Teh</td>
<td>University of Oxford</td>
<td>3140</td>
</tr>
<tr>
<td>Expectation-Maximization for Learning Determinantal Point Processes</td>
<td>Jennifer Gillenwater, University of Pennsylvania, Alex Kulesza, University of Michigan, and Emily Fox, Ben Taskar</td>
<td>University of Washington</td>
<td>3149</td>
</tr>
<tr>
<td>Capturing Semantically Meaningful Word Dependencies with an Admixture of Poisson MRFs</td>
<td>Prem Gopalan, Laurent Charlin, Princeton University, and David Blei</td>
<td>Princeton University, and David Blei</td>
<td>3158</td>
</tr>
<tr>
<td>Streaming, Memory Limited Algorithms for Community Detection</td>
<td>Se-Young Yun, INRIA, Marc Lelarge, and Alexandre Proutiere</td>
<td>Saarland University</td>
<td>3167</td>
</tr>
<tr>
<td>Content-based recommendations with Poisson factorization</td>
<td>Calvin McCarter, Seyoung Kim, Carnegie Mellon University</td>
<td>Carnegie Mellon University</td>
<td>3176</td>
</tr>
<tr>
<td>Bayesian Sampling Using Stochastic Gradient Thermostats</td>
<td>Nan Ding, Google, Youhan Fang, Purdue University, and Ryan Babbush, Harvard University, Changyou Chen, Australian National University, Robert Skeel, Purdue University, Hartmut Neven, Google</td>
<td>Duke University, and Robert Skeel, Purdue University</td>
<td>3185</td>
</tr>
<tr>
<td>Compressive Sensing of Signals from a GMM with Sparse Precision Matrices</td>
<td>Jianbo Yang, Xuejun Liao, Duke University, and Minhua Chen, Lawrence Carin</td>
<td>Duke University</td>
<td>3194</td>
</tr>
<tr>
<td>A Statistical Decision-Theoretic Framework for Social Choice</td>
<td>Hossein Azari Soufiani, Harvard University, David Parkes, and Lirong Xia</td>
<td>Harvard University</td>
<td>3203</td>
</tr>
<tr>
<td>Efficient Minimax Strategies for Square Loss Games</td>
<td>Wouter Koolen, Queensland University of Technology, Alan Malek, and Peter Bartlett</td>
<td>UC Berkeley</td>
<td>3212</td>
</tr>
<tr>
<td>A Residual Bootstrap for High-Dimensional Regression with Near Low-Rank Designs</td>
<td>Miles Lopes, UC Berkeley</td>
<td>UC Berkeley</td>
<td>3221</td>
</tr>
<tr>
<td>Large-Margin Convex Polytope Machine</td>
<td>Alex Kantchelian, Michael Tschantz, UC Berkeley, and Ling Huang, DataVisor, Peter Bartlett, Anthony Joseph, J. Tygar</td>
<td>UC Berkeley</td>
<td>3230</td>
</tr>
</tbody>
</table>
Distributed Variational Inference in Sparse Gaussian Process Regression and Latent Variable Models, Yarin Gal, Mark Wilk, and Carl Rasmussen, University of Cambridge ........................................... 3257

Learning Distributed Representations for Structured Output Prediction, Vivek Srikumar, Christopher Manning, Stanford University ............................................. 3266

Convex Deep Learning via Normalized Kernels, Özlem Aslan, Xinhua Zhang, NICTA and ANU, and Dale Schuurmans, University of Alberta .................................................. 3275

Tight convex relaxations for sparse matrix factorization, Emile Richard, Institut Curie, Guillaume Obozinski, optimization, and Jean-Philippe Vert, Mines ParisTech ........................................... 3284

Learning to Search in Branch and Bound Algorithms, He He, University of Maryland, Hal Daume III, University of Maryland College Park, and Jason Eisner, Johns Hopkins University ........................................ 3293

An Integer Polynomial Programming Based Framework for Lifted MAP Inference, Somdeb Sarkhel, Deepak Venugopal, University of Texas at Dallas, and Parag Singla, IIT Delhi, Vibhav Gogate, UT Dallas ................................. 3302


How transferable are features in deep neural networks?, Jason Yosinski, Cornell University, Jeff Clune, and Yoshua Bengio, U. Montreal, Hod Lipson, Cornell University ........................................ 3320

Accelerated Mini-batch Randomized Block Coordinate Descent Method, Tuo Zhao, Mo Yu, Harbin Institute of Technology, and Yiming Wang, Raman Arora, Johns Hopkins University, Han Liu, supervised learning .................................................. 3329


A Latent Source Model for Online Collaborative Filtering, Guy Bresler, George Chen, and Devavrat Shah, Massachusetts Institute of Technology .................................................. 3347

Tighten after Relax: Minimax-Optimal Sparse PCA in Polynomial Time, Zhaoran Wang, Han Liu, supervised learning .................................................. 3383
Deep Networks with Internal Selective Attention through Feedback Connections, MARIN STOLLENGA, IDSIA/SUPSI USI, JONATHAN MASI, and FAUSTINO GOMEZ, JÜRGEN SCHMIDHUBER ............... 3545

Parallel Feature Selection Inspired by Group Testing, YINGBO ZHOU, SUNY at Buffalo, UTKARSH PORWAL, SUNY Buffalo, and CE ZHANG, University of Wisconsin-Madison, HUNG NGO, University at Buffalo SUNY, XUANLONG NGUYEN, University of Michigan, CHRISTOPHER RE, Stanford University, VENU GOVINDARAJU, SUNY Buffalo ........................................ 3554

Low-Rank Time-Frequency Synthesis, CÉDRIC FÉVOTTE, Centre national de la recherche scientifique, MATTHIEU KOWALSKI, Univ Paris-Sud ... 3563

Pre-training of Recurrent Neural Networks via Linear Autoencoders, LUCA PASA, ALESSANDRO SPERDUTI, Università degli Studi di Padova ......................................................... 3572

Semi-supervised Learning with Deep Generative Models, DIEDERIK KINGMA, University of Amsterdam, SHAKIR MOHAMED, and DANilo REZENDE, Google DeepMind, MAX WELLING .......................... 3581

Signal Aggregate Constraints in Additive Factorial HMMs, with Application to Energy Disaggregation, MINGJUN ZHONG, University of Edinburgh, NIGEL GODDARD, and CHARLES SUTTON, University of Edinburgh ................................................................. 3590

Stochastic variational inference for hidden Markov models, NICK FOTI, JASON XU, and DILLON LAIRD, EMILY FOX, University of Washington 3599

A Wild Bootstrap for Degenerate Kernel Tests, KACPER CHWALKOWSKI, University College London, DINO SEJDI NOVIC, Gatsby Unit UCL, and ARTHUR GRETTON, University College London ............ 3608

Biclustering Using Message Passing, LUKE O’CONNOR, Harvard University, SOHEIL FEIZI, Massachusetts Institute of Technology .............. 3617

Fast Kernel Learning for Multidimensional Pattern Extrapolation, ANDREW WILSON, Carnegie Mellon University, ELAD GILBOA, WUSTL, and JOHN CUNNINGHAM, Columbia University, ARYE NEHORAI, Washington University in St. Louis ................................................................. 3626

Learning on graphs using Orthonormal Representation is Statistically Consistent, RAKESH SHIVANNA, CHIRANJIB BHATTACHARYYA, Indian Institute of Science ........................................ 3635

Spectral k-Support Norm Regularization, ANDREW MCDONALD, MASSIMILIANO PONTIL, and DIMITRIS STAMOS, University College London . 3644

Unsupervised learning of an efficient short-term memory network, PIETRO VERTECHI, Champalimaud Neuroscience Prog, WIELAND BRENDEL, Champalimaud Neuroscience Programme, and CHRISTIAN MACHENS, Champalimaud Centre for the Unknown ........................................... 3653

Quantized Estimation of Gaussian Sequence Models in Euclidean Balls, YUANCHENG ZHU, JOHN LAFFERTY, University of Chicago .......... 3662

Learning a Concept Hierarchy from Multi-labeled Documents, VIET-AN NGUYEN, University of Maryland, JORDAN BOYD-GRABER,
Variational Gaussian Process State-Space Models, Roger Frigola, University of Cambridge, Yutian Chen, Cambridge University, and Carl Rasmussen, University of Cambridge

Fast Prediction for Large-Scale Kernel Machines, Cho-Jui Hsieh, University of Texas at Austin, Si Si, and Inderjit Dhillon, University of Texas

Author Index