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 xlvii

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, Zequn Jie, National Univ of Singapore, Xiaodan Liang, Sun Yat-sen University, Jiashi Feng, National University of Singapo, Xiaojie Jin, National University of Singapore, and Wen Lu, National Univ of Singapore, Shuicheng Yan, National University of Singapo 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, Wittawat Jitkrittum, 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 Tschiatschek, 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, Dhruv 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 Savchynskyy 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 Thibault 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, Duy 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, CentraleSupelec, 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, Ambiata 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 Emiya, 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 Poczos, Carnegie Mellon University 1000

Linear-Memory and Decomposition-Invariant Linearly Convergent Conditional Gradient Algorithm for Structured Polytopes, Dan Garber, Ofer Meshi, TTI-Chicago 1009

Efficient Nonparametric Smoothness Estimation, Shashank Singh, Simon Du, Barnabas Poczos, 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 Nøkland, 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 Poczos, Carnegie Mellon University, Alex Smola 1153

Variance Reduction in Stochastic Gradient Langevin Dynamics, Kumar Dubey, Sashank Reddi, Carnegie Mellon University, Sinead Williamson, Barnabas Poczos, 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 Poczos, 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 Ramamohan, 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 Warick, 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, Argonne 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 Jaitly, 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 Poczos, 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, Yuhuai Wu, University of Toronto, Tong Che, IHES, Zhouhan 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 Vorobeychik, 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 Roossien, 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 Averio 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 Holtmann-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 Pittsburg 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, Dhruv 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 Borgwardt 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, Huishuai 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, Yuhuai 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 Liftable 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 Szepesvari, 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 Gaifman 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, Yuhuai 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 Technolog, 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 Łącki, 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 Lillicrap, Google DeepMind 3628

Matching Networks for One Shot Learning, Oriol Vinyals, Charles Blundell, DeepMind, Timothy Lillicrap, 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, Taiji 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 Abuzaid, 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, Tengyao 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 Coorporation, 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 Technolog, 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 Mooij, 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, l, 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 Technolo, 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 Yurii 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, Wilsun 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 Heess 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 Whye 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 Jaitly, 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