

Advances in Neural Information Processing Systems 29

30th Annual Conference on Neural
Information Processing Systems 2016

Barcelona, Spain
5-10 December 2016

Volume 1 of 7

Editors:

Daniel D. Lee
Ulrike von Luxburg
Roman Garnett

Masashi Sugiyama
Isabelle Guyon

ISBN: 978-1-5108-3881-9

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© (2016) by individual authors and NIPS
All rights reserved.

Printed by Curran Associates, Inc. (2017)

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

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

Phone: (858) 453-1623

Fax: (858) 587-0417

info@nips.cc

Additional copies of this publication are available from:

Curran Associates, Inc.
57 Morehouse Lane
Red Hook, NY 12571 USA
Phone: 845-758-0400
Fax: 845-758-2633
Email: curran@proceedings.com
Web: www.proceedings.com

Contents

Contents	iii
Preface	xli
Donors	xlvi
NIPS foundation	xlix
Committees	li
Reviewers	liv
Scan Order in Gibbs Sampling: Models in Which it Matters and Bounds on How Much, BRYAN HE, CHRISTOPHER SA, IOANNIS MITLIAGKAS, CHRISTOPHER RÉ, Stanford University	1
Deep ADMM-Net for Compressive Sensing MRI, YAN YANG, JIAN SUN, HUIBIN LI, ZONGBEN XU, Xi'an Jiaotong University	10
A scaled Bregman theorem with applications, RICHARD NOCK, Data61 and ANU, ADITYA MENON, CHENG ONG, Data61	19
Swapout: Learning an ensemble of deep architectures, SAURABH SINGH, DEREK HOIEM, DAVID FORSYTH, University of Illinois at Urbana-Champaign	28
On Regularizing Rademacher Observation Losses, RICHARD NOCK, Data61 and ANU	37
Without-Replacement Sampling for Stochastic Gradient Methods, OHAD SHAMIR, Weizmann Institute of Science	46
Fast and Provably Good Seedings for k-Means, OLIVIER BACHEM, MARIO LUCIC, HAMED HASSANI, Eidgenössische Technische Hochschule Zürich, ANDREAS KRAUSE	55
Unsupervised Learning for Physical Interaction through Video Prediction, CHELSEA FINN, Google Inc., IAN GOODFELLOW, SERGEY LEVINE, University of Washington	64
High-Rank Matrix Completion and Clustering under Self-Expressive Models, EHSAN ELHAMIFAR	73
Learning a Probabilistic Latent Space of Object Shapes via 3D Generative-Adversarial Modeling, JIAJUN WU, CHENGKAI ZHANG, TIANFAN XUE, WILLIAM FREEMAN, and JOSH TENENBAUM, Massachusetts Institute of Technology	82

Visual Dynamics: Probabilistic Future Frame Synthesis via Cross Convolutional Networks, TIANFAN XUE, JIAJUN WU, KATHERINE BOUMAN, WILLIAM FREEMAN, Massachusetts Institute of Technology	91
Human Decision-Making under Limited Time, PEDRO ORTEGA, ALAN STOCKER, University of Pennsylvania	100
Incremental Boosting Convolutional Neural Network for Facial Action Unit Recognition, SHIZHONG HAN, ZIBO MENG, AHMED KHAN, YAN TONG, University of South Carolina	109
Natural-Parameter Networks: A Class of Probabilistic Neural Networks, HAO WANG, HKUST, XINGJIAN SHI, DIT-YAN YEUNG	118
Tree-Structured Reinforcement Learning for Sequential Object Localization, ZEQUAN JIE, National Univ of Singapore, XIAODAN LIANG, Sun Yat-sen University, JIASHI FENG, National University of Singapore, XIAOJIE JIN, National University of Singapore, and WEN LU, National Univ of Singapore, SHUICHENG YAN, National University of Singapore	127
Unsupervised Domain Adaptation with Residual Transfer Networks, MINGSHENG LONG, HAN ZHU, JIANMIN WANG, Tsinghua University, MICHAEL JORDAN, OpenAI / UC Berkeley / Gradescope	136
Verification Based Solution for Structured MAB Problems, ZOHAR KARNIN	145
Minimizing Regret on Reflexive Banach Spaces and Nash Equilibria in Continuous Zero-Sum Games, MAXIMILIAN BALANDAT, WALID KRICHENE, CLAIRE TOMLIN, ALEXANDRE BAYEN, UC Berkeley	154
Linear dynamical neural population models through nonlinear embeddings, YUANJUN GAO, EVAN ARCHER, LIAM PANINSKI, JOHN CUNNINGHAM, Columbia University	163
SURGE: Surface Regularized Geometry Estimation from a Single Image, PENG WANG, University of California Los Angeles, XIAOHUI SHEN, Adobe Research, BRYAN RUSSELL, SCOTT COHEN, Adobe Research, and BRIAN PRICE, ALAN YUILLE	172
Interpretable Distribution Features with Maximum Testing Power, WITAWAT JITKRITUM, Gatsby Unit UCL, ZOLTAN SZABO, KACPER CHWIALKOWSKI, Gatsby Unit UCL, ARTHUR GRETTON	181
Sorting out typicality with the inverse moment matrix SOS polynomial, EDOUARD PAUWELS, JEAN-BERNARD LASSERRE, LAAS-CNRS	190
Multi-armed Bandits: Competing with Optimal Sequences, ZOHAR KARNIN, OREN ANAVA, Technion	199
Multivariate tests of association based on univariate tests, RUTH HELLER, Tel-Aviv University, YAIR HELLER	208
Learning What and Where to Draw, SCOTT REED, University of Michigan, ZEYNEP AKATA, Max Planck Institute for Informatics, SANTOSH MOHAN, SAMUEL TENKA, University of Michigan, and BERNT SCHIELE, HONGLAK LEE, University of Michigan	217
The Sound of APALM Clapping: Faster Nonsmooth Nonconvex Optimization with Stochastic Asynchronous PALM, DAMEK	

DAVIS, Cornell University, BRENT EDMUNDS, University of California Los Angeles, MADELEINE UDELL	226
Integrated perception with recurrent multi-task neural networks, HAKAN BILEN, University of Oxford, ANDREA VEDALDI	235
Combining Low-Density Separators with CNNs, YU-XIONG WANG, MARTIAL HEBERT, Carnegie Mellon University	244
CNNpack: Packing Convolutional Neural Networks in the Frequency Domain, YUNHE WANG, Peking University, CHANG XU, SHAN YOU, DACHENG TAO, and CHAO XU	253
Cooperative Graphical Models, JOSIP DJOLONGA, Eidgenössische Technische Hochschule Zürich, STEFANIE JEGELKA, Massachusetts Institute of Technology, SEBASTIAN TSCHIATSCHKEK, Eidgenössische Technische Hochschule Zürich, ANDREAS KRAUSE	262
f-GAN: Training Generative Neural Samplers using Variational Divergence Minimization, SEBASTIAN NOWOZIN, BOTOND CSEKE, Microsoft Research, RYOTA TOMIOKA, MSRC	271
Bayesian Optimization for Probabilistic Programs, TOM RAINFORTH, TUAN LE, University of Oxford, JAN-WILLEM MEENT, Northeastern University, MICHAEL OSBORNE, and FRANK WOOD	280
Hierarchical Question-Image Co-Attention for Visual Question Answering, JIASEN LU, JIANWEI YANG, DHURV BATRA, DEVI PARIKH, Virginia Tech	289
Optimal Sparse Linear Encoders and Sparse PCA, MALIK MAGDON-ISMAIL, Rensselaer, CHRISTOS BOUTSIDIS	298
FPNN: Field Probing Neural Networks for 3D Data, YANGYAN LI, SOEREN PIRK, HAO SU, CHARLES QI, and LEONIDAS GUIBAS, Stanford University	307
CRF-CNN: Modeling Structured Information in Human Pose Estimation, XIAO CHU, WANLI OUYANG, HONGSHENG LI, The Chinese University of Hong Kong, XIAOGANG WANG, Chinese University of Hong Kong .	316
Fairness in Learning: Classic and Contextual Bandits, MATTHEW JOSEPH, MICHAEL KEARNS, JAMIE MORGENSTERN, AARON ROTH, University of Pennsylvania	325
Joint M-Best-Diverse Labelings as a Parametric Submodular Minimization, ALEXANDER KIRILLOV, TU Dresden, ALEXANDER SHEKHOVTSOV, CARSTEN ROTHER, BOGDAN SAVCHYNSKY	334
Domain Separation Networks, KONSTANTINOS BOUSMALIS, Google Brain, GEORGE TRIGEORGIS, NATHAN SILBERMAN, DILIP KRISHNAN, and DUMITRU ERHAN, Google	343
DISCO Nets : DISsimilarity COefficients Networks, DIANE BOUCHACOURT, M. KUMAR, University of Oxford, SEBASTIAN NOWOZIN	352
Multimodal Residual Learning for Visual QA, JIN-HWA KIM, SANG-WOO LEE, DONG-HYUN KWAK, MIN-OH HEO, Seoul National University, and JEONGHEE KIM, JUNG-WOO HA, Naver Labs, BYOUNG-TAK ZHANG, Seoul National University	361

CMA-ES with Optimal Covariance Update and Storage Complexity, OSWIN KRAUSE, DÍDAC ARBONÈS, University of Copenhagen, CHRISTIAN IGEL	370
R-FCN: Object Detection via Region-based Fully Convolutional Networks, JIFENG DAI, Microsoft, YI LI, Tsinghua University, KAIMING HE, JIAN SUN, Microsoft	379
GAP Safe Screening Rules for Sparse-Group Lasso, EUGENE NDIAYE, Télécom ParisTech, OLIVIER FERCOQ, ALEXANDRE GRAMFORT, JOSEPH SALMON	388
Learning and Forecasting Opinion Dynamics in Social Networks, ABIR DE, IIT Kharagpur, ISABEL VALERA, NILOY GANGULY, SOURANGSHU BHATTACHARYA, IIT Kharagpur, and MANUEL RODRIGUEZ, MPI-SWS	397
Gradient-based Sampling: An Adaptive Importance Sampling for Least-squares, RONG ZHU, Chinese Academy of Sciences	406
Collaborative Recurrent Autoencoder: Recommend while Learning to Fill in the Blanks, HAO WANG, HKUST, XINGJIAN SHI, DIT-YAN YEUNG	415
Mutual information for symmetric rank-one matrix estimation: A proof of the replica formula, JEAN BARBIER, MOHAMAD DIA, NICOLAS MACRIS, École polytechnique fédérale de Lausanne, FLORENT KRZAKALA, and THIBAUT LESIEUR, IPHT Saclay, LENKA ZDEBOROVA	424
A Unified Approach for Learning the Parameters of Sum-Product Networks, HAN ZHAO, Carnegie Mellon University, PASCAL POUPART, GEOFF GORDON, Carnegie Mellon University	433
Training and Evaluating Multimodal Word Embeddings with Large-scale Web Annotated Images, JUNHUA MAO, University of California Los Angeles, JIAJING XU, Pinterest, KEVIN JING, ALAN YUILLE ...	442
Stochastic Online AUC Maximization, YIMING YING, LONGYIN WEN, SIWEI LYU, State University of New York at Albany	451
The Generalized Reparameterization Gradient, FRANCISCO RUIZ, Columbia University, MICHALIS TITSIAS, DAVID BLEI, Columbia University ...	460
Coupled Generative Adversarial Networks, MING-YU LIU, Mitsubishi Electric Research Laboratories, ONCEL TUZEL, Mitsubishi Electric Research Labs	469
Exponential Family Embeddings, MAJA RUDOLPH, FRANCISCO RUIZ, Columbia University, STEPHAN MANDT, Disney Research, DAVID BLEI, Columbia University	478
Variational Information Maximization for Feature Selection, SHUYANG GAO, University of Southern California, GREG STEEG, ARAM GALSTYAN	487
Operator Variational Inference, RAJESH RANGANATH, JAAN ALTOSAAR, Princeton University, DUSTIN TRAN, DAVID BLEI, Columbia University	496
Fast learning rates with heavy-tailed losses, VU DINH, Fred Hutchinson Cancer Center, LAM HO, University of California Los Angeles, DU NGUYEN, University of Wisconsin-Madison, BINH NGUYEN, University of Science Vietnam	505

Budgeted stream-based active learning via adaptive submodular maximization, KAITO FUJII, HISASHI KASHIMA, Kyoto University	514
Learning feed-forward one-shot learners, LUCA BERTINETTO, JOAO HENRIQUES, JACK VALMADRE, University of Oxford, PHILIP TORR, and ANDREA VEDALDI	523
Learning User Perceived Clusters with Feature-Level Supervision, TING-YU CHENG, KUAN-HUA LIN, XINYANG GONG, Baidu Inc., KANG-JUN LIU, and SHAN-HUNG WU, National Tsing Hua University	532
Robust Spectral Detection of Global Structures in the Data by Learning a Regularization, PAN ZHANG, Institute of Theoretical Physics Chinese Academy of Sciences	541
Residual Networks Behave Like Ensembles of Relatively Shallow Networks, ANDREAS VEIT, MICHAEL WILBER, SERGE BELONGIE, Cornell University	550
Adversarial Multiclass Classification: A Risk Minimization Perspective, RIZAL FATHONY, U. of Illinois at Chicago, ANQI LIU, KAISER ASIF, BRIAN ZIEBART, University Of Illinois at Chicago	559
Solving Random Systems of Quadratic Equations via Truncated Generalized Gradient Flow, GANG WANG, GEORGIOS GIANNAKIS, University of Minnesota	568
Coin Betting and Parameter-Free Online Learning, FRANCESCO ORABONA, Yahoo Research, DAVID PAL	577
Deep Learning without Poor Local Minima, KENJI KAWAGUCHI, Massachusetts Institute of Technology	586
Testing for Differences in Gaussian Graphical Models: Applications to Brain Connectivity, EUGENE BELILOVSKY, CentraleSupélec, GAEL VAROQUAUX, MATTHEW BLASCHKO, KU Leuven	595
A Constant-Factor Bi-Criteria Approximation Guarantee for k-means++, DENNIS WEI, IBM Research	604
Generating Videos with Scene Dynamics, CARL VONDRICK, Massachusetts Institute of Technology, HAMED PIRSIAVASH, ANTONIO TORRALBA, Massachusetts Institute of Technology	613
Neurally-Guided Procedural Models: Amortized Inference for Procedural Graphics Programs using Neural Networks, DANIEL RITCHIE, ANNA THOMAS, PAT HANRAHAN, NOAH GOODMAN, Stanford University	622
A Powerful Generative Model Using Random Weights for the Deep Image Representation, KUN HE, Huazhong University of Science and Technology, YAN WANG, HUAZHONG UNIVERSITY OF SCIENCE, JOHN HOPCROFT, Cornell University	631
Optimizing affinity-based binary hashing using auxiliary coordinates, RAMIN RAZIPERCHIKOLAEI, MIGUEL CARREIRA-PERPINAN, UC Merced	640
Double Thompson Sampling for Dueling Bandits, HUASEN WU, University of California at Davis, XIN LIU, University of California Davis	649

Generating Images with Perceptual Similarity Metrics based on Deep Networks, ALEXEY DOSOVITSKIY, THOMAS BROX, University of Freiburg	658
Dynamic Filter Networks, BERT BRABANDERE, XU JIA, TINNE TUYTELAARS, KU Leuven, LUC GOOL, ETH Zürich	667
A Simple Practical Accelerated Method for Finite Sums, AARON DEFAZIO, Ambiatà	676
Barzilai-Borwein Step Size for Stochastic Gradient Descent, CONGHUI TAN, SHIQIAN MA, The Chinese University of HK, YU-HONG DAI, YUQIU QIAN, The University of Hong Kong	685
On Graph Reconstruction via Empirical Risk Minimization: Fast Learning Rates and Scalability, GUILLAUME PAPA, Télécom ParisTech, AURÉLIEN BELLET, STEPHAN CLÉMENCON	694
Optimal spectral transportation with application to music transcription, RÉMI FLAMARY, CÉDRIC FÉVOTTE, Centre national de la recherche scientifique, NICOLAS COURTY, VALENTIN ÉMIYA, Aix-Marseille University	703
Regularized Nonlinear Acceleration, DAMIEN SCIEUR, INRIA - ENS, ALEXANDRE D’ASPREMONT, FRANCIS BACH	712
SPALS: Fast Alternating Least Squares via Implicit Leverage Scores Sampling, DEHUA CHENG, Univ. of Southern California, RICHARD PENG, IOAKEIM PERROS, Georgia Institute of Technology, YAN LIU, University of Southern California	721
Single-Image Depth Perception in the Wild, WEIFENG CHEN, ZHAO FU, DAWEI YANG, JIA DENG, University of Michigan	730
Computational and Statistical Tradeoffs in Learning to Rank, ASHISH KHETAN, University of Illinois Urbana-, SEWOONG OH, University of Illinois at Urbana-Champaign	739
Online Convex Optimization with Unconstrained Domains and Losses, ASHOK CUTKOSKY, KWABENA BOAHEN, Stanford University	748
An ensemble diversity approach to supervised binary hashing, MIGUEL CARREIRA-PERPINAN, RAMIN RAZIPERCHIKOLAEI, UC Merced	757
Efficient Globally Convergent Stochastic Optimization for Canonical Correlation Analysis, WEIRAN WANG, TTI-Chicago, JIALEI WANG, University of Chicago, DAN GARBER, NATI SREBRO, TTI-Chicago	766
The Power of Adaptivity in Identifying Statistical Alternatives, KEVIN JAMIESON, DANIEL HAAS, BEN RECHT, UC Berkeley	775
On Explore-Then-Commit strategies, AURELIEN GARIVIER, TOR LATTIMORE, EMILIE KAUFMANN	784
Sublinear Time Orthogonal Tensor Decomposition, ZHAO SONG, UT-Austin, DAVID WOODRUFF, HUAN ZHANG, UC-Davis	793
DECORrelated feature space partitioning for distributed sparse regression, XIANGYU WANG, DAVID DUNSON, Duke University, CHENLEI LENG, University of Warwick	802
Deep Alternative Neural Network: Exploring Contexts as Early as Possible for Action Recognition, JINZHUO WANG, PKU, WENMIN	

WANG, XIONGTAO CHEN, RONGGANG WANG, and WEN GAO, peking university	811
Dual Learning for Machine Translation , DI HE, Microsoft, YINGCE XIA, USTC, TAO QIN, Microsoft, LIWEI WANG, and NENGHAI YU, USTC, TIE-YAN LIU, WEI-YING MA, Microsoft	820
Dialog-based Language Learning , JASON WESTON	829
Joint Line Segmentation and Transcription for End-to-End Handwritten Paragraph Recognition , THEODORE BLUCHE, A2iA	838
Temporal Regularized Matrix Factorization for High-dimensional Time Series Prediction , HSIANG-FU YU, University of Texas at Austin, NIKHIL RAO, Technicolor, INDERJIT DHILLON, University of Texas at Austin ..	847
Active Nearest-Neighbor Learning in Metric Spaces , ARYEH KONTOROVITCH, SIVAN SABATO, Ben-Gurion University of the Negev, RUTH URNER, MPI Tuebingen	856
Proximal Deep Structured Models , SHENLONG WANG, SANJA FIDLER, RAQUEL URTASUN, University of Toronto	865
Faster Projection-free Convex Optimization over the Spectrahedron , DAN GARBER, TTI-Chicago	874
Bayesian Optimization with a Finite Budget: An Approximate Dynamic Programming Approach , REMI LAM, KAREN WILLCOX, Massachusetts Institute of Technology, DAVID WOLPERT	883
SoundNet: Learning Sound Representations from Unlabeled Video , YUSUF AYTAR, CARL VONDRICK, ANTONIO TORRALBA, Massachusetts Institute of Technology	892
Weight Normalization: A Simple Reparameterization to Accelerate Training of Deep Neural Networks , TIM SALIMANS, DIEDERIK KINGMA	901
Efficient Second Order Online Learning by Sketching , HAIPENG LUO, Princeton University, ALEKH AGARWAL, Microsoft, NICOLÒ CESA-BIANCHI, JOHN LANGFORD	910
Dynamic Mode Decomposition with Reproducing Kernels for Koopman Spectral Analysis , YOSHINOBU KAWAHARA, Osaka University	919
Distributed Flexible Nonlinear Tensor Factorization , SHANDIAN ZHE, Purdue University, KAI ZHANG, NEC Labs America Princeton NJ 08540, PENGYUAN WANG, KUANG-CHIH LEE, and ZENGLIN XU, YUAN QI, ZOUBIN GHAHRAMANI	928
The Robustness of Estimator Composition , PINGFAN TANG, JEFF PHILLIPS, University of Utah	937
Efficient and Robust Spiking Neural Circuit for Navigation Inspired by Echolocating Bats , PULKIT TANDON, YASH MALVIYA, IIT Bombay, BIPIN RAJENDRAN, NJIT	946
PerforatedCNNs: Acceleration through Elimination of Redundant Convolutions , MICHAEL FIGURNOV, Skolkovo Inst. of Sc and Tech, AIJAN IBRAIMOVA, Skolkovo Institute of Science and Technology, DMITRY VETROV, PUSHMEET KOHLI	955

Differential Privacy without Sensitivity , KENTARO MINAMI, HITOMI ARAI, ISSEI SATO, The University of Tokyo, HIROSHI NAKAGAWA	964
Optimal Cluster Recovery in the Labeled Stochastic Block Model , SE-YOUNG YUN, Los Alamos National Laboratory, ALEXANDRE PROUTIERE ..	973
Even Faster SVD Decomposition Yet Without Agonizing Pain , ZEYUAN ALLEN-ZHU, YUANZHI LI, Princeton University	982
An algorithm for L1 nearest neighbor search via monotonic embedding , XINAN WANG, University of California San Diego, SANJOY DASGUPTA, University of California, San Diego	991
Gaussian Process Bandit Optimisation with Multi-fidelity Evaluations , KIRTHEVASAN KANDASAMY, GAUTAM DASARATHY, JUNIER OLIVA, JEFF SCHNEIDER, and BARNABAS PO CZOS, Carnegie Mellon University	1000
Linear-Memory and Decomposition-Invariant Linearly Convergent Conditional Gradient Algorithm for Structured Polytopes , DAN GARBER, OFER MESH I, TTI-Chicago	1009
Efficient Nonparametric Smoothness Estimation , SHASHANK SINGH, SIMON DU, BARNABAS PO CZOS, Carnegie Mellon University	1018
A Theoretically Grounded Application of Dropout in Recurrent Neural Networks , YARIN GAL, University of Cambridge, ZOUBIN GHAHRAMANI	1027
Fast ϵ-free Inference of Simulation Models with Bayesian Conditional Density Estimation , GEORGE PAPAMAKARIOS, IAIN MURRAY, University of Edinburgh	1036
Direct Feedback Alignment Provides Learning in Deep Neural Networks , ARILD N ØKLAND, None	1045
Safe and Efficient Off-Policy Reinforcement Learning , REMI MUNOS, THOMAS STEPLETON, Google DeepMind, ANNA HARUTYUNYAN, Vrije Universiteit Brussel, MARC BELLEMARE, Google DeepMind	1054
A Multi-Batch L-BFGS Method for Machine Learning , ALBERT BERAHAS, JORGE NOCEDAL, Northwestern University, MARTIN TAKAC, Lehigh University	1063
Semiparametric Differential Graph Models , PAN XU, QUANQUAN GU, University of Virginia	1072
Rényi Divergence Variational Inference , YINGZHEN LI, University of Cambridge, RICHARD TURNER	1081
Doubly Convolutional Neural Networks , SHUANGFEI ZHAI, Binghamton University, YU CHENG, IBM Research, WEINING LU, Tsinghua University, ZHONGFEI ZHANG, Binghamton University	1090
Density Estimation via Discrepancy Based Adaptive Sequential Partition , DANGNA LI, Stanford university, KUN YANG, Google Inc, WING WONG, Stanford university	1099
How Deep is the Feature Analysis underlying Rapid Visual Categorization? , SVEN EBERHARDT, JONAH CADER, THOMAS SERRE, Brown University	1108

VIME: Variational Information Maximizing Exploration , REIN HOUTHOOFT, Ghent University - iMinds and UC Berkeley and OpenAI, XI CHEN, UC Berkeley and OpenAI, YAN DUAN, UC Berkeley, JOHN SCHULMAN, OpenAI, and FILIP TURCK, Ghent University - iMinds, PIETER ABBEEL, OpenAI / UC Berkeley / Gradescope	1117
Generalized Correspondence-LDA Models (GC-LDA) for Identifying Functional Regions in the Brain , TIMOTHY RUBIN, Indiana University, SANMI KOYEJO, University of Illinois at Urbana-Champaign, MICHAEL JONES, Indiana University, TAL YARKONI, University of Texas at Austin	1126
Solving Marginal MAP Problems with NP Oracles and Parity Constraints , YEXIANG XUE, Cornell University, ZHIYUAN LI, Tsinghua University, STEFANO ERMON, Stanford University, CARLA GOMES, and BART SELMAN, Cornell University	1135
Multi-view Anomaly Detection via Robust Probabilistic Latent Variable Models , TOMOHARU IWATA, NTT, MAKOTO YAMADA	1144
Proximal Stochastic Methods for Nonsmooth Nonconvex Finite-Sum Optimization , SASHANK REDDI, Carnegie Mellon University, SUVRIT SRA, Massachusetts Institute of Technology, BARNABAS POZOS, Carnegie Mellon University, ALEX SMOLA	1153
Variance Reduction in Stochastic Gradient Langevin Dynamics , KUMAR DUBEY, SASHANK REDDI, Carnegie Mellon University, SINEAD WILLIAMSON, BARNABAS POZOS, Carnegie Mellon University, and ALEX SMOLA, ERIC XING, Carnegie Mellon University	1162
Regularization With Stochastic Transformations and Perturbations for Deep Semi-Supervised Learning , MEHDI SAJJADI, MEHRAN JAVANMARDI, TOLGA TASDIZEN, University of Utah	1171
Dense Associative Memory for Pattern Recognition , DMITRY KROTOV, Institute for Advanced Study, JOHN HOPFIELD, Princeton Neuroscience Institute	1180
Causal Bandits: Learning Good Interventions via Causal Inference , FINNIAN LATTIMORE, Australian National University, TOR LATTIMORE, MARK REID, Data61	1189
Refined Lower Bounds for Adversarial Bandits , SÉBASTIEN GERCHINOVITZ, TOR LATTIMORE	1198
Theoretical Comparisons of Positive-Unlabeled Learning against Positive-Negative Learning , GANG NIU, University of Tokyo, MARTHINUS PLESSIS, TOMOYA SAKAI, YAO MA, and MASASHI SUGIYAMA, RIKEN / University of Tokyo	1207
Homotopy Smoothing for Non-Smooth Problems with Lower Complexity than $O(1/\epsilon)$, YI XU, The University of Iowa, YAN YAN, University of Technology Sydney, QIHANG LIN, TIANBAO YANG, University of Iowa	1216
Finite-Sample Analysis of Fixed-k Nearest Neighbor Density Functional Estimators , SHASHANK SINGH, BARNABAS POZOS, Carnegie Mellon University	1225
A state-space model of cross-region dynamic connectivity in MEG/EEG , YING YANG, ELISSA AMINOFF, MICHAEL TARR, ROBERT KASS, Carnegie Mellon University	1234

What Makes Objects Similar: A Unified Multi-Metric Learning Approach, HAN-JIA YE, Nanjing University, DE-CHUAN ZHAN, XUE-MIN SI, YUAN JIANG, Nanjing University, and ZHI-HUA ZHOU	1243
Adaptive Maximization of Pointwise Submodular Functions With Budget Constraint, NGUYEN CUONG, University of Cambridge, HUAN XU, National University of Singapore	1252
Dueling Bandits: Beyond Condorcet Winners to General Tournament Solutions, SIDDARTHA RAMAMOCHAN, Indian Institute of Science, ARUN RAJKUMAR, SHIVANI AGARWAL, University of Pennsylvania ..	1261
Local Similarity-Aware Deep Feature Embedding, CHEN HUANG, Chinese University of HongKong, CHEN LOY, The Chinese University of HK, XIAOOU TANG, The Chinese University of Hong Kong	1270
A Communication-Efficient Parallel Algorithm for Decision Tree, QI MENG, Peking University, GUOLIN KE, TAIFENG WANG, WEI CHEN, and QIWEI YE, Microsoft Research, ZHI-MING MA, Academy of Mathematics and Systems Science Chinese Academy of Sciences, TIE-YAN LIU, Microsoft Research	1279
Convex Two-Layer Modeling with Latent Structure, VIGNESH GANAPATHIRAMAN, University Of Illinois at Chicago, XINHUA ZHANG, University of Illinois Chicago, YAOLIANG YU, Carnegie Mellon University, JUNFENG WEN, University of Alberta	1288
Sampling for Bayesian Program Learning, KEVIN ELLIS, ARMANDO SOLAR-LEZAMA, JOSH TENENBAUM, Massachusetts Institute of Technology ..	1297
Learning Kernels with Random Features, AMAN SINHA, JOHN DUCHI, Stanford University	1306
Optimal Tagging with Markov Chain Optimization, NIR ROSENFELD, Hebrew University of Jerusalem, AMIR GLOBERSON, Tel Aviv University	1315
Crowdsourced Clustering: Querying Edges vs Triangles, RAMYA VINAYAK, HASSIBI BABAK, Caltech	1324
Mixed vine copulas as joint models of spike counts and local field potentials, ARNO ONKEN, STEFANO PANZERI, IIT	1333
Achieving the KS threshold in the general stochastic block model with linearized acyclic belief propagation, EMMANUEL ABBE, COLIN SANDON, Princeton University	1342
Adaptive Concentration Inequalities for Sequential Decision Problems, SHENGJIA ZHAO, ENZE ZHOU, Tsinghua University, ASHISH SABHARWAL, Allen Institute for AI, STEFANO ERMON, Stanford University ...	1351
Nested Mini-Batch K-Means, JAMES NEWLING, FRANÇOIS FLEURET, Idiap Research Institute	1360
Deep Learning Models of the Retinal Response to Natural Scenes, LANE MCINTOSH, NIRU MAHESWARANATHAN, ARAN NAYEBI, SURYA GANGULI, and STEPHEN BACCUS, Stanford University	1369
Preference Completion from Partial Rankings, SURIYA GUNASEKAR, UT Austin, SANMI KOYEJO, University of Illinois at Urbana-Champaign, JOYDEEP GHOSH, UT Austin	1378

Dynamic Network Surgery for Efficient DNNs , YIWEN GUO, Intel Labs China, ANBANG YAO, YURONG CHEN	1387
Learning a Metric Embedding for Face Recognition using the Multibatch Method , OREN TADMOR, YONATAN WEXLER, OrCam, TAL ROSENWEIN, OrCam, SHAI SHALEV-SHWARTZ, and AMNON SHASHUA, OrCam	1396
A Pseudo-Bayesian Algorithm for Robust PCA , TAE-HYUN OH, KAIST, YASUYUKI MATSUSHITA, Osaka University, IN KWEON, KAIST, DAVID WIPF	1398
End-to-End Kernel Learning with Supervised Convolutional Kernel Networks , JULIEN MAIRAL, Inria	1407
Stochastic Variance Reduction Methods for Saddle-Point Problems , P. BALAMURUGAN, FRANCIS BACH	1416
Flexible Models for Microclustering with Application to Entity Resolution , GIACOMO ZANELLA, The University of Warwick, BRENDA BETANCOURT, Duke University, HANNA WALLACH, Microsoft Research, JEFFREY MILLER, and ABBAS ZAIDI, REBECCA STEORTS, Duke University ..	1425
Catching heuristics are optimal control policies , BORIS BELOUSOV, TU Darmstadt, GERHARD NEUMANN, CONSTANTIN ROTHKOPF, JANUARY PETERS	1434
Bayesian optimization under mixed constraints with a slack-variable augmented Lagrangian , VICTOR PICHENY, Institut National de la Recherche Agronomique, ROBERT GRAMACY, Virginia Tech, STEFAN WILD, Arbonne National Lab, SEBASTIEN DIGABEL, École Polytechnique de Montréal ..	1443
Adaptive Neural Compilation , RUDY BUNEL, Oxford University, ALBAN DESMAISON, University of Oxford, PUSHMEET KOHLI, PHILIP TORR, and M. KUMAR, University of Oxford	1452
Synthesis of MCMC and Belief Propagation , SUNG-SOO AHN, KAIST, MISHA CHERTKOV, Los Alamos National Laboratory, JINWOO SHIN, KAIST	1461
Learning Treewidth-Bounded Bayesian Networks with Thousands of Variables , MAURO SCANAGATTA, GIORGIO CORANI, Idsia, CASSIO CAMPOS, Queen’s University Belfast, MARCO ZAFFALON, IDSIA	1470
Unifying Count-Based Exploration and Intrinsic Motivation , MARC BELLEMARE, Google DeepMind, SRINIVASAN SRIRAM, GEORG OSTROVSKI, Google DeepMind, TOM SCHAUL, and DAVID SAXTON, REMI MUNOS, Google DeepMind	1479
Large Margin Discriminant Dimensionality Reduction in Prediction Space , MOHAMMAD SABERIAN, Netflix, JOSE PEREIRA, UC San Diego, CAN XU, Google, NUNO NVASCONCELOS, UC San Diego	1488
Stochastic Structured Prediction under Bandit Feedback , ARTEM SOKOLOV, JULIA KREUTZER, Heidelberg University, CHRISTOPHER LO, STEFAN RIEZLER, Heidelberg University	1497
Simple and Efficient Weighted Minwise Hashing , ANSHUMALI SHRIVASTAVA, Rice University	1506

Truncated Variance Reduction: A Unified Approach to Bayesian Optimization and Level-Set Estimation, ILIJA BOGUNOVIC, EPFL Lausanne, JONATHAN SCARLETT, ANDREAS KRAUSE, VOLKAN CEVHER ...	1515
Structured Sparse Regression via Greedy Hard Thresholding, PRATEEK JAIN, Microsoft Research, NIKHIL RAO, INDERJIT DHILLON, University of Texas at Austin	1524
Understanding Probabilistic Sparse Gaussian Process Approximations, MATTHIAS BAUER, MARK WILK, CARL RASMUSSEN, University of Cambridge	1533
SEBOOST - Boosting Stochastic Learning Using Subspace Optimization Techniques, ELAD RICHARDSON, Technion, ROM HERSKOVITZ, BORIS GINSBURG, MICHAEL ZIBULEVSKY	1542
Generating Long-term Trajectories Using Deep Hierarchical Networks, STEPHAN ZHENG, YISONG YUE, Caltech, PATRICK LUCEY, Stats	1551
Learning Tree Structured Potential Games, VIKAS GARG, TOMMI JAAKKOLA, Massachusetts Institute of Technology	1560
Observational-Interventional Priors for Dose-Response Learning, RICARDO SILVA	1569
Learning from Rational Behavior: Predicting Solutions to Unknown Linear Programs, SHAHIN JABBARI, RYAN ROGERS, AARON ROTH, STEVEN WU, University of Pennsylvania	1578
Identification and Overidentification of Linear Structural Equation Models, BRYANT CHEN, University of California Los Angeles	1587
Adaptive Skills Adaptive Partitions (ASAP), DANIEL MANKOWITZ, Technion, TIMOTHY MANN, Google DeepMind, SHIE MANNOR, Technion	1596
Multiple-Play Bandits in the Position-Based Model, PAUL LAGRÉE, Université Paris Sud, CLAIRE VERNADE, Université Paris Saclay, OLIVIER CAPPE	1605
Optimal Black-Box Reductions Between Optimization Objectives, ZEYUAN ALLEN-ZHU, ELAD HAZAN, Princeton University	1614
On Valid Optimal Assignment Kernels and Applications to Graph Classification, NILS KRIEGE, TU Dortmund, PIERRE-LOUIS GISCARD, RICHARD WILSON, University of York	1623
Robustness of classifiers: from adversarial to random noise, ALHUSSEIN FAWZI, SEYED-MOHSEN MOOSAVI-DEZFOOLI, PASCAL FROSSARD, École polytechnique fédérale de Lausanne	1632
A Non-convex One-Pass Framework for Generalized Factorization Machine and Rank-One Matrix Sensing, MING LIN, JIEPING YE, University of Michigan	1641
Exploiting the Structure: Stochastic Gradient Methods Using Raw Clusters, ZEYUAN ALLEN-ZHU, Princeton University, YANG YUAN, Cornell University, KARTHIK SRIDHARAN, University of Pennsylvania	1650
Combinatorial Multi-Armed Bandit with General Reward Functions, WEI CHEN, WEI HU, Princeton University, FU LI, The	

University of Texas at Austin, JIAN LI, and YU LIU, Tsinghua University, PINYAN LU, Shanghai University of Finance and Economics	1659
Boosting with Abstention, CORINNA CORTES, GIULIA DESALVO, MEHRYAR MOHRI, New York University	1668
Regret of Queueing Bandits, SUBHASHINI KRISHNASAMY, RAJAT SEN, The University of Texas at Austin, RAMESH JOHARI, Stanford University, SANJAY SHAKKOTTAI, The University of Texas at Aus	1677
Deep Learning Games, DALE SCHUURMANS, MARTIN ZINKEVICH, Google	1686
Globally Optimal Training of Generalized Polynomial Neural Networks with Nonlinear Spectral Methods, ANTOINE GAUTIER, QUYNH NGUYEN, MATTHIAS HEIN, Saarland University	1695
Perspective Transformer Nets: Learning Single-View 3D Object Reconstruction without 3D Supervision, XINCHEN YAN, University of Michigan, JIMEI YANG, ERSIN YUMER, Adobe Research, YIJIE GUO, and HONGLAK LEE, University of Michigan	1704
A Credit Assignment Compiler for Joint Prediction, KAI-WEI CHANG, HE HE, University of Maryland, HAL III, JOHN LANGFORD, and STEPHANE ROSS, Google	1713
Accelerating Stochastic Composition Optimization, MENGDI WANG, Princeton University, JI LIU, ETHAN FANG, Pennsylvania State University	1722
Reward Augmented Maximum Likelihood for Neural Structured Prediction, MOHAMMAD NOROUZI, SAMY BENGIO, ZHIFENG CHEN, NAVDEEP JAITLEY, and MIKE SCHUSTER, YONGHUI WU, DALE SCHUURMANS	1731
Consistent Kernel Mean Estimation for Functions of Random Variables, CARL-JOHANN SIMON-GABRIEL, MPI Tuebingen, ADAM SCIBIOR, University of Cambridge, ILYA TOLSTIKHIN, BERNHARD SCHÖLKOPF	1740
Towards Unifying Hamiltonian Monte Carlo and Slice Sampling, YIZHE ZHANG, Duke university, XIANGYU WANG, Duke University, CHANGYOU CHEN, RICARDO HENAO, Duke University, and KAI FAN, Duke university, LAWRENCE CARIN, Duke University	1749
Scalable Adaptive Stochastic Optimization Using Random Projections, GABRIEL KRUMMENACHER, Eidgenössische Technische Hochschule Zürich, BRIAN MCWILLIAMS, Disney Research, YANNIC KILCHER, JOACHIM BUHMANN, Eidgenössische Technische Hochschule Zürich, and NICOLAI MEINSHAUSEN	1758
Variational Inference in Mixed Probabilistic Submodular Models, JOSIP DJOLONGA, SEBASTIAN TSCHIATSCHEK, Eidgenössische Technische Hochschule Zürich, ANDREAS KRAUSE	1767
Correlated-PCA: Principal Components' Analysis when Data and Noise are Correlated, NAMRATA VASWANI, HAN GUO, Iowa State University	1776
The Multi-fidelity Multi-armed Bandit, KIRTHEVASAN KANDASAMY, GAUTAM DASARATHY, BARNABAS POZOS, JEFF SCHNEIDER, Carnegie Mellon University	1785

Anchor-Free Correlated Topic Modeling: Identifiability and Algorithm, KEJUN HUANG, XIAO FU, NICHOLAS SIDIROPOULOS, University of Minnesota	1794
Bootstrap Model Aggregation for Distributed Statistical Learning, JUN HAN, QIANG LIU, Dartmouth College	1803
A scalable end-to-end Gaussian process adapter for irregularly sampled time series classification, STEVEN LI, BENJAMIN MARLIN, UMass Amherst	1812
A Bandit Framework for Strategic Regression, YANG LIU, YILING CHEN, Harvard University	1821
Architectural Complexity Measures of Recurrent Neural Networks, SAIZHENG ZHANG, University of Montreal, YUHUI WU, University of Toronto, TONG CHE, IHES, ZHOUEHAN LIN, and ROLAND MEMISEVIC, University of Montreal, RUSLAN SALAKHUTDINOV, University of Toronto, YOSHUA BENGIO, U. Montreal	1830
Statistical Inference for Cluster Trees, JISU KIM, YEN-CHI CHEN, SIVARAMAN BALAKRISHNAN, ALESSANDRO RINALDO, and LARRY WASSERMAN, Carnegie Mellon University	1839
PAC Reinforcement Learning with Rich Observations, AKSHAY KRISHNAMURTHY, Carnegie Mellon University, ALEKH AGARWAL, Microsoft, JOHN LANGFORD	1848
Improved Deep Metric Learning with Multi-class N-pair Loss Objective, KIHYUK SOHN	1857
Unsupervised Learning of Spoken Language with Visual Context, DAVID HARWATH, ANTONIO TORRALBA, JAMES GLASS, MIT CSAIL	1866
Low-Rank Regression with Tensor Responses, GUILLAUME RABUSSEAU, Aix-Marseille University, HACHEM KADRI	1875
PAC-Bayesian Theory Meets Bayesian Inference, PASCAL GERMAIN, FRANCIS BACH, ALEXANDRE LACOSTE, SIMON LACOSTE-JULIEN, INRIA ..	1884
Data Poisoning Attacks on Factorization-Based Collaborative Filtering, BO LI, Vanderbilt University, YINING WANG, AARTI SINGH, Carnegie Mellon University, YEVGENIY VOROBAYCHIK, Vanderbilt University ..	1893
Learned Region Sparsity and Diversity Also Predicts Visual Attention, ZIJUN WEI, Stony Brook, HOSSEIN ADELI, Stony Brook University, MINH HOAI, Stony Brook, GREGORY ZELINSKY, Stony Brook University, and DIMITRIS SAMARAS	1902
End-to-End Goal-Driven Web Navigation, RODRIGO NOGUEIRA, New York University, KYUNGHYUN CHO, University of Montreal	1911
Automated scalable segmentation of neurons from multispectral images, UYGAR SÜMBÜL, Columbia University, DOUGLAS ROOSSEN, University of Michigan Ann Arbor, FEI CHEN, NICHOLAS BARRY, and EDWARD BOYDEN, Massachusetts Institute of Technology, DAWEN CAI, University of Michigan Ann Arbor, JOHN CUNNINGHAM, LIAM PANINSKI, Columbia University	1920
Privacy Odometers and Filters: Pay-as-you-Go Composition, RYAN ROGERS, University of Pennsylvania, SALIL VADHAN, Harvard University, AARON ROTH, University of Pennsylvania, JONATHAN ULLMAN ..	1929

Minimax Estimation of Maximum Mean Discrepancy with Radial Kernels, ILYA TOLSTIKHIN, BHARATH SRIPERUMBUDUR, Penn State University, BERNHARD SCHÖLKOPF	1938
Adaptive optimal training of animal behavior, JI BAK, JUNG CHOI, ATHENA AKRAMI, ILANA WITTEN, and JONATHAN PILLOW, Princeton University	1947
Hierarchical Object Representation for Open-Ended Object Category Learning and Recognition, HAMIDREZA KASAEI, IEETA University of Aveiro, ANA TOME, LUIS LOPES, University of Aveiro Portugal ..	1956
Relevant sparse codes with variational information bottleneck, MATTHEW CHALK, IST Austria, OLIVIER MARRE, Institut de la vision, GAŠPER TKAČIK, Institute of Science and Technology Austria	1965
Combinatorial Energy Learning for Image Segmentation, JEREMY MAITIN-SHEPARD, VIREN JAIN, MICHAL JANUSZEWSKI, Google, PETER LI, and PIETER ABBEEL, OpenAI / UC Berkeley / Gradescope	1974
Orthogonal Random Features, FELIX YU, ANANDA SURESH, UC San Diego, KRZYSZTOF CHOROMANSKI, DAN HOLTSMANN-RICE, and SANJIV KUMAR, Google	1983
Fast Active Set Methods for Online Spike Inference from Calcium Imaging, JOHANNES FRIEDRICH, LIAM PANINSKI, Columbia University ...	1992
Diffusion-Convolutional Neural Networks, JAMES ATWOOD, DON TOWSLEY, UMass Amherst	2001
Bayesian latent structure discovery from multi-neuron recordings, SCOTT LINDERMAN, Columbia University, RYAN ADAMS, Harvard University, JONATHAN PILLOW, Princeton University	2010
A Probabilistic Programming Approach To Probabilistic Data Analysis, FERAS SAAD, VIKASH MANSINGHKA, Massachusetts Institute of Technology	2019
A Non-parametric Learning Method for Confidently Estimating Patient’s Clinical State and Dynamics, WILLIAM HOILES, University of California Los, MIHAELA SCHAAR	2028
Inference by Reparameterization in Neural Population Codes, RAJKUMAR RAJU, XAQ PITKOW, Rice University	2037
Tensor Switching Networks, CHUAN-YUNG TSAI, Harvard University, ANDREW SAXE, DAVID COX, Harvard University	2046
Stochastic Gradient Richardson-Romberg Markov Chain Monte Carlo, ALAIN DURMUS, Telecom ParisTech, Umut SIMSEKLI, ERIC MOULINES, Ecole Polytechnique, ROLAND BADEAU, and GAËL RICHARD, Telecom ParisTech	2055
Coordinate-wise Power Method, QI LEI, KAI ZHONG, UT AUSTIN, INDERJIT DHILLON, University of Texas at Austin	2064
Learning Influence Functions from Incomplete Observations, XINRAN HE, KE XU, DAVID KEMPE, YAN LIU, University of Southern California	2073
Learning Structured Sparsity in Deep Neural Networks, WEI WEN, CHUNPENG WU, YANDAN WANG, YIRAN CHEN, University of Pittsburgh, and HAI LI, University of Pittsburgh	2082

Sample Complexity of Automated Mechanism Design , NINA BALCAN, TUOMAS SANDHOLM, ELLEN VITERCIK, Carnegie Mellon University	2091
Short-Dot: Computing Large Linear Transforms Distributedly Using Coded Short Dot Products , SANGHAMITRA DUTTA, Carnegie Mellon University, VIVECK CADAMBE, Pennsylvania State University, PULKIT GROVER, Carnegie Mellon University	2100
Brains on Beats , UMUT GÜÇLÜ, JORDY THIELEN, Radboud University, MICHAEL HANKE, Otto-von-Guericke University Magdeburg, MARCEL GERVEN, Radboud University	2109
Learning Transferrable Representations for Unsupervised Domain Adaptation , OZAN SENER, Cornell University, HYUN SONG, Google Research, ASHUTOSH SAXENA, Brain of Things, SILVIO SAVARESE, Stanford University	2118
Stochastic Multiple Choice Learning for Training Diverse Deep Ensembles , STEFAN LEE, Indiana University, SENTHIL PURUSHWALKAM, Carnegie Mellon, MICHAEL COGSWELL, VIRESH RANJAN, Virginia Tech, and DAVID CRANDALL, Indiana University, DHURUV BATRA, Virginia Tech	2127
Active Learning from Imperfect Labelers , SONGBAI YAN, KAMALIKA CHAUDHURI, TARA JAVIDI, University of California San Diego	2136
Learning to Communicate with Deep Multi-Agent Reinforcement Learning , JAKOB FOERSTER, YANNIS ASSAEL, NANDO FREITAS, University of Oxford, SHIMON WHITESON	2145
Value Iteration Networks , AVIV TAMAR, YI WU, GARRETT THOMAS, SERGEY LEVINE, UC Berkeley, and PIETER ABBEEL, OpenAI / UC Berkeley / Gradescope	2154
Blind Regression: Nonparametric Regression for Latent Variable Models via Collaborative Filtering , CHRISTINA LEE, YIHUA LI, DEVAVRAT SHAH, DOGYOON SONG, Massachusetts Institute of Technology	2163
On the Recursive Teaching Dimension of VC Classes , XI CHEN, Columbia University, YU CHENG, U of Southern California, BO TANG, University of Oxford	2172
InfoGAN: Interpretable Representation Learning by Information Maximizing Generative Adversarial Nets , XI CHEN, UC Berkeley and OpenAI, YAN DUAN, UC Berkeley, REIN HOUTHOOFT, Ghent University - iMinds and UC Berkeley and OpenAI, JOHN SCHULMAN, OpenAI, and ILYA SUTSKEVER, PIETER ABBEEL, OpenAI / UC Berkeley / Gradescope	2180
Hardness of Online Sleeping Combinatorial Optimization Problems , SATYEN KALE, CHANSOO LEE, University of Michigan, DAVID PAL	2189
Mixed Linear Regression with Multiple Components , KAI ZHONG, UT AUSTIN, PRATEEK JAIN, Microsoft Research, INDERJIT DHILLON, University of Texas at Austin	2198
Sequential Neural Models with Stochastic Layers , MARCO FRACCARO, DTU, SØREN SØNDERBY, KU, ULRICH PAQUET, OLE WINTHER, DTU	2207
Stochastic Gradient Methods for Distributionally Robust Optimization with f-divergences , HONGSEOK NAMKOONG, JOHN DUCHI, Stanford University	2216

Minimizing Quadratic Functions in Constant Time , KOHEI HAYASHI, AIST, YUICHI YOSHIDA, NII	2225
Improved Techniques for Training GANs , TIM SALIMANS, IAN GOODFELLOW, WOJCIECH ZAREMBA, VICKI CHEUNG, and ALEC RADFORD, OpenAI, XI CHEN, UC Berkeley and OpenAI	2234
DeepMath - Deep Sequence Models for Premise Selection , ALEXANDER ALEMI, Google, FRANCOIS CHOLLET, NIKLAS EEN, Google Inc., GEOFFREY IRVING, and CHRISTIAN SZEGEDY, JOSEF URBAN, Czech Technical University in Prague	2243
Learning Multiagent Communication with Backpropagation , SAINBAYAR SUKHBAATAR, New York University, ARTHUR SZLAM, ROB FERGUS, New York University	2252
Toward Deeper Understanding of Neural Networks: The Power of Initialization and a Dual View on Expressivity , AMIT DANIELY, Google Brain, ROY FROSTIG, Stanford University, YORAM SINGER, Google ...	2261
Learning the Number of Neurons in Deep Networks , JOSE ALVAREZ, NICTA, MATHIEU SALZMANN, École polytechnique fédérale de Lausanne	2270
Finding significant combinations of features in the presence of categorical covariates , LAETITIA PAPAXANTHOS, FELIPE LLINARES, DEAN BODENHAM, Eidgenössische Technische Hochschule Zürich, KARSTEN BORWARDT	2279
Examples are not enough, learn to criticize! Criticism for Interpretability , BEEN KIM, RAJIV KHANNA, UT Austin, SANMI KOYEJO, University of Illinois at Urbana-Champaign	2288
Optimistic Bandit Convex Optimization , MEHRYAR MOHRI, SCOTT YANG, New York University	2297
Safe Policy Improvement by Minimizing Robust Baseline Regret , MAREK PETRIK, MOHAMMAD GHAVAMZADEH, YINLAM CHOW, Stanford University	2306
Graphons, mergeons, and so on! , JUSTIN ELDRIDGE, MIKHAIL BELKIN, YUSU WANG, The Ohio State University	2315
Hierarchical Clustering via Spreading Metrics , AURKO ROY, Georgia Tech, SEBASTIAN POKUTTA, GeorgiaTech	2324
Learning Bayesian networks with ancestral constraints , EUNICE CHEN, YUJIA SHEN, ARTHUR CHOI, ADNAN DARWICHE, University of California Los Angeles	2333
Pruning Random Forests for Prediction on a Budget , FENG NAN, JOSEPH WANG, VENKATESH SALIGRAMA, Boston University	2342
Clustering with Bregman Divergences: an Asymptotic Analysis , CHAOYUE LIU, MIKHAIL BELKIN, The Ohio State University	2351
Variational Autoencoder for Deep Learning of Images, Labels and Captions , YUNCHEN PU, Duke University, ZHE GAN, Duke, RICARDO HENAO, Duke University, XIN YUAN, Bell Labs, and CHUNYUAN LI, Duke, ANDREW STEVENS, LAWRENCE CARIN, Duke University	2360

Review Networks for Caption Generation, ZHILIN YANG, YE YUAN, YUEXIN WU, Carnegie Mellon University, RUSLAN SALAKHUTDINOV, University of Toronto, and WILLIAM COHEN, Carnegie Mellon University	2369
Stein Variational Gradient Descent: A General Purpose Bayesian Inference Algorithm, QIANG LIU, DILIN WANG, Dartmouth College	2378
A Bio-inspired Redundant Sensing Architecture, ANH NGUYEN, JIAN XU, ZHI YANG, University of Minnesota	2387
Contextual semibandits via supervised learning oracles, AKSHAY KRISHNAMURTHY, Carnegie Mellon University, ALEKH AGARWAL, Microsoft, MIRO DUDIK	2396
Blind Attacks on Machine Learners, ALEX BEATSON, ZHAORAN WANG, HAN LIU, Princeton University	2405
Universal Correspondence Network, CHRISTOPHER CHOY, JUNYOUNG GWAK, SILVIO SAVARESE, Stanford University, MANMOHAN CHANDRAKER, NEC Labs America	2414
Satisfying Real-world Goals with Dataset Constraints, GABRIEL GOH, UC Davis, ANDREW COTTER, Google, MAYA GUPTA, MICHAEL FRIEDLANDER, UC Davis	2423
Deep Learning for Predicting Human Strategic Behavior, JASON HARTFORD, JAMES WRIGHT, University of British Columbia, KEVIN LEYTON-BROWN	2432
Phased Exploration with Greedy Exploitation in Stochastic Combinatorial Partial Monitoring Games, SOUGATA CHAUDHURI, AMBUJ TEWARI, University of Michigan	2441
Eliciting Categorical Data for Optimal Aggregation, CHIEN-JU HO, Cornell University, RAFAEL FRONGILLO, YILING CHEN, Harvard University	2450
Measuring the reliability of MCMC inference with bidirectional Monte Carlo, ROGER GROSSE, SIDDHARTH ANCHA, University of Toronto, DANIEL ROY	2459
Breaking the Bandwidth Barrier: Geometrical Adaptive Entropy Estimation, WEIHAO GAO, SEWOONG OH, PRAMOD VISWANATH, University of Illinois at Urbana-Champaign	2468
Selective inference for group-sparse linear models, FAN YANG, RINA BARBER, University of Chicago, PRATEEK JAIN, Microsoft Research, JOHN LAFFERTY	2477
Graph Clustering: Block-models and model free results, YALI WAN, MARINA MEILA, University of Washington	2486
Maximizing Influence in an Ising Network: A Mean-Field Optimal Solution, CHRISTOPHER LYNN, DANIEL LEE, University of Pennsylvania	2495
Hypothesis Testing in Unsupervised Domain Adaptation with Applications in Alzheimer’s Disease, HAO ZHOU, SATHYA RAVI, University of Wisconsin Madiso, VAMSI ITHAPU, STERLING JOHNSON, and GRACE WAHBA, University of Wisconsin Madison, VIKAS SINGH, UW Madison	2504
Geometric Dirichlet Means Algorithm for topic inference, MIKHAIL YUROCHKIN, XUANLONG NGUYEN, University of Michigan	2513

Structured Prediction Theory Based on Factor Graph Complexity, CORINNA CORTES, VITALY KUZNETSOV, Courant Institute, MEHRYAR MOHRI, SCOTT YANG, New York University	2522
Improved Dropout for Shallow and Deep Learning, ZHE LI, The University of Iowa, BOQING GONG, University of Central Florida, TIANBAO YANG, University of Iowa	2531
Constraints Based Convex Belief Propagation, YANIV TENZER, The Hebrew University, ALEXANDER SCHWING, University of Toronto, KEVIN GIMPEL, TTI-Chicago, TAMIR HAZAN	2540
Error Analysis of Generalized Nyström Kernel Regression, HONG CHEN, University of Texas, HAIFENG XIA, Huazhong Agricultural University, WEIDONG CAI, University of Sydney, HENG HUANG, University of Texas Arlington	2549
A Probabilistic Framework for Deep Learning, ANKIT PATEL, Baylor College of Medicine and Rice University, TAN NGUYEN, RICHARD BARANIUK, Rice University	2558
General Tensor Spectral Co-clustering for Higher-Order Data, TAO WU, Purdue University, AUSTIN BENSON, Stanford University, DAVID GLEICH, Purdue University	2567
Cyclades: Conflict-free Asynchronous Machine Learning, XINGHAO PAN, MAXIMILIAN LAM, STEPHEN TU, DIMITRIS PAPAILIOPOULOS, UC Berkeley, and CE ZHANG, Stanford University, MICHAEL JORDAN, OpenAI / UC Berkeley / Gradescope, KANNAN RAMCHANDRAN, UC Berkeley, CHRISTOPHER RE, Stanford University, BEN RECHT, UC Berkeley	2576
Single Pass PCA of Matrix Products, SHANSHAN WU, UT Austin, SRINADH BHOJANAPALLI, TTI Chicago, SUJAY SANGHAVI, ALEXANDROS DIMAKIS, The University of Texas at Aus	2585
Stochastic Variational Deep Kernel Learning, ANDREW WILSON, ZHITING HU, Carnegie Mellon University, RUSLAN SALAKHUTDINOV, University of Toronto, ERIC XING, Carnegie Mellon University	2594
Interaction Screening: Efficient and Sample-Optimal Learning of Ising Models, MARC VUFFRAY, SIDHANT MISRA, ANDREY LOKHOV, MISHA CHERTKOV, Los Alamos National Laboratory	2603
Long-term Causal Effects via Behavioral Game Theory, PANOS TOULIS, University of Chicago, DAVID PARKES, Harvard University	2612
Measuring Neural Net Robustness with Constraints, OSBERT BASTANI, Stanford University, YANI IOANNOU, University of Cambridge, LEONIDAS LAMPROPOULOS, University of Pennsylvania, DIMITRIOS VYTINIOTIS, and ADITYA NORI, Microsoft Research, ANTONIO CRIMINISI ...	2621
Reshaped Wirtinger Flow for Solving Quadratic System of Equations, HUIHUAI ZHANG, YINGBIN LIANG, Syracuse University	2630
Nearly Isometric Embedding by Relaxation, JAMES MCQUEEN, MARINA MEILA, University of Washington, DOMINIQUE JONCAS, Google	2639
Probabilistic Inference with Generating Functions for Poisson Latent Variable Models, KEVIN WINNER, UMass CICS, DANIEL SHELDON, UMass Amherst	2648

Causal meets Submodular: Subset Selection with Directed Information, YUXUN ZHOU, UC Berkeley, COSTAS SPANOS	2657
Depth from a Single Image by Harmonizing Overcomplete Local Network Predictions, AYAN CHAKRABARTI, TTI-Chicago, JINGYU SHAO, University of California Los Angeles, GREG SHAKHNAROVICH, TTI-Chicago ...	2666
Deep Neural Networks with Inexact Matching for Person Re-Identification, ARULKUMAR SUBRAMANIAM, MOITREYA CHATTERJEE, ANURAG MITTAL, IIT Madras	2675
Global Analysis of Expectation Maximization for Mixtures of Two Gaussians, JI XU, Columbia university, DANIEL HSU, ARIAN MALEKI, Columbia University	2684
Estimating the class prior and posterior from noisy positives and unlabeled data, SHANATNU JAIN, MARTHA WHITE, PREDRAG RADIVOJAC, Indiana University	2693
Kronecker Determinantal Point Processes, ZELDA MARIET, SUVRIT SRA, Massachusetts Institute of Technology	2702
Finite Sample Prediction and Recovery Bounds for Ordinal Embedding, LALIT JAIN, University of Michigan, KEVIN JAMIESON, UC Berkeley, ROBERT NOWAK, University of Wisconsin Madison	2711
Feature-distributed sparse regression: a screen-and-clean approach, JIYAN YANG, Stanford University, MICHAEL MAHONEY, MICHAEL SAUNDERS, Stanford University, YUEKAI SUN, University of Michigan	2720
Learning Bound for Parameter Transfer Learning, WATARU KUMAGAI, Kanagawa University	2729
Learning under uncertainty: a comparison between R-W and Bayesian approach, HE HUANG, MARTIN PAULUS, LIBR	2738
Bi-Objective Online Matching and Submodular Allocations, HOSSEIN ESFANDIARI, University of Maryland, NITISH KORULA, Google Research, VAHAB MIRROKNI, Google	2747
Quantized Random Projections and Non-Linear Estimation of Cosine Similarity, PING LI, Rutgers University and Amazon, MICHAEL MITZENMACHER, Harvard University, MARTIN SLAWSKI	2756
The non-convex Burer-Monteiro approach works on smooth semidefinite programs, NICOLAS BOUMAL, VLAD VORONINSKI, Massachusetts Institute of Technology, AFONSO BANDEIRA	2765
Dimensionality Reduction of Massive Sparse Datasets Using Coresets, DAN FELDMAN, MIKHAIL VOLKOV, DANIELA RUS, Massachusetts Institute of Technology	2774
Using Social Dynamics to Make Individual Predictions: Variational Inference with a Stochastic Kinetic Model, ZHEN XU, WEN DONG, SUNY at Buffalo, SARGUR SRIHARI	2783
On statistical learning via the lens of compression, OFIR DAVID, Technion - Israel institute of technology, SHAY MORAN, AMIR YEHUDAYOFF, Technion - Israel institue of Technology	2792

Generative Shape Models: Joint Text Recognition and Segmentation with Very Little Training Data, XINGHUA LOU, Vicarious FPC Inc, KEN KANSKY, WOLFGANG LEHRACH, Vicarious, CC LAAN, and BHASKARA MARTHI, D. PHOENIX, DILEEP GEORGE	2801
Image Restoration Using Very Deep Convolutional Encoder-Decoder Networks with Symmetric Skip Connections, XIAO-JIAO MAO, Nanjing University, CHUNHUA SHEN, YU-BIN YANG	2810
Object based Scene Representations using Fisher Scores of Local Subspace Projections, MANDAR DIXIT, UC San Diego, NUNO VASCONCELOS	2819
Active Learning with Oracle Epiphany, TZU-KUO HUANG, Uber Advanced Technologies Center, LIHONG LI, Microsoft Research, ARA VARTANIAN, University of Wisconsin-Madison, SALEEMA AMERSHI, Microsoft, and XIAOJIN ZHU, University of Wisconsin-Madison	2828
Statistical Inference for Pairwise Graphical Models Using Score Matching, MING YU, The University of Chicago, MLADEN KOLAR, VARUN GUPTA, University of Chicago	2837
Improved Error Bounds for Tree Representations of Metric Spaces, SAMIR CHOWDHURY, The Ohio State University, FACUNDO MEMOLI, ZANE SMITH, The Ohio State University	2846
Can Peripheral Representations Improve Clutter Metrics on Complex Scenes?, ARTURO DEZA, MIGUEL ECKSTEIN, UCSB	2855
On Multiplicative Integration with Recurrent Neural Networks, YUHUI WU, University of Toronto, SAIZHENG ZHANG, YING ZHANG, University of Montreal, YOSHUA BENGIO, U. Montreal, and RUSLAN SALAKHUTDINOV, University of Toronto	2864
Learning HMMs with Nonparametric Emissions via Spectral Decompositions of Continuous Matrices, KIRTHEVASAN KANDASAMY, MARUAN AL-SHEDIVAT, ERIC XING, Carnegie Mellon University	2873
Regret Bounds for Non-decomposable Metrics with Missing Labels, NAGARAJAN NATARAJAN, PRATEEK JAIN, Microsoft Research	2882
Robust k-means: a Theoretical Revisit, ALEXANDROS GEORGOGIANNIS, TECHNICAL UNIVERSITY OF CRETE	2891
Bayesian optimization for automated model selection, GUSTAVO MALKOMES, Washington University, CHARLES SCHAFF, Washington University in St. Louis, ROMAN GARNETT, Washington University	2900
A Probabilistic Model of Social Decision Making based on Reward Maximization, KOOSHA KHALVATI, University of Washington, SEONGMIN PARK, Cognitive Neuroscience Center, JEAN-CLAUDE DREHER, Centre de Neurosciences Cognitives, RAJESH RAO, University of Washington	2909
Balancing Suspense and Surprise: Timely Decision Making with Endogenous Information Acquisition, AHMED ALAA, University of California Los Angeles, MIHAELA SCHAAR	2918
Fast and Flexible Monotonic Functions with Ensembles of Lattices, KEVIN CANINI, ANDREW COTTER, Google, MAYA GUPTA, MAHDI FARD, and JAN PFEIFER, Google	2927

Conditional Generative Moment-Matching Networks, YONG REN, Tsinghua University, JUNE ZHU, JIALIAN LI, YUCEN LUO, Tsinghua University	2936
Stochastic Gradient MCMC with Stale Gradients, CHANGYOU CHEN, NAN DING, Google, CHUNYUAN LI, Duke, YIZHE ZHANG, Duke university, and LAWRENCE CARIN, Duke University	2945
Composing graphical models with neural networks for structured representations and fast inference, MATTHEW JOHNSON, Massachusetts Institute of Technology, DAVID DUVENAUD, Harvard University, ALEX WILTSCHKO, Harvard University and Twitter, RYAN ADAMS, Harvard University, and SANDEEP DATTA, Harvard Medical School	2954
Noise-Tolerant Life-Long Matrix Completion via Adaptive Sampling, NINA BALCAN, HONGYANG ZHANG, Carnegie Mellon University	2963
Combinatorial semi-bandit with known covariance, RÉMY DEGENNE, Université Paris Diderot, VIANNEY PERCHET, Ensae - Criteo Labs	2972
Matrix Completion has No Spurious Local Minimum, RONG GE, JASON LEE, UC Berkeley, TENGYU MA, Princeton University	2981
The Multiscale Laplacian Graph Kernel, RISI KONDOR, HORACE PAN, UChicago	2990
Adaptive Averaging in Accelerated Descent Dynamics, WALID KRICHENE, ALEXANDRE BAYEN, UC Berkeley, PETER BARTLETT, OpenAI / UC Berkeley / Gradescope	2999
Sub-sampled Newton Methods with Non-uniform Sampling, PENG XU, JIYAN YANG, Stanford University, FARBOD ROOSTA-KHORASANI, University of California Berkeley, CHRISTOPHER RE, Stanford University, and MICHAEL MAHONEY	3008
Stochastic Gradient Geodesic MCMC Methods, CHANG LIU, Tsinghua University, JUNE ZHU, YANG SONG, Stanford University	3017
Variational Bayes on Monte Carlo Steroids, ADITYA GROVER, STEFANO ERMON, Stanford University	3026
Showing versus doing: Teaching by demonstration, MARK HO, Brown University, MICHAEL LITTMAN, JAMES MACGLASHAN, Brown University, FIERY CUSHMAN, Harvard University, and JOE AUSTERWEIL, University of Wisconsin-Madison	3035
Combining Fully Convolutional and Recurrent Neural Networks for 3D Biomedical Image Segmentation, JIANXU CHEN, LIN YANG, YIZHE ZHANG, MARK ALBER, and DANNY CHEN, University of Notre Dame	3044
Maximization of Approximately Submodular Functions, THIBAUT HOREL, YARON SINGER, Harvard University	3053
A Comprehensive Linear Speedup Analysis for Asynchronous Stochastic Parallel Optimization from Zeroth-Order to First-Order, XIANGRU LIAN, University of Rochester, HUAN ZHANG, CHO-JUI HSIEH, UC-Davis, YIJUN HUANG, and JI LIU	3062
Learning Infinite RBMs with Frank-Wolfe, WEI PING, UC Irvine, QIANG LIU, Dartmouth College, ALEX IHLER	3071
Estimating the Size of a Large Network and its Communities from a Random Sample, LIN CHEN, AMIN KARBASI, FORREST CRAWFORD, Yale University	3080

Learning Sensor Multiplexing Design through Back-propagation, AYAN CHAKRABARTI, TTI-Chicago	3089
On Robustness of Kernel Clustering, BOWEI YAN, University of Texas at Austin, PURNAMRITA SARKAR, U.C. Berkeley	3098
High resolution neural connectivity from incomplete tracing data using nonnegative spline regression, KAMERON HARRIS, University of Washington, STEFAN MIHALAS, Allen Institute for Brain Science, ERIC SHEA-BROWN, University of Washington	3107
MoCap-guided Data Augmentation for 3D Pose Estimation in the Wild, GREGORY ROGEZ, Inria, CORDELIA SCHMID	3116
New Lifiable Classes for First-Order Probabilistic Inference, SEYED KAZEMI, UBC, ANGELIKA KIMMIG, KU Leuven, GUY BROECK, University of California Los Angeles, DAVID POOLE, UBC	3125
The Parallel Knowledge Gradient Method for Batch Bayesian Optimization, JIAN WU, PETER FRAZIER, Cornell University	3134
Improved Regret Bounds for Oracle-Based Adversarial Contextual Bandits, VASILIS SYRGKANIS, HAIPENG LUO, Princeton University, AKSHAY KRISHNAMURTHY, Carnegie Mellon University, ROBERT SCHAPIRE ..	3143
Consistent Estimation of Functions of Data Missing Non- Monotonically and Not at Random, ILYA SHPITSER, Johns Hopkins University	3152
Optimistic Gittins Indices, ELI GUTIN, Massachusetts Institute of Tec, VIVEK FARIAS, Massachusetts Institute of Technology	3161
Finite-Dimensional BFRY Priors and Variational Bayesian Inference for Power Law Models, JUHO LEE, POSTECH, LANCELOT JAMES, HKUST, SEUNGJIN CHOI, POSTECH	3170
Launch and Iterate: Reducing Prediction Churn, QUENTIN CORMIER, Google, MAHDI FARD, KEVIN CANINI, MAYA GUPTA	3179
“Congruent” and “Opposite” Neurons: Sisters for Multisensory Integration and Segregation, WENHAO ZHANG, Institute of Neuroscience Chinese Academy of Sciences, HE WANG, K. WONG, HKUST, SI WU	3188
Learning shape correspondence with anisotropic convolutional neural networks, DAVIDE BOSCAINI, University of Lugano, JONATHAN MASCI, EMANUELE RODOLÀ, MICHAEL BRONSTEIN, University of Lugano ..	3197
Pairwise Choice Markov Chains, STEPHEN RAGAIN, JOHAN UGANDER, Stanford University	3206
NESTT: A Nonconvex Primal-Dual Splitting Method for Distributed and Stochastic Optimization, DAVOOD HAJINEZHAD, MINGYI HONG, Iowa State University, TUO ZHAO, Johns Hopkins University, ZHAORAN WANG, Princeton University	3215
Clustering with Same-Cluster Queries, HASSAN ASHTIANI, SHRINU KUSHAGRA, University of Waterloo, SHAI BEN-DAVID, U. Waterloo	3224
Attend, Infer, Repeat: Fast Scene Understanding with Generative Models, S. ESLAMI, Google DeepMind, NICOLAS HEESS, THEOPHANE WEBER, YUVAL TASSA, and DAVID SZEPEVARI, KORAY KAVUKCUOGLU, Google DeepMind, GEOFFREY HINTON, Google	3233

Parameter Learning for Log-supermodular Distributions, TATIANA SHPAKOVA, Inria - ENS Paris, FRANCIS BACH	3242
Deconvolving Feedback Loops in Recommender Systems, AYAN SINHA, Purdue, DAVID GLEICH, KARTHIK RAMANI, Purdue University	3251
Structured Matrix Recovery via the Generalized Dantzig Selector, SHENG CHEN, ARINDAM BANERJEE, University of Minnesota	3260
Confusions over Time: An Interpretable Bayesian Model to Characterize Trends in Decision Making, HIMABINDU LAKKARAJU, JURE LESKOVEC, Stanford University	3269
Automatic Neuron Detection in Calcium Imaging Data Using Convolutional Networks, NOAH APTHORPE, ALEXANDER RIORDAN, ROBERT AGUILAR, JAN HOMANN, and YI GU, DAVID TANK, H. SEUNG, Princeton University	3278
Designing smoothing functions for improved worst-case competitive ratio in online optimization, REZA EGHBALI, University of Washington, MARYAM FAZEL, University of Washington	3287
Convergence guarantees for kernel-based quadrature rules in misspecified settings, MOTONOBU KANAGAWA, BHARATH SRIPERUMBUDUR, Penn State University, KENJI FUKUMIZU	3296
Unsupervised Learning from Noisy Networks with Applications to Hi-C Data , BO WANG, JUNJIE ZHU, Stanford University, OANA URSU, Dept. of Genetics Stanford University, ARMIN POURSHAFEIE, Stanford University, and SERAFIM BATZOGLOU, Dept. of Computer Science Stanford University, ANSHUL KUNDAJE, Stanford University	3305
A Non-generative Framework and Convex Relaxations for Unsupervised Learning, ELAD HAZAN, TENGYU MA, Princeton University	3314
Equality of Opportunity in Supervised Learning, MORITZ HARDT, Google Brain, ERIC PRICE, University of Texas at Austin, NATI SREBRO, TTI-Chicago	3323
Scaled Least Squares Estimator for GLMs in Large-Scale Problems, MURAT A. ERDOGDU, MOHSEN BAYATI, Stanford University, LEE DICKER, Rutgers University and Amazon	3332
Interpretable Nonlinear Dynamic Modeling of Neural Trajectories, YUAN ZHAO, IL PARK, Stony Brook University	3341
Search Improves Label for Active Learning, ALINA BEYGELZIMER, Yahoo Inc, DANIEL HSU, Columbia University, JOHN LANGFORD, CHICHENG ZHANG, University of California San Diego	3350
Higher-Order Factorization Machines, MATHIEU BLONDEL, AKINORI FUJINO, NTT, NAONORI UEDA, MASAKAZU ISHIHATA, Hokkaido University .	3359
Exponential expressivity in deep neural networks through transient chaos, BEN POOLE, SUBHANEIL LAHIRI, Stanford University, MAITHRA RAGHU, Cornell University, JASCHA SOHL-DICKSTEIN, and SURYA GANGULI, Stanford University	3368
Split LBI: An Iterative Regularization Path with Structural Sparsity, CHENDI HUANG, XINWEI SUN, JIECHAO XIONG, Peking University, YUAN YAO	3377

An equivalence between high dimensional Bayes optimal inference and M-estimation, MADHU ADVANI, SURYA GANGULI, Stanford University	3386
Synthesizing the preferred inputs for neurons in neural networks via deep generator networks, ANH NGUYEN, University of Wyoming, ALEXEY DOSOVITSKIY, University of Freiburg, JASON YOSINSKI, Cornell, THOMAS BROX, University of Freiburg, and JEFF CLUNE	3395
Deep Submodular Functions: Definitions and Learning, BRIAN DOLHANSKY, University of Washington, JEFF BILMES, University of Washington Seattle	3404
Discriminative Gaiifman Models, MATHIAS NIEPERT	3413
Leveraging Sparsity for Efficient Submodular Data Summarization, ERIK LINDGREN, University of Texas at Austin, SHANSHAN WU, UT Austin, ALEXANDROS DIMAKIS, The University of Texas at Aus	3422
Local Minimax Complexity of Stochastic Convex Optimization, YUANCHENG ZHU, SABYASACHI CHATTERJEE, University of Chicago, JOHN DUCHI, Stanford University, JOHN LAFFERTY	3431
Stochastic Optimization for Large-scale Optimal Transport, AUDE GENEVAY, Université Paris Dauphine, MARCO CUTURI, GABRIEL PEYRÉ, FRANCIS BACH	3440
On Mixtures of Markov Chains, RISHI GUPTA, Stanford University, RAVI KUMAR, SERGEI VASSILVITSKII, Google	3449
Linear Contextual Bandits with Knapsacks, SHIPRA AGRAWAL, Columbia University, NIKHIL DEVANUR, Microsoft Research	3458
Reconstructing Parameters of Spreading Models from Partial Observations, ANDREY LOKHOV, Los Alamos National Laboratory	3467
Spatiotemporal Residual Networks for Video Action Recognition, CHRISTOPH FEICHTENHOFER, AXEL PINZ, Graz University of Technology, RICHARD WILDES, York University Toronto	3476
Path-Normalized Optimization of Recurrent Neural Networks with ReLU Activations, BEHNAM NEYSHABUR, TTI-Chicago, YUHUI WU, RUSLAN SALAKHUTDINOV, University of Toronto, NATI SREBRO, TTI-Chicago	3485
Strategic Attentive Writer for Learning Macro-Actions, ALEXANDER VEZHNEVETS, Google DeepMind, VOLODYMYR MNIH, SIMON OSINDERO, Google DeepMind, ALEX GRAVES, and ORIOL VINYALS, JOHN AGAPIOU, KORAY KAVUKCUOGLU, Google DeepMind	3494
The Limits of Learning with Missing Data, BRIAN BULLINS, ELAD HAZAN, Princeton University, TOMER KOREN, Google Brain	3503
RETAIN: An Interpretable Predictive Model for Healthcare using Reverse Time Attention Mechanism, EDWARD CHOI, Georgia Institute of Technology, MOHAMMAD BAHADORI, JOSHUA KULAS, Georgia Institute of Technology, ANDY SCHUETZ, and WALTER STEWART, Sutter Health, JIMENG SUN	3512
Total Variation Classes Beyond 1d: Minimax Rates, and the Limitations of Linear Smoothers, VEERANJANEYULU SADHANALA, YU-XIANG WANG, Carnegie Mellon University, RYAN TIBSHIRANI	3521

Community Detection on Evolving Graphs, ARIS ANAGNOSTOPOULOS, JAKUB ŁACKI, Sapienza University of Rome, SILVIO LATTANZI, Google, STEFANO LEONARDI, Sapienza University of Rome, and MOHAMMAD MAHDIAN, Google Research New York	3530
Online and Differentially-Private Tensor Decomposition, YINING WANG, Carnegie Mellon University, ANIMA ANANDKUMAR, UC Irvine	3539
Dimension-Free Iteration Complexity of Finite Sum Optimization Problems, YOSSI ARJEVANI, OHAD SHAMIR, Weizmann Institute of Science	3548
Towards Conceptual Compression, KAROL GREGOR, FREDERIC BESSE, Google DeepMind, DANILO REZENDE, IVO DANIHELKA, and DAAN WIERSTRA, Google DeepMind	3556
Exact Recovery of Hard Thresholding Pursuit, XIAOTONG YUAN, Nanjing University of Informat, PING LI, TONG ZHANG, Rutgers University and Amazon	3565
Data Programming: Creating Large Training Sets, Quickly, ALEXANDER RATNER, CHRISTOPHER SA, SEN WU, DANIEL SELSAM, and CHRISTOPHER RÉ, Stanford University	3574
Generalization of ERM in Stochastic Convex Optimization: The Dimension Strikes Back, VITALY FELDMAN	3583
Dynamic matrix recovery from incomplete observations under an exact low-rank constraint, LIANGBEI XU, MARK DAVENPORT, Georgia Institute of Technology	3592
Fast Distributed Submodular Cover: Public-Private Data Summarization, BAHARAN MIRZASOLEIMAN, Eidgenössische Technische Hochschule Zürich, MORTEZA ZADIMOGHADDAM, Google Research, AMIN KARBASI, Yale University	3601
Estimating Nonlinear Neural Response Functions using GP Priors and Kronecker Methods, CRISTINA SAVIN, IST Austria, GAŠPER TKAČIK, Institute of Science and Technology Austria	3610
Lifelong Learning with Weighted Majority Votes, ANASTASIA PENTINA, IST Austria, RUTH URNER, MPI Tuebingen	3619
Scaling Memory-Augmented Neural Networks with Sparse Reads and Writes, JACK RAE, Google DeepMind, JONATHAN HUNT, IVO DANIHELKA, TIM HARLEY, Google DeepMind, and ANDREW SENIOR, GREG WAYNE, ALEX GRAVES, TIMOTHY LILICRAP, Google DeepMind	3628
Matching Networks for One Shot Learning, ORIOL VINYALS, CHARLES BLUNDELL, DeepMind, TIMOTHY LILICRAP, KORAY KAVUKCUOGLU, and DAAN WIERSTRA, Google DeepMind	3637
Tight Complexity Bounds for Optimizing Composite Objectives, BLAKE WOODWORTH, Toyota Technological Institute, NATI SREBRO, TTI-Chicago	3646
Graphical Time Warping for Joint Alignment of Multiple Curves, YIZHI WANG, Virginia Tech, DAVID MILLER, The Pennsylvania State University, KIRA POSKANZER, University of California San Francisco, YUE WANG, Virginia Tech, and LIN TIAN, The University of California Davis, GUOQIANG YU, Virginia Tech	3655

Unsupervised Risk Estimation Using Only Conditional Independence Structure , JACOB STEINHARDT, PERCY LIANG, Stanford University	3664
MetaGrad: Multiple Learning Rates in Online Learning , TIM ERVEN, WOUTER KOOLEN	3673
Hierarchical Deep Reinforcement Learning: Integrating Temporal Abstraction and Intrinsic Motivation , TEJAS KULKARNI, KARTHIK NARASIMHAN, ARDAVAN SAEEDI, JOSH TENENBAUM, Massachusetts Institute of Technology	3682
High Dimensional Structured Superposition Models , QILONG GU, ARINDAM BANERJEE, University of Minnesota	3691
Joint quantile regression in vector-valued RKHSs , MAXIME SANGNIER, LTCI CNRS Télécom ParisTech, OLIVIER FERCOQ, FLORENCE D'ALCHÉ-BUC	3700
The Forget-me-not Process , KIERAN MILAN, Google DeepMind, JOEL VENESS, JAMES KIRKPATRICK, Google DeepMind, MICHAEL BOWLING, and ANNA KOOP, University of Alberta, DEMIS HASSABIS	3709
Wasserstein Training of Restricted Boltzmann Machines , GREGOIRE MONTAVON, KLAUS-ROBERT MÜLLER, MARCO CUTURI	3718
Communication-Optimal Distributed Clustering , JIECAO CHEN, Indiana University Bloomington, HE SUN, The University of Bristol, DAVID WOODRUFF, QIN ZHANG, Indiana University	3727
Probing the Compositionality of Intuitive Functions , ERIC SCHULZ, University College London, JOSH TENENBAUM, Massachusetts Institute of Technology, DAVID DUVENAUD, Harvard University, MAARTEN SPEEKENBRINK, University College London, and SAM GERSHMAN, Harvard University	3736
Ladder Variational Autoencoders , CASPER SØNDERBY, University of Copenhagen, TAPANI RAIKO, LARS MAALØE, Technical University of Denmark, SØREN SØNDERBY, KU, and OLE WINTHER, Technical University of Denmark	3745
The Multiple Quantile Graphical Model , ALNUR ALI, ZICO KOLTER, Carnegie Mellon University, RYAN TIBSHIRANI	3754
Threshold Learning for Optimal Decision Making , NATHAN LEPORA, University of Bristol	3763
Unsupervised Feature Extraction by Time-Contrastive Learning and Nonlinear ICA , AAPO HYVÄRINEN, HIROSHI MORIOKA, University of Helsinki	3772
Can Active Memory Replace Attention? , ŁUKASZ KAISER, SAMY BENGIO	3781
Minimax Optimal Alternating Minimization for Kernel Nonparametric Tensor Learning , TAJI SUZUKI, HEISHIRO KANAGAWA, HAYATO KOBAYASHI, NOBUYUKI SHIMIZU, and YUKIHIRO TAGAMI	3790
The Product Cut , THOMAS LAURENT, Loyola Marymount University, JAMES BRECHT, CSULB, XAVIER BRESSON, ARTHUR SZLAM	3799

Learning Sparse Gaussian Graphical Models with Overlapping Blocks, MOHAMMAD HOSSEINI, SU-IN LEE, University of Washington	3808
Yggdrasil: An Optimized System for Training Deep Decision Trees at Scale, FIRAS ABUZOID, Massachusetts Institute of Technology, JOSEPH BRADLEY, Databricks, FEYNMAN LIANG, Cambridge University Engineering Department, ANDREW FENG, and LEE YANG, Yahoo!, MATEI ZAHARIA, Massachusetts Institute of Technology, AMEET TALWALKAR, University of California Los Angeles	3817
Average-case hardness of RIP certification, TENG YAO WANG, University of Cambridge, QUENTIN BERTHET, YANIV PLAN, University of British Columbia	3826
A forward model at Purkinje cell synapses facilitates cerebellar anticipatory control, IVAN HERREROS-ALONSO, Universitat Pompeu Fabra, XERXES ARSIWALLA, PAUL VERSCHURE	3835
Convolutional Neural Networks on Graphs with Fast Localized Spectral Filtering, MICHAËL DEFFERRARD, École polytechnique fédérale de Lausanne, XAVIER BRESSON, PIERRE VANDERGHEYNST, École polytechnique fédérale de Lausanne	3844
CliqueCNN: Deep Unsupervised Exemplar Learning, MIGUEL BAUTISTA, ARTSIOM SANAKOYEU, EKATERINA TIKHONCHEVA, Heidelberg University, BJÖRN OMMER	3853
Large-Scale Price Optimization via Network Flow, SHINJI ITO, NEC Corporation, RYOHEI FUJIMAKI	3862
Online Pricing with Strategic and Patient Buyers, MICHAL FELDMAN, TAU, TOMER KOREN, Technion—Israel Inst. of Technology, ROI LIVNI, Huji, YISHAY MANSOUR, Microsoft, and AVIV ZOHAR, huji	3871
Global Optimality of Local Search for Low Rank Matrix Recovery, SRINADH BHOJANAPALLI, TTI Chicago, BEHNAM NEYSHABUR, NATI SREBRO, TTI-Chicago	3880
Phased LSTM: Accelerating Recurrent Network Training for Long or Event-based Sequences, DANIEL NEIL, MICHAEL PFEIFFER, Institute of Neuroinformatics, SHIH-CHII LIU	3889
Improving PAC Exploration Using the Median Of Means, JASON PAZIS, Massachusetts Institute of Technology, RONALD PARR, JONATHAN HOW, Massachusetts Institute of Technology	3898
Infinite Hidden Semi-Markov Modulated Interaction Point Process, MATT ZHANG, Nicta, PENG LIN, TING GUO, Data61, YANG WANG, and FANG CHEN, Data61 CSIRO	3907
Cooperative Inverse Reinforcement Learning, DYLAN HADFIELD-MENELL, STUART RUSSELL, UC Berkeley, PIETER ABBEEL, OpenAI / UC Berkeley / Gradescope, ANCA DRAGAN, UC Berkeley	3916
Spatio-Temporal Hilbert Maps for Continuous Occupancy Representation in Dynamic Environments, RANSALU SENANAYAKE, LIONEL OTT, The University of Sydney, SIMON O'CALLAGHAN, NICTA, FABIO RAMOS, The University of Sydney	3925
Select-and-Sample for Spike-and-Slab Sparse Coding, ABDUL-SABOOR SHEIKH, SAP Labs Berlin, JÖRG LÜCKE	3934

Tractable Operations for Arithmetic Circuits of Probabilistic Models, YUJIA SHEN, ARTHUR CHOI, ADNAN DARWICHE, University of California Los Angeles	3943
Greedy Feature Construction, DINO OGLIC, University of Bonn, THOMAS GAERTNER, The University of Nottingham	3952
Mistake Bounds for Binary Matrix Completion, MARK HERBSTER, STEPHEN PASTERIS, University College London, MASSIMILIANO PONTIL	3961
Data driven estimation of Laplace-Beltrami operator, FREDERIC CHAZAL, INRIA, ILARIA GIULINI, BERTRAND MICHEL	3970
Tracking the Best Expert in Non-stationary Stochastic Environments, CHEN-YU WEI, YI-TE HONG, CHI-JEN LU, Academia Sinica	3979
Learning to learn by gradient descent by gradient descent, MARCIN ANDRYCHOWICZ, Google Deepmind, MISHA DENIL, SERGIO GOMEZ, MATTHEW HOFFMAN, and DAVID PFAU, Google DeepMind, TOM SCHAUL, NANDO FREITAS, Google	3988
Kernel Observers: Systems-Theoretic Modeling and Inference of Spatiotemporally Evolving Processes, HASSAN KINGRAVI, Pindrop Security Services, HARSHAL MASKE, GIRISH CHOWDHARY, University of Illinois at Urbana-Champaign	3997
Quantum Perceptron Models, ASHISH KAPOOR, NATHAN WIEBE, Microsoft Research, KRYSTA SVORE	4006
Guided Policy Search via Approximate Mirror Descent, WILLIAM MONTGOMERY, SERGEY LEVINE, University of Washington	4015
The Power of Optimization from Samples, ERIC BALKANSKI, Harvard University, AVIAD RUBINSTEIN, UC Berkeley, YARON SINGER, Harvard University	4024
Deep Exploration via Bootstrapped DQN, IAN OSBAND, CHARLES BLUNDELL, DeepMind, ALEXANDER PRITZEL, BENJAMIN VAN ROY, Stanford University	4033
A Multi-step Inertial Forward-Backward Splitting Method for Non-convex Optimization, JINGWEI LIANG, GREYC ENSICAEN, JALAL FADILI, GABRIEL PEYRÉ	4042
Scaling Factorial Hidden Markov Models: Stochastic Variational Inference without Messages, YIN NG, PAWEL CHILINSKI, RICARDO SILVA, University College London	4051
Convolutional Neural Fabrics, SHREYAS SAXENA, INRIA, JAKOB VERBEEK	4060
Adaptive Newton Method for Empirical Risk Minimization to Statistical Accuracy, ARYAN MOKHTARI, University of Pennsylvania, HADI DANESHMAND, Eidgenössische Technische Hochschule Zürich, AURELIEN LUCCHI, THOMAS HOFMANN, and ALEJANDRO RIBEIRO, University of Pennsylvania	4069
A Sparse Interactive Model for Matrix Completion with Side Information, JIN LU, GUANNAN LIANG, JIANGWEN SUN, JINBO BI, University of Connecticut	4078

Coresets for Scalable Bayesian Logistic Regression, JONATHAN HUGGINS, TREVOR CAMPBELL, TAMARA BRODERICK, Massachusetts Institute of Technology	4087
Agnostic Estimation for Misspecified Phase Retrieval Models, MATEY NEYKOV, ZHAORAN WANG, HAN LIU, Princeton University	4096
Linear Relaxations for Finding Diverse Elements in Metric Spaces, ADITYA BHASKARA, University of Utah, MEHRDAD GHADIRI, Sharif University of Technology, VAHAB MIRROKNI, Google, OLA SVENSSON, École polytechnique fédérale de Lausanne	4105
Binarized Neural Networks, ITAY HUBARA, Technion, MATTHIEU COURBARIAUX, Université de Montréal, DANIEL SOUDRY, Columbia University, RAN EL-YANIV, Technion, and YOSHUA BENGIO, Université de Montréal	4114
Local Maxima in the Likelihood of Gaussian Mixture Models: Structural Results and Algorithmic Consequences, CHI JIN, YUCHEN ZHANG, UC Berkeley, SIVARAMAN BALAKRISHNAN, Carnegie Mellon University, MARTIN WAINWRIGHT, UC Berkeley, and MICHAEL JORDAN, OpenAI / UC Berkeley / Gradescope	4123
Memory-Efficient Backpropagation Through Time, AUDRUNAS GRUSLYS, REMI MUNOS, Google DeepMind, IVO DANIHELKA, MARC LANCTOT, Google DeepMind, and ALEX GRAVES	4132
Bayesian Optimization with Robust Bayesian Neural Networks, JOST SPRINGENBERG, AARON KLEIN, STEFAN FALKNER, FRANK HUTTER, University of Freiburg	4141
Learnable Visual Markers, OLEG GRINCHUK, VADIM LEBEDEV, Skolkovo Institute of Science and Technology, VICTOR LEMPITSKY	4150
Fast Algorithms for Robust PCA via Gradient Descent, XINYANG YI, UT Austin, DOHYUNG PARK, University of Texas at Austin, YUDONG CHEN, Cornell University, CONSTANTINE CARAMANIS, UT Austin	4159
One-vs-Each Approximation to Softmax for Scalable Estimation of Probabilities, MICHALIS TITSIAS	4168
Learning Deep Embeddings with Histogram Loss, EVGENIYA USTINOVA, Skoltech, VICTOR LEMPITSKY	4177
Spectral Learning of Dynamic Systems from Nonequilibrium Data, HAO WU, Free University of Berlin, FRANK NOE	4186
Fast Mixing Markov Chains for Strongly Rayleigh Measures, DPPs, and Constrained Sampling, CHENGTAO LI, STEFANIE JEGELKA, SUVRIT SRA, Massachusetts Institute of Technology	4195
Mapping Estimation for Discrete Optimal Transport, MICHAEL PERROT, University of Saint-Etienne laboratoire Hubert Curien, NICOLAS COURTY, RÉMI FLAMARY, AMAURY HABRARD, University of Saint-Etienne Laboratoire Hubert Curien	4204
Batched Gaussian Process Bandit Optimization via Determinantal Point Processes, TARUN KATHURIA, Microsoft Research, AMIT DESHPANDE, PUSHMEET KOHLI	4213
Protein contact prediction from amino acid co-evolution using convolutional networks for graph-valued images, VLADIMIR	

GOLKOV, Technical University of Munich, MARCIN SKWARK, Vanderbilt University, ANTONIJ GOLKOV, University of Augsburg, ALEXEY DOSOVITSKIY, and THOMAS BROX, University of Freiburg, JENS MEILER, Vanderbilt University, DANIEL CREMERS, Technical University of Munich	4222
Linear Feature Encoding for Reinforcement Learning , ZHAO SONG, Duke University, RONALD PARR, XUEJUN LIAO, LAWRENCE CARIN, Duke University	4231
A Minimax Approach to Supervised Learning , FARZAN FARNIA, DAVID TSE, Stanford University	4240
Edge-exchangeable graphs and sparsity , DIANA CAI, University of Chicago, TREVOR CAMPBELL, TAMARA BRODERICK, Massachusetts Institute of Technology	4249
A Locally Adaptive Normal Distribution , GEORGIOS ARVANITIDIS, DTU, LARS HANSEN, SØREN HAUBERG	4258
Completely random measures for modelling block-structured sparse networks , TUE HERLAU, MIKKEL SCHMIDT, DTU, MORTEN MØRUP, Technical University of Denmark	4267
Sparse Support Recovery with Non-smooth Loss Functions , KÉVIN DEGRAUX, Université catholique de Louva, GABRIEL PEYRÉ, CNRS DMA École Normale Supérieure Paris France, JALAL FADILI, LAURENT JACQUES, Université catholique de Louvain	4276
Neurons Equipped with Intrinsic Plasticity Learn Stimulus Intensity Statistics , TRAVIS MONK, University of Oldenburg, CRISTINA SAVIN, IST Austria, JÖRG LÜCKE	4285
Learning values across many orders of magnitude , HADO HASSELT, ARTHUR GUEZ, MATTEO HESSEL, Google DeepMind, VOLODYMYR MNIH, and DAVID SILVER	4294
Adaptive Smoothed Online Multi-Task Learning , KEERTHIRAM MURUGESAN, HANXIAO LIU, JAIME CARBONELL, YIMING YANG, Carnegie Mellon University	4303
Safe Exploration in Finite Markov Decision Processes with Gaussian Processes , MATTEO TURCHETTA, FELIX BERKENKAMP, Eidgenössische Technische Hochschule Zürich, ANDREAS KRAUSE	4312
Probabilistic Linear Multistep Methods , ONUR TEYMUR, Imperial College London, KONSTANTINOS ZYGALAKIS, University of Edinburgh, BEN CALDERHEAD	4321
Stochastic Three-Composite Convex Minimization , ALP YURTSEVER, École polytechnique fédérale de Lausanne, BANG VU, VOLKAN CEVHER	4329
Using Fast Weights to Attend to the Recent Past , JIMMY BA, University of Toronto, GEOFFREY HINTON, Google, VOLODYMYR MNIH, JOEL LEIBO, Google DeepMind, and CATALIN IONESCU, Google	4338
Maximal Sparsity with Deep Networks? , BO XIN, YIZHOU WANG, Peking University, WEN GAO, peking university, BAOYUAN WANG, Microsoft Research, and DAVID WIPF	4347
Man is to Computer Programmer as Woman is to Homemaker? Debiasing Word Embeddings , TOLGA BOLUKBASI, Boston	

University, KAI-WEI CHANG, JAMES ZOU, Harvard University, VENKATESH SALIGRAMA, Boston University, and ADAM KALAI, Microsoft Research	4356
beta-risk: a New Surrogate Risk for Learning from Weakly Labeled Data, VALENTINA ZANTEDESCHI, UJM Saint-Etienne France, RÉMI EMONET, MARC SEBBAN	4365
Learning Additive Exponential Family Graphical Models via $\ell_{2,1}$-norm Regularized M-Estimation, XIAOTONG YUAN, Nanjing University of Informat, PING LI, TONG ZHANG, Rutgers University and Amazon, QINGSHAN LIU, and GUANGCAN LIU, NUIST	4374
Backprop KF: Learning Discriminative Deterministic State Estimators, TUOMAS HAARNOJA, ANURAG AJAY, UC Berkeley, SERGEY LEVINE, University of Washington, PIETER ABBEEL, OpenAI / UC Berkeley / Gradescope	4383
LightRNN: Memory and Computation-Efficient Recurrent Neural Networks, XIANG LI, NJUST, TAO QIN, Microsoft, JIAN YANG, TIE-YAN LIU, Microsoft Research	4392
Fast recovery from a union of subspaces, CHINMAY HEGDE, Iowa State University, PIOTR INDYK, LUDWIG SCHMIDT, Massachusetts Institute of Technology	4401
Incremental Variational Sparse Gaussian Process Regression, CHING-AN CHENG, Georgia Institute of Technolog, BYRON BOOTS, Georgia Institute of Technology	4410
A Consistent Regularization Approach for Structured Prediction, CARLO CILIBERTO, ALESSANDRO RUDI, LORENZO ROSASCO, Massachusetts Institute of Technology	4419
Clustering Signed Networks with the Geometric Mean of Laplacians, PEDRO MERCADO, FRANCESCO TUDISCO, MATTHIAS HEIN, Saarland University	4428
An urn model for majority voting in classification ensembles, VÍCTOR SOTO, Columbia University, ALBERTO SUAREZ, GONZALO MARTÍNEZ-MUÑOZ	4437
Avoiding Imposters and Delinquents: Adversarial Crowdsourcing and Peer Prediction, JACOB STEINHARDT, GREGORY VALIANT, MOSES CHARIKAR, Stanford University	4446
Fast and accurate spike sorting of high-channel count probes with KiloSort, MARIUS PACHITARIU, NICK STEINMETZ, University College London, SHABNAM KADIR, MATTEO CARANDINI, and KENNETH HARRIS, University College London	4455
Combining Adversarial Guarantees and Stochastic Fast Rates in Online Learning, WOUTER KOOLEN, PETER GRUNWALD, CWI, TIM ERVEN	4464
Ancestral Causal Inference, SARA MAGLIACANE, VU University Amsterdam, TOM CLAASSEN, JORIS MOOLJ, Radboud University Nijmegen ...	4473
More Supervision, Less Computation: Statistical-Computational Tradeoffs in Weakly Supervised Learning, XINYANG YI, UT Austin, ZHAORAN WANG, ZHUORAN YANG, Princeton University, CONSTANTINE CARAMANIS, UT Austin, and HAN LIU, Princeton University	4482

Tagger: Deep Unsupervised Perceptual Grouping , KLAUS GREFF, IDSIA, ANTTI RASMUS, MATHIAS BERGLUND, TELE HAO, The Curious AI Company, and JÜRGEN SCHMIDHUBER, HARRI VALPOLA, The Curious AI Company	4491
An Efficient Streaming Algorithm for the Submodular Cover Problem , ASHKAN NOROUZI-FARD, ABBAS BAZZI, École polytechnique fédérale de Lausanne, MARWA HALABI, I, ILIJA BOGUNOVIC, EPFL Lausanne, and YA-PING HSIEH, VOLKAN CEVHER	4500
Interaction Networks for Learning about Objects, Relations and Physics , PETER BATTAGLIA, Google DeepMind, RAZVAN PASCANU, MATTHEW LAI, Google DeepMind, DANILO REZENDE, and KORAY KAVUKCUOGLU, Google DeepMind	4509
Efficient state-space modularization for planning: theory, behavioral and neural signatures , DANIEL MCNAMEE, DANIEL WOLPERT, MATE LENGYEL, University of Cambridge	4518
Provable Efficient Online Matrix Completion via Non-convex Stochastic Gradient Descent , CHI JIN, UC Berkeley, SHAM KAKADE, University of Washington, PRANEETH NETRAPALLI, Microsoft Research	4527
Online Bayesian Moment Matching for Topic Modeling with Unknown Number of Topics , WEI-SHOU HSU, University of Waterloo, PASCAL POUPART	4536
Computing and maximizing influence in linear threshold and triggering models , JUSTIN KHIM, University of Pennsylvania, VARUN JOG, PO-LING LOH, Berkeley	4545
Coevolutionary Latent Feature Processes for Continuous-Time User-Item Interactions , YICHEN WANG, Georgia Tech, NAN DU, Georgia Institute of Technology, RAKSHIT TRIVEDI, Georgia Institute of Technology, LE SONG, Georgia Institute of Technology	4554
Optimal Learning for Multi-pass Stochastic Gradient Methods , JUNHONG LIN, Istituto Italiano di Tecnologia, LORENZO ROSASCO, Massachusetts Institute of Technology	4563
Generative Adversarial Imitation Learning , JONATHAN HO, STEFANO ERMON, Stanford University	4572
Latent Attention For If-Then Program Synthesis , XINYUN CHEN, Shanghai Jiaotong University, CHANG LIU, University of Maryland, RICHARD SHIN, DAWN SONG, UC Berkeley, and MINGCHENG CHEN, University of Illinois Urbana-Champaign	4581
Dual Space Gradient Descent for Online Learning , TRUNG LE, University of Pedagogy Ho Chi Minh city, TU NGUYEN, VU NGUYEN, DINH PHUNG, Deakin University	4590
Riemannian SVRG: Fast Stochastic Optimization on Riemannian Manifolds , HONGYI ZHANG, Massachusetts Institute of Technology, SASHANK REDDI, Carnegie Mellon University, SUVRIT SRA, Massachusetts Institute of Technology	4599
Professor Forcing: A New Algorithm for Training Recurrent Networks , ALEX LAMB, Montreal, ANIRUDH GOYAL, YING ZHANG, SAIZHENG ZHANG, and AARON COURVILLE, University of Montreal, YOSHUA BENGIO, U. Montreal	4608

Learning brain regions via large-scale online structured sparse dictionary learning , ELVIS DOHMATOB, Inria, ARTHUR MENSCH, inria, GAEL VAROQUAUX, BERTRAND THIRION	4617
Efficient Neural Codes under Metabolic Constraints , ZHUO WANG, XUE-XIN WEI, ALAN STOCKER, DANIEL LEE, University of Pennsylvania ...	4626
Approximate maximum entropy principles via Goemans-Williamson with applications to provable variational methods , ANDREJ RISTESKI, YUANZHI LI, Princeton University	4635
Efficient High-Order Interaction-Aware Feature Selection Based on Conditional Mutual Information , ALEXANDER SHISHKIN, ANASTASIA BEZZUBTSEVA, ALEXEY DRUTSA, ILIA SHISHKOV, and EKATERINA GLADKIKH, Yandex, GLEB GUSEV, Yandex LLC, PAVEL SERDYUKOV, Yandex	4644
Bayesian Intermittent Demand Forecasting for Large Inventories , MATTHIAS SEEGER, DAVID SALINAS, VALENTIN FLUNKERT, Amazon	4653
Visual Question Answering with Question Representation Update (QRU) , RUIYU LI, JIAYA JIA, The Chinese University of Hong Kong	4662
Learning Parametric Sparse Models for Image Super-Resolution , YONGBO LI, WEISHENG DONG, XUEMEI XIE, GUANGMING SHI, Xidian University, and XIN LI, WVU, DONGLAI XU, Teesside University	4671
Blazing the trails before beating the path: Sample-efficient Monte-Carlo planning , JEAN-BASTIEN GRILL, MICHAL VALKO, Inria Lille - Nord Europe, REMI MUNOS, Google DeepMind	4680
Asynchronous Parallel Greedy Coordinate Descent , YANG YOU, UC Berkeley, XIANGRU LIAN, University of Rochester, JI LIU, HSIANG-FU YU, and INDERJIT DHILLON, University of Texas at Austin, JAMES DEMMEL, UC Berkeley, CHO-JUI HSIEH, UC-Davis	4689
Iterative Refinement of the Approximate Posterior for Directed Belief Networks , REX HJELM, University of New Mexico, RUSLAN SALAKHUTDINOV, University of Toronto, KYUNGHYUN CHO, University of Montreal, NEBOJSA JOJIC, Microsoft Research, and VINCE CALHOUN, Mind Research Network, JUNYOUNG CHUNG, University of Montreal	4698
Assortment Optimization Under the Mallows model , ANTOINE DESIR, Columbia University, VINEET GOYAL, SRIKANTH JAGABATHULA, DANNY SEGEV	4707
Disease Trajectory Maps , PETER SCHULAM, RAMAN ARORA, Johns Hopkins University	4716
Multistage Campaigning in Social Networks , MEHRDAD FARAJTABAR, Georgia Tech, XIAOJING YE, Georgia State University, SAHAR HARATI, Emory University, LE SONG, and HONGYUAN ZHA, Georgia Institute of Technology	4725
Learning in Games: Robustness of Fast Convergence , DYLAN FOSTER, Cornell University, ZHIYUAN LI, Tsinghua University, THODORIS LYKOURIS, Cornell University, KARTHIK SRIDHARAN, University of Pennsylvania, and EVA TARDOS, Cornell University	4734
Improving Variational Autoencoders with Inverse Autoregressive Flow , DIEDERIK KINGMA, TIM SALIMANS, RAFAL JOZEFOWICZ, OpenAI, XI CHEN, UC Berkeley and OpenAI, and ILYA SUTSKEVER, MAX WELLING .	4743

Algorithms and matching lower bounds for approximately-convex optimization , ANDREJ RISTESKI, YUANZHI LI, Princeton University	4752
Unified Methods for Exploiting Piecewise Linear Structure in Convex Optimization , TYLER JOHNSON, CARLOS GUESTRIN, University of Washington	4761
Kernel Bayesian Inference with Posterior Regularization , YANG SONG, Stanford University, JUNE ZHU, YONG REN, Tsinghua University	4770
Neural Universal Discrete Denoiser , TAESUP MOON, DGIST, SEONWOO MIN, BYUNGHAN LEE, SUNGROH YOON	4779
Optimal Architectures in a Solvable Model of Deep Networks , JONATHAN KADMON, Hebrew University, HAIM SOMPOLINSKY	4788
Conditional Image Generation with PixelCNN Decoders , AARON OORD, Google Deepmind, NAL KALCHBRENNER, ORIOL VINYALS, LASSE ESPEHOLT, and ALEX GRAVES, KORAY KAVUKCUOGLU, Google DeepMind	4797
Supervised Learning with Tensor Networks , EDWIN STOUDENMIRE, Univ of California Irvine, DAVID SCHWAB, Northwestern University	4806
Multi-step learning and underlying structure in statistical models , MAIA FRASER, University of Ottawa	4815
Structure-Blind Signal Recovery , DMITRY OSTROVSKY, Univ. Grenoble Alpes, ZAID HARCHAOUI, NYU Courant Institute, ANATOLI JUDITSKY, ARKADI NEMIROVSKI, Gerogia Institute of Technology	4824
An Architecture for Deep, Hierarchical Generative Models , PHILIP BACHMAN	4833
Feature selection in functional data classification with recursive maxima hunting , JOSÉ TORRECILLA, Universidad Autónoma de Madrid, ALBERTO SUAREZ	4842
Achieving budget-optimality with adaptive schemes in crowdsourcing , ASHISH KHETAN, University of Illinois Urbana-, SEWOONG OH, University of Illinois at Urbana-Champaign	4851
Near-Optimal Smoothing of Structured Conditional Probability Matrices , MOEIN FALAHATGAR, University of California San Diego, MESROB OHANNESSIAN, ALON ORLITSKY, UC San Diego	4860
Supervised Word Mover’s Distance , GAO HUANG, CHUAN GUO, MATT KUSNER, YU SUN, and KILIAN WEINBERGER, Cornell University, FEI SHA, University of Southern California	4869
Exploiting Tradeoffs for Exact Recovery in Heterogeneous Stochastic Block Models , AMIN JALALI, QIYANG HAN, IOANA DUMITRIU, MARYAM FAZEL, University of Washington	4878
Full-Capacity Unitary Recurrent Neural Networks , SCOTT WISDOM, THOMAS POWERS, University of Washington, JOHN HERSHEY, JONATHAN LEROUX, and LES ATLAS, University of Washington	4887
Threshold Bandits, With and Without Censored Feedback , JACOB ABERNETHY, KAREEM AMIN, University of Michigan, RUIHAO ZHU, Massachusetts Institute of Technology	4896

Understanding the Effective Receptive Field in Deep Convolutional Neural Networks, WENJIE LUO, YUJIA LI, RAQUEL URTASUN, RICHARD ZEMEL, University of Toronto	4905
Learning Supervised PageRank with Gradient-Based and Gradient-Free Optimization Methods, LEV BOGOLUBSKY, PAVEL DVURECHENSKY, Weierstrass Institute for Appl, ALEXANDER GASNIKOV, GLEB GUSEV, Yandex LLC, and YURI NESTEROV, ANDREY RAIGORODSKII, ALEKSEY TIKHONOV, MAKSIM ZHUKOVSKII	4914
k*-Nearest Neighbors: From Global to Local, OREN ANAVA, Kfir LEVY, Technion	4923
Normalized Spectral Map Synchronization, YANYAO SHEN, UT Austin, QIXING HUANG, Toyota Technological Institute at Chicago, NATI SREBRO, TTI-Chicago, SUJAY SANGHAVI	4932
Beyond Exchangeability: The Chinese Voting Process, MOONTAE LEE, SEOK JIN, DAVID MIMNO, Cornell University	4941
A posteriori error bounds for joint matrix decomposition problems, NICOLO COLOMBO, University College London, NIKOS VLASSIS, Adobe Research	4950
A Bayesian method for reducing bias in neural representational similarity analysis, MING CAI, Princeton University, NICOLAS SCHUCK, Princeton Neuroscience Institute Princeton University, JONATHAN PILLOW, YAEL NIV, Princeton University	4958
Online ICA: Understanding Global Dynamics of Nonconvex Optimization via Diffusion Processes, CHRIS LI, ZHAORAN WANG, HAN LIU, Princeton University	4967
Following the Leader and Fast Rates in Linear Prediction: Curved Constraint Sets and Other Regularities, RUITONG HUANG, University of Alberta, TOR LATTIMORE, ANDRÁS GYÖRGY, CSABA SZEPESVARI, U. Alberta	4976
SDP Relaxation with Randomized Rounding for Energy Disaggregation, KIARASH SHALOUDEGI, ANDRÁS GYÖRGY, CSABA SZEPESVARI, U. Alberta, WILSON XU, University of Alberta	4985
Recovery Guarantee of Non-negative Matrix Factorization via Alternating Updates, YUANZHI LI, YINGYU LIANG, ANDREJ RISTESKI, Princeton University	4994
Unsupervised Learning of 3D Structure from Images, DANILO REZENDE, S. ESLAMI, SHAKIR MOHAMED, PETER BATTAGLIA, Google DeepMind, and MAX JADERBERG, NICOLAS HEES	5003
Poisson–Gamma dynamical systems, AARON SCHEIN, UMass Amherst, HANNA WALLACH, Microsoft Research, MINGYUAN ZHOU	5012
Gaussian Processes for Survival Analysis, TAMARA FERNANDEZ, University of Oxford, NICOLAS RIVERA, King’s College London, YEE WHY TEH	5021
Dual Decomposed Learning with Factorwise Oracle for Structural SVM of Large Output Domain, IAN YEN, HUANG XIANGRU, KAI ZHONG, ZHANG RUOHAN, and PRADEEP RAVIKUMAR, INDERJIT DHILLON, University of Texas at Austin	5030

Optimal Binary Classifier Aggregation for General Losses, AKSHAY BALSUBRAMANI, YOAV FREUND, UC San Diego	5039
Disentangling factors of variation in deep representation using adversarial training, MICHAEL MATHIEU, JUNBO ZHAO, ADITYA RAMESH, PABLO SPRECHMANN, and YANN LECUN, New York University ...	5047
A primal-dual method for conic constrained distributed optimization problems, NECDET AYBAT, ERFAN HAMEDANI, Penn State University	5056
Fundamental Limits of Budget-Fidelity Trade-off in Label Crowdsourcing, FARSHAD , BABAK HASSIBI, Caltech	5065
Learning to Poke by Poking: Experiential Learning of Intuitive Physics, PULKIT AGRAWAL, ASHVIN NAIR, UC Berkeley, PIETER ABBEEL, OpenAI / UC Berkeley / Gradescope, JITENDRA MALIK, UC Berkeley, and SERGEY LEVINE, University of Washington	5074
A Neural Transducer, NAVDEEP JAITLEY, QUOC LE, ORIOL VINYALS, ILYA SUTSKEVER, and DAVID SUSSILLO, Google, SAMY BENGIO	5083
Learning Deep Parsimonious Representations, RENJIE LIAO, UofT, ALEXANDER SCHWING, RICHARD ZEMEL, RAQUEL URTASUN, University of Toronto	5092
Author Index	5101