

Neural Information Processing Systems

Advances in Neural Information Processing Systems 27

28th Annual Conference on Neural Information
Processing Systems 2014

December 8-13, 2014
Montreal, Canada

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: 978-1-5108-0041-0

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© (2014) by Neural Information Processing Systems
All rights reserved.

Printed by Curran Associates, Inc. (2015)

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	xxxix
Donors	xxxv
NIPS foundation	xxxvi
Committees	xxxvii
Reviewers	xl
Kernel Mean Estimation via Spectral Filtering , KRIKAMOL MUANDET, MPI for Intelligent Systems, BHARATH SRIPERUMBUDUR, supervised learning, and BERNHARD SCHOELKOPF, Max Planck Institute for Biological Cybernetics	1
Semi-Separable Hamiltonian Monte Carlo for Inference in Bayesian Hierarchical Models , YICHUAN ZHANG, CHARLES SUTTON, University of Edinburgh	10
Communication Efficient Distributed Machine Learning with the Parameter Server , MU LI, DAVID ANDERSEN, Carnegie Mellon University, and ALEX SMOLA, Yahoo! Research, KAI YU, Baidu IDL	19
The Infinite Mixture of Infinite Gaussian Mixtures , HALID YEREBAKAN, IUPUI, BARTEK RAJWA, Purdue University, and MURAT DUNDAR, IUPUI	28
Robust Classification Under Sample Selection Bias , ANQI LIU, Univ of Illinois at Chicago, BRIAN ZIEBART, University of Illinois Chicago	37
Zeta Hull Pursuits: Learning Nonconvex Data Hulls , YUANJUN XIONG, The Chinese University of HK, WEI LIU, IBM T. J. Watson Research Center, and DELI ZHAO, Advanced Algorithm Research Group HTC Beijing, XIAOOU TANG, Chinese University of Hong Kong	46
Grouping-Based Low-Rank Trajectory Completion and 3D Reconstruction , AIKATERINI FRAGKIADAKI, University of California Berk, MARTA SALAS, and PABLO ARBELAEZ, JITENDRA MALIK, UC Berkeley	55
Sparse Space-Time Deconvolution for Calcium Image Analysis , FERRAN DIEGO ANDILLA, HCI IWR Uni Heidelberg, FRED HAMPRECHT	64
Restricted Boltzmann machines modeling human choice , TAKAYUKI OSOGAMI, MAKOTO OTSUKA, IBM Research - Tokyo	73
Multiscale Fields of Patterns , PEDRO FELZENSZWALB, JOHN OBERLIN, Brown University	82

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 TUO 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 RAVANBAKSH, 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 BOTTONNE, 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, ZHUJIN 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 Inst	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 MICHAL VALKO, REMI MUNOS, INRIA Lille - Nord Europe	613
Repeated Contextual Auctions with Strategic Buyers, KAREEM AMIN, University of Pennsylvania, AFSHIN ROSTAMIZADEH, Google Research, and UMAR 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
Optimization Methods for Sparse Pseudo-Likelihood Graphical Model Selection, SANG-YUN OH, Lawrence Berkeley National Lab, ONKAR DALAL, and KSHITIJ KHARE, University of Florida, BALA RAJARATNAM, Stanford University	667
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, KFIR 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, DEEPARNAB CHAKRABARTY, PRATEEK JAIN, Microsoft Research, and PRAVESH KOTHARI, UT Austin	802
Exploiting easy data in online optimization, AMIR SANI, INRIA, GERGELY 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, VIKTORIIA 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 SCAMAN, 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, HONGYU 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, XIAOJIAN 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 .	891
SerialRank: Spectral Ranking using Seriation, FAJWEL FOGEL, ALEXANDRE D’ASPREMONT, CNRS - ENS, and MILAN VOJNOVIC, 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 LINDENBAOUM, SHAUL MARKOVITCH, Technion	999
Constant Nullspace Strong Convexity and Fast Convergence of Proximal Methods under High-Dimensional Settings , EN-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 SZÖRÉNYI, 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, Univeristy 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

Learning Multiple Tasks in Parallel with a Shared Annotator, HAIM COHEN, KOBY CRAMMER, Technion	1170
Reducing the Rank in Relational Factorization Models by Including Observable Patterns, MAXIMILIAN NICKEL, Massachusetts Institute of Technology, XUEYAN JIANG, LMU Munich, and VOLKER TRESP, Siemens AG	1179
Clustering from Labels and Time-Varying Graphs, SHIAU HONG LIM, National University of Singapore, YUDONG CHEN, University of California Berkeley, and HUAN XU, National University of Singapore	1188
From Stochastic Mixability to Fast Rates, NISHANT MEHTA, Australian National University, ROBERT WILLIAMSON, NICTA	1197
Recovery of Coherent Data via Low-Rank Dictionary Pursuit, GUANGCAN LIU, Rutgers, PING LI, Rutgers University	1206
Inferring sparse representations of continuous signals with continuous orthogonal matching pursuit, KARIN KNUDSON, JACOB YATES, and ALEXANDER HUK, University of Texas at Austin, JONATHAN PILLOW, Princeton University	1215
Analysis of Variational Bayesian Latent Dirichlet Allocation: Weaker Sparsity Than MAP, SHINICHI NAKAJIMA, Nikon, ISSEI SATO, University of Tokyo, and MASASHI SUGIYAMA, Tokyo Institute of Technology, KAZUHO WATANABE, Toyohashi Tech, HIROKO KOBAYASHI, Nikon Corporation	1224
Discovering, Learning and Exploiting Relevance, CEM TEKIN, MIHAELA SCHAAR, University of California Los Angeles	1233
Divide-and-Conquer Learning by Anchoring a Conical Hull, TIANYI ZHOU, JEFFREY BILMES, and CARLOS GUESTRIN, University of Washington	1242
Extended and Unscented Gaussian Processes, DANIEL STEINBERG, EDWIN BONILLA, NICTA	1251
Spectral Methods meet EM: A Provably Optimal Algorithm for Crowdsourcing, YUCHEN ZHANG, XI CHEN, UC Berkeley, and DENGYONG ZHOU, Microsoft Research, MICHAEL JORDAN, UC Berkeley	1260
Exploiting Linear Structure Within Convolutional Networks for Efficient Evaluation, EMILY DENTON, WOJCIECH ZAREMBA, and JOAN BRUNA, YANN LECUN, ROB FERGUS, New York University	1269
Learning to Discover Efficient Mathematical Identities, WOJCIECH ZAREMBA, New York University, KAROL KURACH, and ROB FERGUS, New York University	1278
The Large Margin Mechanism for Differentially Private Maximization, KAMALIKA CHAUDHURI, University of California San Diego, DANIEL HSU, Columbia University, and SHUANG SONG, UC San Diego	1287
DFacTo: Distributed Factorization of Tensors, JOON CHOI, Purdue University, S. VISHWANATHAN, optimization	1296
Localized Data Fusion for Kernel k-Means Clustering with Application to Cancer Biology, MEHMET GÖNEN, ADAM MARGOLIN, Oregon Health & Science University	1305

Conditional Swap Regret and Conditional Correlated Equilibrium, MEHRYAR MOHRI, Courant Institute and Google, SCOTT YANG, Courant Institute	1314
Mode Estimation for High Dimensional Discrete Tree Graphical Models , CHAO CHEN, Rutgers University, HAN LIU, supervised learning, and DIMITRIS METAXAS, Rutgers University, TIANQI ZHAO, Princeton University	1323
Large-scale L-BFGS using MapReduce, WEIZHU CHEN, ZHENGHAO WANG, and JINGREN ZHOU, Microsoft	1332
Submodular Attribute Selection for Action Recognition in Video, JINGJING ZHENG, University of Maryland, ZHUOLIN JIANG, Noah’s Ark Lab Huawei Technologies, and RAMA CHELLAPPA, University of Maryland College Park, JONATHON PHILLIPS, National Institute of Standards and Technology ...	1341
Efficient Structured Matrix Rank Minimization, ADAMS YU, WANLI MA, and YAOLIANG YU, JAIME CARBONELL, Carnegie Mellon University, SUVRIT SRA, MPI-IS CMU	1350
On Integrated Clustering and Outlier Detection, LIONEL OTT, LINSEY PANG, Sydney University, and FABIO RAMOS, University of Sydney, SANJAY CHAWLA, Sydney University	1359
A Drifting-Games Analysis for Online Learning and Applications to Boosting, HAIPENG LUO, Princeton University, ROBERT SCHAPIRE, Microsoft Research	1368
Projecting Markov Random Field Parameters for Fast Mixing, XIANGHANG LIU, JUSTIN DOMKE, NICTA	1377
Automatic Discovery of Cognitive Skills to Improve the Prediction of Student Learning, ROBERT LINDSEY, MOHAMMAD KHAJAH, University of Colorado, and MICHAEL MOZER, ”U Colorado Boulder”	1386
Near-Optimal-Sample Estimators for Spherical Gaussian Mixtures, ANANDA SURESH, University of California San, ALON ORLITSKY, University of California San Diego, and JAYADEV ACHARYA, Massachusetts Institute of Technology, ASHKAN JAFARPOUR, University of California San Diego	1395
Automated Variational Inference for Gaussian Process Models, TRUNG NGUYEN, ANU & NICTA, EDWIN BONILLA, NICTA	1404
Learning Mixtures of Submodular Functions for Image Collection Summarization, SEBASTIAN TSCHIATSCHKEK, Graz University of Technology, RISHABH IYER, ”University of Washington Seattle”, and HAOCHEN WEI, JEFFREY BILMES, University of Washington	1413
Robust Tensor Decomposition with Gross Corruption, QUANQUAN GU, CS UIUC, HUAN GUI, University of Illinois at Urbana, and JIAWEI HAN, University of Illinois at Urbana-Champaign	1422
Provable Tensor Factorization with Missing Data, PRATEEK JAIN, Microsoft Research, SEWOONG OH, University of Illinois at Urbana-Champaign	1431
Parallel Successive Convex Approximation for Nonsmooth Nonconvex Optimization, MEISAM RAZAVIYAYN, MINGYI HONG, and ZHI-QUAN LUO, University of Minnesota, JONG-SHI PANG, University of Southern California	1440

Using Convolutional Neural Networks to Recognize Rhythm Stimuli from Electroencephalography Recordings , SEBASTIAN STOBER, Western University London ON, DANIEL CAMERON, and JESSICA GRAHN	1449
Blossom Tree Graphical Models , ZHE LIU, JOHN LAFFERTY, University of Chicago	1458
Model-based Reinforcement Learning and the Eluder Dimension , IAN OSBAND, BENJAMIN VAN ROY, Stanford University	1466
Minimax-optimal Inference from Partial Rankings , BRUCE HAJEK, SEWOONG OH, and JIAMING XU, University of Illinois at Urbana-Champaign ..	1475
Spectral Methods for Indian Buffet Process Inference , HSIAO-YU TUNG, Carnegie Mellon University, ALEX SMOLA, Yahoo! Research	1484
On the Statistical Consistency of Plug-in Classifiers for Non-decomposable Performance Measures , HARIKRISHNA NARASIMHAN, Indian Institute of Science Bangalore, ROHIT VAISH, Indian Insititute of Science, and SHIVANI AGARWAL, Indian Institute of Science	1493
Top Rank Optimization in Linear Time , NAN LI, Nanjing University, RONG JIN, unsupervised learning, and ZHI-HUA ZHOU, Nanjing University	1502
Spectral Methods for Supervised Topic Models , YINING WANG, Carnegie Mellon University, JUNE ZHU, Tsinghua University	1511
Graphical Models for Recovering Probabilistic and Causal Queries from Missing Data , KARTHIKA MOHAN, JUDEA PEARL, University of California Los Angeles	1520
Sparse PCA with Oracle Property , QUANQUAN GU, CS UIUC, ZHAORAN WANG, Princeton University, and HAN LIU, supervised learning	1529
Unsupervised Transcription of Piano Music , TAYLOR BERG-KIRKPATRICK, JACOB ANDREAS, and DAN KLEIN, UC Berkeley	1538
Decoupled Variational Gaussian Inference , MOHAMMAD KHAN, École polytechnique fédérale de Lausanne	1547
Estimation with Norm Regularization , ARINDAM BANERJEE, SHENG CHEN, and FARIDEH FAZAYELI, VIDYASHANKAR SIVAKUMAR, University of Minnesota	1556
Decomposing Parameter Estimation Problems , KHALED REFAAT, ARTHUR CHOI, and ADNAN DARWICHE, University of California Los Angeles ..	1565
Stochastic Proximal Gradient Descent with Acceleration Techniques , ATSUSHI NITANDA, NTT DATA Mathematical Systems Inc. ...	1574
Learning to Optimize via Information-Directed Sampling , DAN RUSSO, BENJAMIN VAN ROY, Stanford University	1583
Covariance shrinkage for autocorrelated data , DANIEL BARTZ, KLAUS-ROBERT MÜLLER, TU Berlin	1592
Do Convnets Learn Correspondence? , JONATHAN LONG, University of California - Ber, NING ZHANG, and TREVOR DARRELL, UC Berkeley	1601
The Blinded Bandit: Learning with Adaptive Feedback , OFER DEKEL, Microsoft Research, ELAD HAZAN, Technion, and TOMER KOREN, Technion Israel Institute of Technology	1610

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, Ambiatà, 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 KLOPP, 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, MITESH 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 NÆSSETH, 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

QUIC & DIRTY: A Quadratic Approximation Approach for Dirty Statistical Models, CHO-JUI HSIEH, University of Texas at Austin, INDERJIT DHILLON, University of Texas, and PRADEEP RAVIKUMAR, University of Texas-Austin, STEPHEN BECKER, Paris 6, PEDER OLSEN	2006
General Stochastic Networks for Classification, MATTHIAS ZÖHRER, Graz University of Technology, FRANZ PERNKOPF, Signal Processing and Speech Communication Laboratory Graz Austria	2015
Spatio-temporal Representations of Uncertainty in Spiking Neural Networks, CRISTINA SAVIN, IST Austria, SOPHIE DENÈVE, ENS Paris	2024
Attentional Neural Network: Feature Selection Using Cognitive Feedback, QIAN WANG, Tsinghua University, JIAXING ZHANG, Microsoft Research, and SEN SONG, Tsinghua University, ZHENG ZHANG, NYU Shanghai	2033
Convolutional Neural Network Architectures for Matching Natural Language Sentences, BAOTIAN HU, Harbin Institute of Technology Shenzhen Graduate School, ZHENG DONG LU, "Noah's Ark Lab Huawei Technologies", and HANG LI, Noah's Ark Lab Huawei Technologies Co. Ltd. Sha Tin Hong Kong, QINGCAI CHEN, Harbin Institute of Technology Shenzhen Graduate School	2042
Scalable Non-linear Learning with Adaptive Polynomial Expansions, ALEKH AGARWAL, Microsoft, ALINA BEYGELZIMER, Yahoo, and DANIEL HSU, Columbia University, JOHN LANGFORD, Microsoft Research, MATUS TELGARSKY, UC San Diego	2051
On the relations of LFPs & Neural Spike Trains, DAVID CARLSON, JANA BORG, and KAFUI DZIRASA, LAWRENCE CARIN, Duke University	2060
Diverse Sequential Subset Selection for Supervised Video Summarization, BOQING GONG, WEI-LUN CHAO, University of Southern California, and KRISTEN GRAUMAN, U. Texas, FEI SHA, University of Southern California	2069
Self-Paced Learning with Diversity, LU JIANG, Carnegie Mellon University, DEYU MENG, Xi'an Jiaotong University, and SHOU-I YU, ZHENZHONG LAN, Carnegie Mellon University, SHIGUANG SHAN, ALEXANDER HAUPTMANN, Carnegie Mellon University	2078
Feature Cross-Substitution in Adversarial Classification, BO LI, YEVGENIY VOROBAYCHIK, Vanderbilt University	2087
Deep Recursive Neural Networks for Compositionality in Language, OZAN IRSOY, CLAIRE CARDIE, Cornell University	2096
Inference by Learning: Speeding-up Graphical Model Optimization via a Coarse-to-Fine Cascade of Pruning Classifiers, BRUNO CONEJO, Caltech, NIKOS KOMODAKIS, Ecole des Ponts ParisTech, and SEBASTIEN LEPRINCE, JEAN AVOUAC, Caltech	2105
A Filtering Approach to Stochastic Variational Inference, NEIL HOULSBY, University of Cambridge, DAVID BLEI	2114
Optimizing F-Measures by Cost-Sensitive Classification, SHAMEEM PARAMBATH, UTC, NICOLAS USUNIER, University Technologique de Compigne, and YVES GRANDVALET, Université de Technologie de Compigne	2123
Two-Layer Feature Reduction for Sparse-Group Lasso via Decomposition of Convex Sets, JIE WANG, JIEPING YE, Arizona State University	2132

Improved Multimodal Deep Learning with Variation of Information , KIHYUK SOHN, WENLING SHANG, and HONGLAK LEE, University of Michigan	2141
PEWA: Patch-based Exponentially Weighted Aggregation for image denoising , CHARLES KERVRANN, Inria	2150
Elementary Estimators for Graphical Models , EUNHO YANG, UT Austin, AURELIE LOZANO, IBM Research, and PRADEEP RAVIKUMAR, University of Texas-Austin	2159
Beyond the Birkhoff Polytope: Convex Relaxations for Vector Permutation Problems , CONG LIM, Univ. of Wisconsin-Madison, STEPHEN WRIGHT, University of Wisconsin	2168
Neural Word Embedding as Implicit Matrix Factorization , OMER LEVY, YOAV GOLDBERG, Bar-Ilan University	2177
Multi-Resolution Cascades for Multiclass Object Detection , MOHAMMAD SABERIAN, Yahoo! Labs, NUNO VASCONCELOS, UC San Diego	2186
Median Selection Subset Aggregation for Parallel Inference , XIANGYU WANG, Duke University, PEICHAO PENG, University of Pennsylvania, and DAVID DUNSON, Duke Statistics	2195
Recurrent Models of Visual Attention , VOLODYMYR MNIH, Google Inc., NICOLAS HEESS, and ALEX GRAVES, Google DeepMind, KORAY KAVUKCUOGLU	2204
Tree-structured Gaussian Process Approximations , THANG BUI, University of Cambridge, RICHARD TURNER, neuroscience	2213
Active Learning and Best-Response Dynamics , MARIA-FLORINA BALCAN, Carnegie Mellon University, CHRISTOPHER BERLIND, Georgia Tech, and AVRIM BLUM, Carnegie Mellon University, EMMA COHEN, KAUSHIK PATNAIK, Georgia Institute of Technology, LE SONG, Georgia Tech	2222
Analog Memories in a Balanced Rate-Based Network of E-I Neurons , DYLAN FESTA, GUILLAUME HENNEQUIN, and MATE LENGYEL, University of Cambridge	2231
Fast Sampling-Based Inference in Balanced Neuronal Networks , GUILLAUME HENNEQUIN, University of Cambridge, LAURENCE AITCHISON, Gatsby, and MATE LENGYEL, University of Cambridge	2240
Spectral Learning of Mixture of Hidden Markov Models , CEM SUBAKAN, JOHANNES TRAA, University of Illinois, and PARIS SMARAGDIS, University of Illinois Urbana-Champaign	2249
Subspace Embeddings for the Polynomial Kernel , HAIM AVRON, International Business Machines (IBM), HUY NGUYEN, University of California Berkeley, and DAVID WOODRUFF, IBM Almaden	2258
A Boosting Framework on Grounds of Online Learning , TOFIGH NAGHIBI, BEAT PFISTER, Eidgenössische Technische Hochschule Zürich	2267
A Dual Algorithm for Olfactory Computation in the Locust Brain , SINA TOOTOONIAN, University fo Cambridge, MATE LENGYEL, University of Cambridge	2276
Advances in Learning Bayesian Networks of Bounded Treewidth , SIQI NIE, Rensselaer Polytechnic Institute, DENIS MAUA, USP, and CASSIO CAMPOS, Queen's University Belfast, QIANG JI	2285

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 Aus, 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, PRITISH MOHAPATRA, IIT 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 Unviersity, CHRISTOPHER STOCK, Harvard College, and RYAN ADAMS, Harvard University	2330
Randomized Experimental Design for Causal Graph Discovery, HUINING HU, Mcgill, ZHENTAO LI, Ecole 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 TEPPU NIINIMÄKI, University of Helsinki	2357
Depth Map Prediction from a Single Image using a Multi-Scale Deep Network, DAVID EIGEN, CHRISTIAN PUHRSCHE, 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 BALDI, 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, JIANGWEN SUN, University of Connecticut	2411
Multivariate f-divergence Estimation With Confidence, KEVIN MOON, ALFRED HERO, University of Michigan	2420

Generalized Unsupervised Manifold Alignment , ZHEN CUI, ICT CAS, HONG CHANG, Institute of Computing Technology Chinese Academy of Science, and SHIGUANG SHAN, XILIN CHEN	2429
Smoothed Gradients for Stochastic Variational Inference , STEPHAN MANDT, Princeton University, DAVID BLEI, Princeton	2438
Recursive Context Propagation Network for Semantic Scene Labeling , ABHISHEK SHARMA, University of Maryland, ONCEL TUZEL, and MING-YU LIU, Mitsubishi Electric Research Laboratories	2447
Sparse Random Feature Algorithm as Coordinate Descent in Hilbert Space , EN-HSU YEN, UT-Austin, TING-WEI LIN, and SHOU-DE LIN, National Taiwan University, PRADEEP RAVIKUMAR, University of Texas-Austin, INDERJIT DHILLON, University of Texas	2456
Optimal Teaching for Limited-Capacity Human Learners , KAUSTUBH PATIL, XIAOJIN ZHU, University of Wisconsin-Madison, and LUKASZ KOPEC, University College London, BRADLEY LOVE	2465
Shaping Social Activity by Incentivizing Users , MEHRDAD FARAJTABAR, Georgia Institute of Technolog, NAN DU, Georgia Institute of Technology, and MANUEL RODRIGUEZ, Max Plank Institute for Intelligent Systems, ISABEL VALERA, Universidad Carlos III de Madrid, HONGYUAN ZHA, LE SONG, Georgia Tech	2474
Analysis of Brain States from Multi-Region LFP Time-Series , KYLE ULRICH, DAVID CARLSON, Duke University, and WENZHAO LIAN, Duke university, JANA BORG, KAFUI DZIRASA, LAWRENCE CARIN, Duke University	2483
Reputation-based Worker Filtering in Crowdsourcing , SRIKANTH JAGABATHULA, LAKSHMINARAYANAN SUBRAMANIAN, and ASHWIN VENKATARAMAN, New York University	2492
Multi-Class Deep Boosting , VITALY KUZNETSOV, Courant Institute, MEHRYAR MOHRI, Courant Institute and Google, and UMAR SYED, Google ..	2501
A Differential Equation for Modeling Nesterov’s Accelerated Gradient Method: Theory and Insights , WEIJIE SU, STEPHEN BOYD, and EMMANUEL CANDES, Stanford University	2510
Difference of Convex Functions Programming for Reinforcement Learning , BILAL PIOT, Université Lille 3, MATTHIEU GEIST, SUPELEC, and OLIVIER PIETQUIN, Université de Lille 1	2519
Design Principles of the Hippocampal Cognitive Map , KIMBERLY STACHENFELD, MATTHEW BOTVINICK, Princeton University, and SAMUEL GERSHMAN, Massachusetts Institute of Technology	2528
Deep Symmetry Networks , ROBERT GENS, University of Washington, PEDRO DOMINGOS, U. Washington	2537
Nonparametric Bayesian inference on multivariate exponential families , WILLIAM VEGA-BROWN, MAREK DONIEC, and NICHOLAS ROY, Massachusetts Institute of Technology	2546
Optimal rates for k-NN density and mode estimation , SANJOY DASGUPTA, UC San Diego, SAMORY KPOTUFE, Princeton University ORFE ..	2555
Feedforward Learning of Mixture Models , MATTHEW LAWLOR, STEVEN ZUCKER, Yale University	2564

Diverse Randomized Agents Vote to Win , ALBERT JIANG, Trinity University, LEANDRO MARCOLINO, University of Southern California, and ARIEL PROCACCIA, TUOMAS SANDHOLM, NISARG SHAH, Carnegie Mellon University, MILIND TAMBE, University of Southern California	2573
Ranking via Robust Binary Classification , HYOKUN YUN, Amazon, PARAMESWARAN RAMAN, Purdue University, and S. VISHWANATHAN, optimization	2582
Distributed Balanced Clustering via Mapping Coresets , MOHAMMADHOSSEIN BATENI, ADITYA BHASKARA, and SILVIO LATTANZI, VAHAB MIRROKNI, Google Research NYC	2591
Augur: Data-Parallel Probabilistic Modeling , JEAN-BAPTISTE TRISTAN, Oracle Labs, DANIEL HUANG, Harvard University, and JOSEPH TASSAROTTI, Carnegie Mellon University, ADAM POCOCK, STEPHEN GREEN, GUY STEELE, Oracle Labs	2600
Learning Mixtures of Ranking Models , PRANJAL AWASTHI, Princeton University, AVRIM BLUM, and OR SHEFFET, ARAVINDAN VIJAYARAGHAVAN, Carnegie Mellon University	2609
Controlling privacy in recommender systems , YU XIN, CSAIL MIT, TOMMI JAAKKOLA, Massachusetts Institute of Technology	2618
Convolutional Kernel Networks , JULIEN MAIRAL, INRIA, PIOTR KONIUSZ, and ZAID HARCHAOU, Inria, CORDELIA SCHMID, Inria Alpes	2627
Fairness in Multi-Agent Sequential Decision-Making , CHONGJIE ZHANG, Mass. Institute of Technology, JULIE SHAH, Massachusetts Institute of Technology	2636
Submodular meets Structured: Finding Diverse Subsets in Exponentially-Large Structured Item Sets , ADARSH PRASAD, UT Austin, STEFANIE JEGELKA, UC Berkeley, and DHRUV BATRA, Virginia Tech	2645
Do Deep Nets Really Need to be Deep? , JIMMY BA, University of Toronto, RICH CARUANA, Microsoft Research	2654
Dynamic Rank Factor Model for Text Streams , SHAOBO HAN, LIN DU, and ESTHER SALAZAR, LAWRENCE CARIN, Duke University	2663
Generative Adversarial Nets , IAN GOODFELLOW, University of Montreal, JEAN POUGET-ABADIE, and MEHDI MIRZA, University of Montreal, BING XU, DAVID WARDE-FARLEY, SHERJIL OZAIR, AARON COURVILLE, University of Montreal, YOSHUA BENGIO, U. Montreal	2672
Testing Unfaithful Gaussian Graphical Models , DE SOH, SEKHAR TATIKONDA, Yale University	2681
Global Sensitivity Analysis for MAP Inference in Graphical Models , JASPER BOCK, Ghent University SYSTeMS, CASSIO CAMPOS, Queen’s University Belfast, and ALESSANDRO ANTONUCCI, IDSIA	2690
Deconvolution of High Dimensional Mixtures via Boosting, with Application to Diffusion-Weighted MRI of Human Brain , CHARLES ZHENG, FRANCO PESTILLI, and ARIEL ROKEM, Stanford University	2699
Efficient Minimax Signal Detection on Graphs , JING QIAN, VENKATESH SALIGRAMA, Boston University	2708

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 NATARAJAN, 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 Aus, CONSTANTINE CARAMANIS, UT Austin, and SUJAY SANGHAVI, Univ. of Austin	2753
Deterministic Symmetric Positive Semidefinite Matrix Completion, WILLIAM BISHOP, Carnegie Mellon, BYRON YU, Carnegie Mellon University	2762
Multi-Step Stochastic ADMM in High Dimensions: Applications to Sparse Optimization and Matrix Decomposition, HANIE SEDGHI, University 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 DEREZINSKI, 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
On Model Parallelization and Scheduling Strategies for Distributed Machine Learning, SEUNGHAK LEE, JIN KYU KIM, and XUN ZHENG, Carnegie Mellon University, QIRONG HO, Institute for Infocomm Research ASTAR, GARTH GIBSON, Carnegie Mellon University, ERIC XING, Carnegie-Mellon University	2834

Scalable Inference for Neuronal Connectivity from Calcium Imaging , ALYSON FLETCHER, SUNDEEP RANGAN	2843
Structure learning of antiferromagnetic Ising models , GUY BRESLER, DAVID GAMARNIK, and DEVAVRAT 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 Tec, MUNTHER DAHLEH, Massachusetts Institute of Tehcnology	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, Université de Montréal, and KYUNGHYUN CHO, 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

Scaling-up Importance Sampling for Markov Logic Networks, DEEPAK VENUGOPAL, University of Texas at Dallas, VIBHAV GOGATE, UT Dallas	2978
Optimal Neural Codes for Control and Estimation, ALEX SUSEMIHL, Berlin Institute of Technology, RON MEIR, Technion, and MANFRED OPPER, "Computer Science TU Berlin"	2987
Graph Clustering With Missing Data: Convex Algorithms and Analysis, RAMYA VINAYAK, SAMET OYMAK, and BABAK HASSIBI, Caltech	2996
Scale Adaptive Blind Deblurring, HAICHAO ZHANG, JIANCHAO YANG, Adobe Research	3005
Weighted importance sampling for off-policy learning with linear function approximation, ASHIQUE MAHMOOD, University of Alberta, HADO HASSELT, and RICHARD SUTTON, University of Alberta	3014
Information-based learning by agents in unbounded state spaces, SHARIQ MOBIN, JAMES ARNEMANN, and FRITZ SOMMER, UC Berkeley	3023
Exponential Concentration of a Density Functional Estimator, SHASHANK SINGH, BARNABAS POCZOS, Carnegie Mellon University	3032
Scalable Kernel Methods via Doubly Stochastic Gradients, BO DAI, BO XIE, and NIAO HE, Georgia Institute of Technology, YINGYU LIANG, Georgia Tech, ANANT RAJ, IIT Kanpur, MARIA-FLORINA BALCAN, Carnegie Mellon University, LE SONG, Georgia Tech	3041
Fast Training of Pose Detectors in the Fourier Domain, JOÃO HENRIQUES, PEDRO MARTINS, and RUI CASEIRO, JORGE BATISTA, University of Coimbra	3050
An Accelerated Proximal Coordinate Gradient Method, QIHANG LIN, University of Iowa, ZHAOSONG LU, Simon Frasar University, and LIN XIAO, MSR Redmond	3059
Communication-Efficient Distributed Dual Coordinate Ascent, MARTIN JAGGI, VIRGINIA SMITH, UC Berkeley, and MARTIN TAKAC, Lehigh University, JONATHAN TERHORST, SANJAY KRISHNAN, UC Berkeley, THOMAS HOFMANN, MICHAEL JORDAN, UC Berkeley	3068
Simple MAP Inference via Low-Rank Relaxations, ROY FROSTIG, SIDA WANG, and PERCY LIANG, CHRISTOPHER MANNING, Stanford University	3077
A* Sampling, CHRIS MADDISON, University of Toronto, DANIEL TARLOW, and TOM MINKA, Microsoft Research	3086
A Bayesian model for identifying hierarchically organised states in neural population activity, PATRICK PUTZKY, FLORIAN FRANZEN, and GIACOMO BASSETTO, JAKOB MACKE, Max Planck Institute for Biological Cybernetics	3095
Sequence to Sequence Learning with Neural Networks, ILYA SUTSKEVER, ORIOL VINYALS, and QUOC LE, Google	3104
Improved Distributed Principal Component Analysis, YINGYU LIANG, Georgia Tech, MARIA-FLORINA BALCAN, Carnegie Mellon University, and VANDANA KANCHANAPALLY, Georgia Institute of Technology, DAVID WOODRUFF, IBM Almaden	3113

Sparse Polynomial Learning and Graph Sketching, MURAT KOCAOGLU, UT Austin, KARTHIKEYAN SHANMUGAM, and ALEXANDROS DIMAKIS, The University of Texas at Austin, ADAM KLIVANS, UT Austin	3122
Tight Continuous Relaxation of the Balanced k-Cut Problem, SYAMA SUNDAR RANGAPURAM, PRAMOD MUDRAKARTA, and MATTHIAS HEIN, Saarland University	3131
Mondrian Forests: Efficient Online Random Forests, BALAJI LAKSHMINARAYANAN, Gatsby Unit UCL, DANIEL ROY, University of Cambridge, and YEE WHYEE TEH, University of Oxford	3140
Expectation-Maximization for Learning Determinantal Point Processes, JENNIFER GILLENWATER, University of Pennsylvania, ALEX KULESZA, University of Michigan, and EMILY FOX, BEN TASKAR, University of Washington	3149
Capturing Semantically Meaningful Word Dependencies with an Admixture of Poisson MRFs, DAVID INOUE, University of Texas at Austin, PRADEEP RAVIKUMAR, University of Texas-Austin, and INDERJIT DHILLON, University of Texas	3158
Streaming, Memory Limited Algorithms for Community Detection, SE-YOUNG YUN, INRIA, MARC LELARGE, and ALEXANDRE PROUTIERE	3167
Content-based recommendations with Poisson factorization, PREM GOPALAN, LAURENT CHARLIN, Princeton University, and DAVID BLEI, Princeton	3176
A Statistical Decision-Theoretic Framework for Social Choice, HOSSEIN AZARI SOUFIANI, Harvard University, DAVID PARKES, and LIRONG XIA, Harvard University	3185
Compressive Sensing of Signals from a GMM with Sparse Precision Matrices, JIANBO YANG, XUEJUN LIAO, Duke University, and MINHUA CHEN, LAWRENCE CARIN, Duke University	3194
Bayesian Sampling Using Stochastic Gradient Thermostats, NAN DING, Google, YOUHAN FANG, Purdue University, and RYAN BABBUSH, Harvard University, CHANGYOU CHEN, Australian National University, ROBERT SKEEL, Purdue University, HARTMUT NEVEN, Google	3203
On Sparse Gaussian Chain Graph Models, CALVIN MCCARTER, SEYOUNG KIM, Carnegie Mellon University	3212
Orbit Regularization, RENATO NEGRINHO, ANDRE MARTINS, Carnegie Mellon University	3221
Efficient Minimax Strategies for Square Loss Games, WOUTER KOOLEN, Queensland University of Technology, ALAN MALEK, and PETER BARTLETT, UC Berkeley	3230
A Residual Bootstrap for High-Dimensional Regression with Near Low-Rank Designs, MILES LOPES, UC Berkeley	3239
Large-Margin Convex Polytope Machine, ALEX KANTCHELIAN, MICHAEL TSCHANTZ, UC Berkeley, and LING HUANG, DataVisor, PETER BARTLETT, ANTHONY JOSEPH, J. TYGAR, UC Berkeley	3248

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
Conditional Random Field Autoencoders for Unsupervised Structured Prediction , WALEED AMMAR, CHRIS DYER, and NOAH SMITH, Carnegie Mellon University	3311
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
Deep Learning for Real-Time Atari Game Play Using Offline Monte-Carlo Tree Search Planning , XIAOXIAO GUO, University of Michigan, SATINDER SINGH, U Michigan, and HONGLAK LEE, RICHARD LEWIS, XIAOSHI WANG, University of Michigan	3338
A Latent Source Model for Online Collaborative Filtering , GUY BRESLER, GEORGE CHEN, and DEVAVRAT SHAH, Massachusetts Institute of Technology	3347
Distributed Bayesian Posterior Sampling via Moment Sharing , MINJIE XU, Tsinghua University, BALAJI LAKSHMINARAYANAN, Gatsby Unit UCL, and YEE WHYIE TEH, University of Oxford, JUNE ZHU, BO ZHANG, Tsinghua University	3356
Learning with Pseudo-Ensembles , PHIL BACHMAN, OUAIS ALSHARIF, McGill University, and DOINA PRECUP, McGill Montreal	3365
Learning Time-Varying Coverage Functions , NAN DU, Georgia Institute of Technology, YINGYU LIANG, Georgia Tech, and MARIA-FLORINA BALCAN, Carnegie Mellon University, LE SONG, Georgia Tech	3374
Tighten after Relax: Minimax-Optimal Sparse PCA in Polynomial Time , ZHAORAN WANG, HUANRAN LU, Princeton University, and HAN LIU, supervised learning	3383

Discriminative Metric Learning by Neighborhood Gerrymandering , SHUBHENDU TRIVEDI, TTI-Chicago, DAVID MCALLESTER, Toyota Tech Institute Chicago, and GREG SHAKHNAROVICH, TTI-Chicago	3392
Finding a sparse vector in a subspace: Linear sparsity using alternating directions , QING QU, JU SUN, and JOHN WRIGHT, Columbia University	3401
Asynchronous Anytime Sequential Monte Carlo , BROOKS PAIGE, FRANK WOOD, and ARNAUD DOUCET, YEE WHYIE TEH, University of Oxford	3410
Discrete Graph Hashing , WEI LIU, IBM T. J. Watson Research Center, CUN MU, Columbia University, and SANJIV KUMAR, Google, SHIH-FU CHANG, IBM T. J. Watson Research Center	3419
Feedback Detection for Live Predictors , STEFAN WAGER, Stanford University, NICK CHAMANDY, and OMKAR MURALIDHARAN, AMIR NAJMI, Google	3428
Rates of Convergence for Nearest Neighbor Classification , KAMALIKA CHAUDHURI, University of California San Diego, SANJOY DASGUPTA, UC San Diego	3437
Consistency of weighted majority votes , DANIEL BEREND, ARYEH KONTOROVITCH, Ben Gurion University	3446
Beta-Negative Binomial Process and Exchangeable Random Partitions for Mixed-Membership Modeling , MINGYUAN ZHOU, University of Texas at Austin	3455
Zero-shot recognition with unreliable attributes , DINESH JAYARAMAN, UT Austin, KRISTEN GRAUMAN, U. Texas	3464
Concavity of reweighted Kikuchi approximation , PO-LING LOH, University of Pennsylvania, ANDRE WIBISONO, "University of California Berkeley"	3473
Online Decision-Making in General Combinatorial Spaces , ARUN RAJKUMAR, SHIVANI AGARWAL, Indian Institute of Science	3482
Fast Multivariate Spatio-temporal Analysis via Low Rank Tensor Learning , MOHAMMAD BAHADORI, University of Southern California, QI QI, Rose, and YAN LIU, University of Southern California	3491
Clustered factor analysis of multineuronal spike data , LARS BUESING, TIMOTHY MACHADO, and JOHN CUNNINGHAM, LIAM PANINSKI, Columbia University	3500
Algorithms for CVaR Optimization in MDPs , YINLAM CHOW, Stanford University, MOHAMMAD GHAVAMZADEH, Adobe Research & INRIA	3509
Factoring Variations in Natural Images with Deep Gaussian Mixture Models , AARON VAN DEN OORD, BENJAMIN SCHRAUWEN, Ugent	3518
Partition-wise Linear Models , HIDEKAZU OIWA, The University of Tokyo, RYOHEI FUJIMAKI, NECLA	3527
LSDA: Large Scale Detection through Adaptation , JUDY HOFFMAN, SERGIO GUADARRAMA, and ERIC TZENG, UC Berkeley,	

RONGHANG HU, Tsinghua University, JEFF DONAHUE, ROSS GIRSHICK, TREVOR DARRELL, UC Berkeley, KATE SAENKO, UMass Lowell	3536
Deep Networks with Internal Selective Attention through Feedback Connections, MARIJN STOLLENGA, IDSIA/SUPSI USI, JONATHAN MASCI, 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 CHWIALKOWSKI, University College London, DINO SEJDINOVIC, 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,	

University of Maryland College Park, and PHILIP RESNIK, University of Maryland, JONATHAN CHANG, Facebook	3671
Variational Gaussian Process State-Space Models, ROGER FRIGOLA, University of Cambridge, YUTIAN CHEN, Cambridge University, and CARL RASMUSSEN, University of Cambridge	3680
Fast Prediction for Large-Scale Kernel Machines, CHO-JUI HSIEH, University of Texas at Austin, SI SI, and INDERJIT DHILLON, University of Texas	3689
Author Index	