

Advances in Neural Information Processing Systems 30

31st Annual Conference on Neural
Information Processing Systems
(NIPS 2017)

Long Beach, California, USA
4-9 December 2017

Volume 1 of 10

Editors:

Ulrike von Luxburg
Samy Bengio
Rob Fergus
Roman Garnett

Isabelle Guyon
Hanna Wallach
S.V.N. Vishwanathan

ISBN: 978-1-5108-6096-4

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© (2017) by individual authors and NIPS
All rights reserved.

Printed by Curran Associates, Inc. (2018)

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	xlix
Donors	lv
NIPS foundation	lviii
Committees	lx
Reviewers	lxv
Wider and Deeper, Cheaper and Faster: Tensorized LSTMs for Sequence Learning , ZHEN HE, University College London, and SHAOBING GAO, Sichuan University, LIANG XIAO, DAXUE LIU, HANGEN HE, National University of Defense Technology, DAVID BARBER, University College London	1
Concentration of Multilinear Functions of the Ising Model with Applications to Network Data , CONSTANTINOS DASKALAKIS, and NISHANTH DIKKALA, GAUTAM KAMATH, Massachusetts Institute of Technology	12
Deep Subspace Clustering Networks , PAN JI, University of Adelaide, and TONG ZHANG, HONGDONG LI, Australian National University, MATHIEU SALZMANN, École polytechnique fédérale de Lausanne, IAN REID, The University of Adelaide	24
Attentional Pooling for Action Recognition , ROHIT GIRDHAR, and DEVA RAMANAN, Carnegie Mellon University	34
On the Consistency of Quick Shift , HEINRICH JIANG, Google	46
Breaking the Nonsmooth Barrier: A Scalable Parallel Method for Composite Optimization , FABIAN PEDREGOSA, and REMI LEBLOND, INRIA, SIMON LACOSTE-JULIEN, Université de Montréal	56
Dual-Agent GANs for Photorealistic and Identity Preserving Profile Face Synthesis , JIAN ZHAO, National University of Singapore, and LIN XIONG, Panasonic R&D Center Singapore, KARLEKAR JAYASHREE, Panasonic Singapore, JIANSHU LI, National University of Singapore, FANG ZHAO, National University of Singapore, ZHECAN WANG, Franklin. W. Olin College of Engineering, SUGIRI PRANATA, SHENGMEI SHEN, Panasonic Singapore, SHUICHENG YAN, JIASHI FENG, National University of Singapore	66
Dilated Recurrent Neural Networks , SHIYU CHANG, IBM T.J. Watson Research Center, and YANG ZHANG, WEI HAN, University of Illinois at Urbana-Champaign, MO YU, IBM Research, XIAOXIAO GUO, WEI TAN, Citadel LLC, XIAODONG CUI, MICHAEL WITBROCK, IBM T. J. Watson	

Research Center, MARK HASEGAWA-JOHNSON, THOMAS HUANG, University of Illinois at Urbana-Champaign	77
Hunt For The Unique, Stable, Sparse And Fast Feature Learning On Graphs , SAURABH VERMA, and ZHI-LI ZHANG, University of Minnesota . . .	88
Scalable Generalized Linear Bandits: Online Computation and Hashing , KWANG-SUNG JUN, and ANIRUDDHA BHARGAVA, UW-Madison, ROBERT NOWAK, University of Wisconsin-Madison, REBECCA WILLETT, UW-Madison	99
Probabilistic Models for Integration Error in the Assessment of Functional Cardiac Models , CHRIS OATES, Newcastle University, and STEVEN NIEDERER, Kings College London, ANGELA LEE, King's College London, FRANCOIS-XAVIER BRIOL, University of Warwick, MARK GIROLAMI, Imperial College London	110
Machine Learning with Adversaries: Byzantine Tolerant Gradient Descent , PEVA BLANCHARD, and EL MHAMDI, École polytechnique fédérale de Lausanne, RACHID GUERRAOUI, JULIEN STAINER	119
Dynamic Safe Interruptibility for Decentralized Multi-Agent Reinforcement Learning , EL MHAMDI, École polytechnique fédérale de Lausanne, and RACHID GUERRAOUI, HADRIEN HENDRIKX, École polytechnique fédérale de Lausanne, ALEXANDRE MAURER	130
Interactive Submodular Bandit , LIN CHEN, Yale University, and ANDREAS KRAUSE, AMIN KARBASI, Yale University	141
Learning to See Physics via Visual De-animation , JIAJUN WU, and ERIKA LU, Massachusetts Institute of Technology, PUSHMEET KOHLI, WILLIAM FREEMAN, JOSH TENENBAUM, Massachusetts Institute of Technology	153
Label Efficient Learning of Transferable Representations across Domains and Tasks , ZELUN LUO, Stanford University, and YULIANG ZOU, University of Michigan, JUDY HOFFMAN, FEI-FEI LI, Stanford University	165
Decoding with Value Networks for Neural Machine Translation , DI HE, Microsoft Research Asia, and HANQING LU, Carnegie Mellon University, YINGCE XIA, University of Science and Technology of China, TAO QIN, Microsoft Research, LIWEI WANG, TIEYAN LIU	178
Parametric Simplex Method for Sparse Learning , HAOTIAN PANG, and HAN LIU, ROBERT VANDERBEI, Princeton University, TUO ZHAO, Georgia Tech	188
Group Sparse Additive Machine , HONG CHEN, and XIAOQIAN WANG, University of Texas at Arlington, CHENG DENG, School of Electronic Engineering Xidian University China, HENG HUANG, University of Pittsburgh ...	198
Uprooting and Rerooting Higher-Order Graphical Models , MARK ROWLAND, and ADRIAN WELLER, University of Cambridge	209
The Unreasonable Effectiveness of Structured Random Orthogonal Embeddings , KRZYSZTOF CHOROMANSKI, and MARK ROWLAND, ADRIAN WELLER, University of Cambridge	219
From Parity to Preference-based Notions of Fairness in Classification , MUHAMMAD ZAFAR, MPI-SWS, and ISABEL VALERA,	

MPI for Intelligent Systems, MANUEL RODRIGUEZ, KRISHNA GUMMADI, MPI-SWS, ADRIAN WELLER, University of Cambridge	229
Inferring Generative Model Structure with Static Analysis, PAROMA VARMA, and BRYAN HE, PAYAL BAJAJ, NISHITH KHANDWALA, IMON BANERJEE, DANIEL RUBIN, CHRISTOPHER RE, Stanford University	240
Structured Embedding Models for Grouped Data, MAJA RUDOLPH, and FRANCISCO RUIZ, Columbia University, SUSAN ATHEY, Stanford University, DAVID BLEI, Columbia University	251
A Linear-Time Kernel Goodness-of-Fit Test, WITTAWAT JITKRITTUM, and WENKAI XU, Gatsby Unit UCL, ZOLTAN SZABO, Ecole Polytechnique, KENJI FUKUMIZU, ARTHUR GRETTON	262
Cortical microcircuits as gated-recurrent neural networks, RUI COSTA, and YANNIS ASSAEL, BRENDAN SHILLINGFORD, NANDO FREITAS, TIM VOGELS, University of Oxford	272
k-Support and Ordered Weighted Sparsity for Overlapping Groups: Hardness and Algorithms, CONG LIM, UW-Madison, and STEPHEN WRIGHT, University of Wisconsin-Madison	284
A simple model of recognition and recall memory, NISHEETH SRIVASTAVA, IIT Kanpur, and EDWARD VUL, University of California San Diego	293
On Structured Prediction Theory with Calibrated Convex Surrogate Losses, ANTON OSOKIN, CS HSE, and FRANCIS BACH, SAC, SIMON LACOSTE-JULIEN, Université de Montréal	302
Best of Both Worlds: Transferring Knowledge from Discriminative Learning to a Generative Visual Dialog Model, JIASEN LU, Georgia Tech, and ANITHA KANNAN, JIANWEI YANG, DEVI PARIKH, DHRUV BATRA, Georgia Tech	314
MaskRNN: Instance Level Video Object Segmentation, YUAN-TING HU, University of Illinois at Urbana-Champaign, and JIA-BIN HUANG, Virginia Tech, ALEXANDER SCHWING, University of Illinois at Urbana-Champaign	325
Gated Recurrent Convolution Neural Network for OCR, JIANFENG WANG, BUPT, and XIAOLIN HU, Tsinghua University	335
Towards Accurate Binary Convolutional Neural Network, XIAOFAN LIN, and CONG ZHAO, WEI PAN, DJI	345
Semi-Supervised Learning for Optical Flow with Generative Adversarial Networks, WEI-SHENG LAI, UC Merced, and JIA-BIN HUANG, Virginia Tech, MING-HSUAN YANG, UC Merced	354
Learning a Multi-View Stereo Machine, ABHISHEK KAR, and CHRISTIAN HÄNE, JITENDRA MALIK, UC Berkeley	365
Phase Transitions in the Pooled Data Problem, JONATHAN SCARLETT, Ecole Polytechnique Federale De Lausanne, and VOLKAN CEVHER .	377
Universal Style Transfer via Feature Transforms, YIJUN LI, University of California Merc, and CHEN FANG, Adobe Research, JIMEI YANG, ZHAOWEN WANG, Adobe Research, XIN LU, Adobe, MING-HSUAN YANG, UC Merced	386

On the Model Shrinkage Effect of Gamma Process Edge Partition Models, IKU OHAMA, Hokkaido University, and ISSEI SATO, The University of Tokyo/RIKEN, TAKUYA KIDA, HIROKI ARIMURA, Hokkaido University	397
Pose Guided Person Image Generation, LIQIAN MA, and XU JIA, KU Leuven, QIANRU SUN, MPI, BERNT SCHIELE, TINNE TUYTELAARS, LUC GOOL, KU Leuven	406
Inference in Graphical Models via Semidefinite Programming Hierarchies , MURAT ERDOGDU, Microsoft Research, and YASH DESHPANDE, Massachusetts Institute of Technology, ANDREA MONTANARI, Stanford University	417
Variable Importance Using Decision Trees, SAYED KAZEMITABAR, and ARASH AMINI, University of California Los Angeles, ADAM BLONIARZ, Google, AMEET TALWALKAR, University of California Los Angeles	426
Preventing Gradient Explosions in Gated Recurrent Units, SEKITOSHI KANAI, and YASUHIRO FUJIWARA, NTT Software Innovation Center, SOTETSU IWAMURA, NTT Software Innovation center	436
On the Power of Truncated SVD for General High-rank Matrix Estimation Problems, SIMON DU, and YINING WANG, AARTI SINGH, Carnegie Mellon University	446
f-GANs in an Information Geometric Nutshell, RICHARD NOCK, Data61 and ANU, and ZAC CRANKO, Australian National University, ADITYA MENON, LIZHEN QU, BOB WILLIAMSON, Australian National University	457
Toward Multimodal Image-to-Image Translation , JUN-YAN ZHU, and RICHARD ZHANG, DEEPAK PATHAK, UC Berkeley, TREVOR DARRELL, Berkeley, ALEXEI EFROS, UC Berkeley, OLIVER WANG, ELI SHECHTMAN	466
Mixture-Rank Matrix Approximation for Collaborative Filtering, DONGSHENG LI, and CHAO CHEN, IBM Research - China, WEI LIU, Tencent AI Lab, TUN LU, NING GU, Fudan University, STEPHEN CHU, IBM Research - China	478
Non-monotone Continuous DR-submodular Maximization: Structure and Algorithms, AN BIAN, and KFIR LEVY, Eidgenössische Technische Hochschule Zürich, ANDREAS KRAUSE, JOACHIM BUHMANN, Eidgenössische Technische Hochschule Zürich	487
Learning with Average Top-k Loss, YANBO FAN, Chinese Academy of Science and University at Albany SUNY, and SIWEI LYU, SUNY Albany, YIMING YING, University at Albany SUNY, BAOGANG HU, Chinese Academy of Sciences	498
Learning multiple visual domains with residual adapters, SYLVESTRE-ALVISE REBUFFI, University of Oxford, and HAKAN BILEN, University of Edinburgh, ANDREA VEDALDI, University of Oxford	507
Dijkstra’s Algorithm, ADMM, and Coordinate Descent: Connections, Insights, and Extensions, RYAN TIBSHIRANI	518
Learning Spherical Convolution for Fast Features from 360°Imagery, YU-CHUAN SU, UT Austin, and KRISTEN GRAUMAN, SAC	530
MarrNet: 3D Shape Reconstruction via 2.5D Sketches, JIAJUN WU, Massachusetts Institute of Technology, and YIFAN WANG, ShanghaiTech University, TIANFAN XUE, Massachusetts Institute of Technology, XINGYUAN	

SUN, Shanghai Jiao Tong University, WILLIAM FREEMAN, JOSH TENENBAUM, Massachusetts Institute of Technology	541
Multimodal Learning and Reasoning for Visual Question Answering , ILIJA ILIEVSKI, National University of Singapo, and JIASHI FENG, National University of Singapore	552
Adversarial Surrogate Losses for Ordinal Regression , RIZAL FATHONY, and MOHAMMAD BASHIRI, BRIAN ZIEBART, University of Illinois Chicago	564
Hypothesis Transfer Learning via Transformation Functions , SIMON DU, and JAYANTH KOUSHIK, AARTI SINGH, BARNABAS PO CZOS, Carnegie Mellon University	575
Controllable Invariance through Adversarial Feature Learning , QIZHE XIE, and ZIHANG DAI, YULUN DU, EDUARD HOVY, GRAHAM NEUBIG, Carnegie Mellon University	586
Convergence Analysis of Two-layer Neural Networks with ReLU Activation , YUANZHI LI, Princeton University, and YANG YUAN, Cornell University	598
Doubly Accelerated Stochastic Variance Reduced Dual Averaging Method for Regularized Empirical Risk Minimization , TOMOYA MURATA, NTT DATA Mathematical Systems, and TAJI SUZUKI, University of Tokyo/JST-PRESTO/RIKEN AIP	609
Langevin Dynamics with Continuous Tempering for Training Deep Neural Networks , NANYANG YE, University of Cambridge, and ZHANXING ZHU, Peking University and Beijing Institute of Big Data Research, RAFAL MANTIUK, University of Cambridge	619
Efficient Online Linear Optimization with Approximation Algorithms , DAN GARBER	628
Geometric Descent Method for Convex Composite Minimization , SHIXIANG CHEN, Chinese University of HongKong, and SHIQIAN MA, UC Davis, WEI LIU, Tencent AI Lab	637
Diffusion Approximations for Online Principal Component Estimation and Global Convergence , CHRIS LI, and MENGDI WANG, Princeton University, TONG ZHANG	646
Avoiding Discrimination through Causal Reasoning , NIKI KILBERTUS, MPI-IS Tübingen, and MATEO ROJAS-CARULLA, MPI Tuebingen Cambridge, GIAMBATTISTA PARASCANDOLO, Max Planck Institute for Intelligent Systems and and Max Planck ETH CLS, MORITZ HARDT, DOMINIK JANZING, MPI Tübingen, BERNHARD SCHÖLKOPF, Max Planck Institute for Intelligent Systems	657
Nonparametric Online Regression while Learning the Metric , ILJA KUZBORSKIJ, École polytechnique fédérale de Lausanne, and NICOLÒ CESA-BIANCHI	668
Recycling Privileged Learning and Distribution Matching for Fairness , NOVI QUADRIANTO, University of Sussex and HSE, and VIKTORIIA SHARMANSKA, University of Sussex	678
Safe and Nested Subgame Solving for Imperfect-Information Games , NOAM BROWN, and TUOMAS SANDHOLM, Carnegie Mellon University	690

Unsupervised Image-to-Image Translation Networks , MING-YU LIU, NVIDIA, and THOMAS BREUEL, JAN KAUTZ, NVIDIA	701
Coded Distributed Computing for Inverse Problems , YAOQING YANG, and PULKIT GROVER, SOUMMYA KAR, Carnegie Mellon University	710
A Screening Rule for l1-Regularized Ising Model Estimation , ZHAOBIN KUANG, University of Wisconsin Madison, and SINONG GENG, UW-Madison, DAVID PAGE, University of Wisconsin Madison	721
Improved Dynamic Regret for Non-degenerate Functions , LIJUN ZHANG, and TIANBAO YANG, University of Iowa, JINFENG YI, RONG JIN, Alibaba Group, ZHI-HUA ZHOU, Nanjing University	733
Learning Efficient Object Detection Models with Knowledge Distillation , GUOBIN CHEN, University of Missouri, and WONGUN CHOI, XIANG YU, TONY HAN, University of Missouri, MANMOHAN CHANDRAKER ..	743
One-Sided Unsupervised Domain Mapping , SAGIE BENAÏM, Tel Aviv University, and LIOR WOLF, Facebook AI Research and TAU	753
Deep Mean-Shift Priors for Image Restoration , SIAVASH BIGDELI, and MATTHIAS ZWICKER, PAOLO FAVARO, MEIGUANG JIN, University of Bern	764
Greedy Algorithms for Cone Constrained Optimization with Convergence Guarantees , FRANCESCO LOCATELLO, Eidgenössische Technische Hochschule Zürich, and MICHAEL TSCHANNEN, GUNNAR RAETSCH, Eidgenössische Technische Hochschule Zürich, MARTIN JAGGI	774
A New Theory for Matrix Completion , GUANGCAN LIU, and QINGSHAN LIU, XIAOTONG YUAN, NUIST	786
Robust Hypothesis Test for Nonlinear Effect with Gaussian Processes , JEREMIAH LIU, and BRENT COULL, Harvard University	796
Lower bounds on the robustness to adversarial perturbations , JONATHAN PECK, and JORIS ROELS, BART GOOSSENS, YVAN SAEYS, Ghent University	805
Minimizing a Submodular Function from Samples , ERIC BALKANSKI, Harvard, and YARON SINGER, Harvard University	815
Introspective Classification with Convolutional Nets , LONG JIN, and JUSTIN LAZAROW, ZHUOWEN TU, University of California San Diego	824
Label Distribution Learning Forests , WEI SHEN, and KAI ZHAO, YILU GUO, Shanghai University, ALAN YUILLE, Johns Hopkins University	835
Unsupervised learning of object frames by dense equivariant image labelling , JAMES THEWLIS, University of Oxford, and HAKAN BILEN, University of Edinburgh, ANDREA VEDALDI, University of Oxford	845
Compression-aware Training of Deep Networks , JOSE ALVAREZ, Data61, and MATHIEU SALZMANN, École polytechnique fédérale de Lausanne ...	857
Multiscale Semi-Markov Dynamics for Intracortical Brain-Computer Interfaces , DANIEL MILSTEIN, and JASON PACHECO, Brown University, LEIGH HOCHBERG, Brown MGH VA Harvard, JOHN SIMERAL, Brown University VA, BEATA JAROSIEWICZ, Stanford University, ERIK SUDDERTH, UC Irvine	869

PredRNN: Recurrent Neural Networks for Predictive Learning using Spatiotemporal LSTMs, YUNBO WANG, and MINGSHENG LONG, JIANMIN WANG, ZHIFENG GAO, Tsinghua University, PHILIP YU, University of Illinois Chicago	880
Detrended Partial Cross Correlation for Brain Connectivity Analysis, JAIME IDE, Yale University, and FÁBIO CAPPABIANCO, FABIO FARIA, Federal University of Sao Paulo, CHIANG-SHAN LI, Yale University	890
Contrastive Learning for Image Captioning, BO DAI, and DAHUA LIN, The Chinese University of Hong Kong	899
Safe Model-based Reinforcement Learning with Stability Guarantees, FELIX BERKENKAMP, and MATTEO TURCHETTA, Eidgenössische Technische Hochschule Zürich, ANGELA SCHOELLIG, ANDREAS KRAUSE	909
Online multiclass boosting, YOUNG JUNG, and JACK GOETZ, Statistics department University of Michigan, AMBUJ TEWARI, University of Michigan	920
Matching on Balanced Nonlinear Representations for Treatment Effects Estimation, SHENG LI, and YUN FU, Northeastern University	930
Learning Overcomplete HMMs, VATSAL SHARAN, Stanford University, and SHAM KAKADE, University of Washington, PERCY LIANG, GREGORY VALIANT, Stanford University	941
GP CaKe: Effective brain connectivity with causal kernels, LUCA AMBROGIONI, and MAX HINNE, MARCEL GERVEN, ERIC MARIS, Donders Institute	951
Decoupling "when to update" from "how to update", ERAN MALACH, Hebrew University of Jerusalem, and SHAI SHALEV-SHWARTZ	961
Self-Normalizing Neural Networks, GÜNTER KLAMBAUER, and THOMAS UNTERTHINER, ANDREAS MAYR, Johannes Kepler University Linz, SEPP HOCHREITER, LIT AI Lab / University Linz	972
Learning to Pivot with Adversarial Networks, GILLES LOUPPE, New York University, and MICHAL KAGAN, KYLE CRANMER, New York University	982
SchNet: A continuous-filter convolutional neural network for modeling quantum interactions, KRISTOF SCHÜTT, TU Berlin, and PIETER-JAN KINDERMANS, HUZIEL SAUCEDA, Fritz-Haber-Institut der Max-Planck-Gesellschaft, STEFAN CHMIELA, TU Berlin, ALEXANDRE TKATCHENKO, University of Luxembourg, KLAUS-ROBERT MÜLLER	992
Active Bias: Training More Accurate Neural Networks by Emphasizing High Variance Samples, HAW-SHIUAN CHANG, UMass Amherst, and ERIK LEARNED-MILLER, CICS UMass Amherst, ANDREW MCCALLUM, UMass Amherst	1003
Differentiable Learning of Submodular Functions, JOSIP DJOLONGA, Eidgenössische Technische Hochschule Zürich, and ANDREAS KRAUSE	1014
Inductive Representation Learning on Large Graphs, WILLIAM HAMILTON, and REX YING, JURE LESKOVEC, Stanford University	1025
Subset Selection and Summarization in Sequential Data, EHSAN ELHAMIFAR, and M. KALUZA, Northeastern University	1036

Question Asking as Program Generation , ANSELM ROTHE, and BRENDEN LAKE, TODD GURECKIS, New York University	1047
Revisiting Perceptron: Efficient and Label-Optimal Learning of Halfspaces , SONGBAI YAN, and CHICHENG ZHANG, University of California San Diego	1057
Gradient Descent Can Take Exponential Time to Escape Saddle Points , SIMON DU, Carnegie Mellon University, and CHI JIN, University of California at Berkeley, JASON LEE, University of Southern California, MICHAEL JORDAN, University of California at Berkeley, AARTI SINGH, BARNABAS POCZOS, Carnegie Mellon University	1068
Union of Intersections (UoI) for Interpretable Data Driven Discovery and Prediction , KRISTOFER BOUCHARD, and ALEJANDRO BUJAN, FARBOD ROOSTA-KHORASANI, SHASHANKA UBARU, University of Minnesota, MR PRABHAT, ANTOINE SNIJDERS, JIAN-HUA MAO, EDWARD CHANG, MICHAEL MAHONEY, SHARMODEEP BHATTACHARYA	1079
One-Shot Imitation Learning , ROCKY DUAN, UC Berkeley, and MARCIN ANDRYCHOWICZ, OpenAI, BRADLY STADIE, JONATHAN HO, OpenAI UC Berkeley, JONAS SCHNEIDER, OpenAI, ILYA SUTSKEVER, PIETER ABBEEL, University of California at Berkeley, WOJCIECH ZAREMBA, OpenAI ..	1088
Learning the Morphology of Brain Signals Using Alpha-Stable Convolutional Sparse Coding , MAINAK JAS, Télécom ParisTech, and TOM TOUR, UMUT SIMSEKLI, Telecom ParisTech, ALEXANDRE GRAMFORT, INRIA	1100
Integration Methods and Optimization Algorithms , DAMIEN SCIEUR, INRIA - ENS, and VINCENT ROULET, INRIA-ENS, FRANCIS BACH, SAC, ALEXANDRE D'ASPREMONT, ENS CNRS	1110
Sharpness, Restart and Acceleration , VINCENT ROULET, INRIA-ENS, and ALEXANDRE D'ASPREMONT, ENS CNRS	1120
Learning Koopman Invariant Subspaces for Dynamic Mode Decomposition , NAOYA TAKEISHI, The University of Tokyo, and YOSHINOBU KAWAHARA, Osaka University / RIKEN, TAKEHISA YAIRI, The University of Tokyo	1131
Soft-to-Hard Vector Quantization for End-to-End Learning Compressible Representations , EIRIKUR AGUSTSSON, and FABIAN MENTZER, Eidgenössische Technische Hochschule Zürich, MICHAEL TSCHANNEN, LUKAS CAVIGELLI, RADU TIMOFTE, LUCA BENINI, LUC GOOL, Eidgenössische Technische Hochschule Zürich	1142
Learning spatiotemporal piecewise-geodesic trajectories from longitudinal manifold-valued data , STEPHANIE ALLASSONNIERE, Universite Paris Descartes, and JULIETTE CHEVALLIER, CMAP, STEPHANE OUDARD, HEGP	1153
Improving Regret Bounds for Combinatorial Semi-Bandits with Probabilistically Triggered Arms and Its Applications , QINSHI WANG, Princeton University, and WEI CHEN	1162
Predictive-State Decoders: Encoding the Future into Recurrent Networks , ARUN VENKATRAMAN, and NICHOLAS RHINEHART, WEN SUN, LERREL PINTO, MARTIAL HEBERT, Carnegie Mellon University, BYRON BOOTS, Georgia Tech, KRIS KITANI, J. BAGNELL, Carnegie Mellon University	1173

Optimistic posterior sampling for reinforcement learning: worst-case regret bounds , SHIPRA AGRAWAL, and RANDY JIA, Columbia University	1185
Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results , ANTTI TARVAINEN, and HARRI VALPOLA, The Curious AI Company	1196
Matching neural paths: transfer from recognition to correspondence search , NIKOLAY SAVINOV, and LUBOR LADICKY, MARC POLLEFEYS, Eidgenössische Technische Hochschule Zürich	1206
Linearly constrained Gaussian processes , CARL JIDLING, and NIKLAS WAHLSTRÖM, Uppsala University, ADRIAN WILLS, University of Newcastle Australia, THOMAS SCHON, Uppsala University	1216
Fixed-Rank Approximation of a Positive-Semidefinite Matrix from Streaming Data , JOEL TROPP, Caltech, and ALP YURTSEVER, École polytechnique fédérale de Lausanne, MADELEINE UDELL, Cornell, VOLKAN CEVHER	1226
Multi-Modal Imitation Learning from Unstructured Demonstrations using Generative Adversarial Nets , KAROL HAUSMAN, and YEVGEN CHEBOTAR, STEFAN SCHAAL, GAURAV SUKHATME, JOSEPH LIM, University of Southern California	1236
Learning to Inpaint for Image Compression , MOHAMMAD BAIG, Dartmouth College, and VLADLEN KOLTUN, Intel Labs, LORENZO TORRESANI, Dartmouth College	1247
Adaptive Bayesian Sampling with Monte Carlo EM , ANIRBAN ROYCHOWDHURY, Ohio State University, and SRINIVASAN PARTHASARATHY, The Ohio State University	1257
ADMM without a Fixed Penalty Parameter: Faster Convergence with New Adaptive Penalization , YI XU, The University of Iowa, and MINGRUI LIU, University of Iowa, QIHANG LIN, Univ Iowa faculty, TIANBAO YANG, University of Iowa	1268
Shape and Material from Sound , ZHOUTONG ZHANG, Massachusetts Institute of Technology, and QIUJIA LI, University of Cambridge, ZHENGJIA HUANG, ShanghaiTech, JIAJUN WU, JOSH TENENBAUM, WILLIAM FREEMAN, Massachusetts Institute of Technology	1279
Flexible statistical inference for mechanistic models of neural dynamics , JAN-MATTHIS LUECKMANN, research center caesar, and PEDRO GONCALVES, research center caesar an associate of the Max Planck Society Bonn Germany, GIACOMO BASSETTO, KAAO OECAL, MARCEL NONNENMACHER, research center caesar, JAKOB MACKE	1290
Online Prediction with Selfish Experts , TIM ROUGHGARDEN, and OKKE SCHRIJVERS, Stanford University	1301
Tensor Biclustering , SOHEIL FEIZI, and HAMID JAVADI, DAVID TSE, Stanford University	1312
DPSCREEN: Dynamic Personalized Screening , KARTIK AHUJA, and WILLIAM ZAME, University of California Los Angeles, MIHAELA SCHAAR, University of Oxford	1322

Learning Unknown Markov Decision Processes: A Thompson Sampling Approach, YI OUYANG, University of Southern Califor, and MUKUL GAGRANI, ASHUTOSH NAYYAR, RAHUL JAIN, University of Southern California	1334
Testing and Learning on Distributions with Symmetric Noise Invariance, HO LAW, University Of Oxford, and CHRISTOPHER YAU, DINO SEJDINOVIC, University of Oxford	1344
A Dirichlet Mixture Model of Hawkes Processes for Event Sequence Clustering, HONGTENG XU, Georgia Tech, and HONGYUAN ZHA, Georgia Institute of Technology	1355
Deanonymization in the Bitcoin P2P Network, GIULIA FANTI, University of Illinois at Urba, and PRAMOD VISWANATH, University of Illinois at Urbana-Champaign	1365
Accelerated consensus via Min-Sum Splitting, PATRICK REBESCHINI, University of Oxford, and SEKHAR TATIKONDA, Yale	1375
Generalized Linear Model Regression under Distance-to-set Penalties, JASON XU, University of California Los Angeles, and ERIC CHI, NC State, KENNETH LANGE, University of California Los Angeles	1386
Adaptive stimulus selection for optimizing neural population responses, BENJAMIN COWLEY, and RYAN WILLIAMSON, Carnegie Mellon University, KATERINA CLEMENS, MATTHEW SMITH, University of Pittsburgh, BYRON YU, Carnegie Mellon University	1397
Nonbacktracking Bounds on the Influence in Independent Cascade Models, EMMANUEL ABBE, and SANJEEV KULKARNI, EUN LEE, Princeton University	1408
Learning with Feature Evolvable Streams, BO-JIAN HOU, LAMDA, and LIJUN ZHANG, ZHI-HUA ZHOU, Nanjing University	1418
Online Convex Optimization with Stochastic Constraints, HAO YU, Univ. of Southern California, and MICHAEL NEELY, Univ. Southern California, XIAOHAN WEI, Univ. of Southern California	1429
Max-Margin Invariant Features from Transformed Unlabelled Data, DIPAN PAL, and ASHWIN KANNAN, GAUTAM ARAKALGUD, MARIOS SAVVIDES, Carnegie Mellon University	1440
Regularized Modal Regression with Applications in Cognitive Impairment Prediction, XIAOQIAN WANG, and HONG CHEN, University of Texas at Arlington, WEIDONG CAI, School of Information Technologies University of Sydney Australia, DINGGANG SHEN, UNC, HENG HUANG, University of Pittsburgh	1449
Translation Synchronization via Truncated Least Squares, XIANGRU HUANG, University of Texas at Austin, and ZHENXIAO LIANG, Tsinghua University, CHANDRAJIT BAJAJ, The University of Texas at Austin, QIXING HUANG, University of Texas at Austin	1460
From which world is your graph, CHENG LI, College of William and Mary, and FELIX WONG, ZHENMING LIU, College of William and Mary, VARUN KANADE	1470
A New Alternating Direction Method for Linear Programming, SINONG WANG, and NESS SHROFF, The Ohio State University	1481

Regret Analysis for Continuous Dueling Bandit, WATARU KUMAGAI, RIKEN	1490
Best Response Regression, OMER PORAT, Technion – Israel Institute of, and MOSHE TENNENHOLTZ, Technion - Israel Institute of Technology	1500
TernGrad: Ternary Gradients to Reduce Communication in Distributed Deep Learning, WEI WEN, Duke University, and CONG XU, Hewlett Packard Labs, FENG YAN, University of Nevada Reno, CHUNPENG WU, Duke University, YANDAN WANG, University of Pittsburgh, YIRAN CHEN, HAI LI, Duke University	1510
Learning Affinity via Spatial Propagation Networks, SIFEI LIU, UC Merced, and SHALINI MELLO, JINWEI GU, Nvidia, GUANGYU ZHONG, Dalian University of Technology and UC Merced, MING-HSUAN YANG, UC Merced, JAN KAUTZ, NVIDIA	1521
Linear regression without correspondence, DANIEL HSU, and KEVIN SHI, Columbia University, XIAORUI SUN, UC Berkeley	1532
NeuralFDR: Learning Discovery Thresholds from Hypothesis Features, FEI XIA, and MARTIN ZHANG, Stanford University, JAMES ZOU, DAVID TSE, Stanford University	1542
Cost efficient gradient boosting, SVEN PETER, University Heidelberg, and FERRAN DIEGO, Bosch, FRED HAMPRECHT, BOAZ NADLER	1552
Probabilistic Rule Realization and Selection, HAIZI YU, University of Illinois at Urbana-Champaign, and TIANXI LI, University of Michigan, LAV VARSHNEY, University of Illinois at Urbana-Champaign	1563
Nearest-Neighbor Sample Compression: Efficiency, Consistency, Infinite Dimensions, ARYEH KONTOROVITCH, and SIVAN SABATO, ROI WEISS	1574
A Scale Free Algorithm for Stochastic Bandits with Bounded Kurtosis, TOR LATTIMORE	1585
Learning Multiple Tasks with Multilinear Relationship Networks, MINGSHENG LONG, and ZHANGJIE CAO, JIANMIN WANG, Tsinghua University, PHILIP YU, University of Illinois Chicago	1595
Deep Hyperalignment, MUHAMMAD YOUSEFNEZHAD, Nanjing University of Aeronautics and Astronautics, and DAOQIANG ZHANG	1605
Online to Offline Conversions, Universality and Adaptive Minibatch Sizes, KFIR LEVY, Eidgenössische Technische Hochschule Zürich	1614
Stochastic Optimization with Variance Reduction for Infinite Datasets with Finite Sum Structure, ALBERTO BIETTI, Inria, and JULIEN MAIRAL	1624
Deep Learning with Topological Signatures, CHRISTOPH HOFER, University of Salzburg, and ROLAND KWITT, MARC NIETHAMMER, UNC Chapel Hill, ANDREAS UHL, University of Salzburg	1635
Predicting User Activity Level In Point Processes With Mass Transport Equation, YICHEN WANG, Georgia Tech, and XIAOJING YE, IBM TJ Watson Research Center, HONGYUAN ZHA, Georgia Institute of Technology, LE SONG, Georgia Tech	1646

Submultiplicative Glivenko-Cantelli and Uniform Convergence of Revenues , NOGA ALON, Tel Aviv University and Microsoft Research, and MOSHE BABAI OFF, Microsoft Research, YANNAI GONCZAROWSKI, The Hebrew University of Jerusalem and Microsoft Research, YISHAY MANSOUR, Tel Aviv University, SHAY MORAN, UCSD Simons Institute for TC Berkeley and Max Planck Saarbrücken Germany, AMIR YEHUDAYOFF, Technion - Israel Institute of Technology	1657
Deep Dynamic Poisson Factorization Model , CHENGYUE GONG, Peking University, and WIN-BIN HUANG, peking university	1667
Positive-Unlabeled Learning with Non-Negative Risk Estimator , RYUICHI KIRYO, and GANG NIU, The University of Tokyo / RIKEN, MARTINUS PLESSIS, MASASHI SUGIYAMA, RIKEN/The University of Tokyo	1676
Optimal Sample Complexity of M-wise Data for Top-K Ranking , MINJE JANG, KAIST, and SUNGHYUN KIM, ETRI, CHANGHO SUH, KAIST, SEWOONG OH, University of Illinois at Urbana-Champaign	1687
Counterfactual Gaussian Processes for Reliable Decision-making and What-if Reasoning , PETER SCHULAM, Johns Hopkins University, and SUCHI SARIA, JOHNS HOPKINS UNIVERSITY	1698
QSGD: Communication-Efficient SGD via Gradient Quantization and Encoding , DAN ALISTARH, and DEMJAN GRUBIC, Eidgenössische Technische Hochschule Zürich, JERRY LI, Massachusetts Institute of Technology, RYOTA TOMIOKA, MSRC, MILAN VOJNOVIC, LSE	1710
Convergent Block Coordinate Descent for Training Tikhonov Regularized Deep Neural Networks , ZIMING ZHANG, and MATTHEW BRAND, Mitsubishi Electric Research Laboratories	1722
Train longer, generalize better: closing the generalization gap in large batch training of neural networks , ELAD HOFFER, Technion, and ITAY HUBARA, DANIEL SOUDRY, Columbia University	1732
Flexpoint: An Adaptive Numerical Format for Efficient Training of Deep Neural Networks , URS KÖSTER, and TRISTAN WEBB, Intel Corporation, XIN WANG, INTEL CORPORATION, MARCEL NASSAR, Intel Corporation, ARJUN BANSAL, WILLIAM CONSTABLE, OGUZ ELIBOL, STEWART HALL, LUKE HORN OF, AMIR KHOSROWSHAHI, CAREY KLOSS, Intel, RUBY PAI, INTEL CORPORATION, NAVEEN RAO, Intel	1743
Model evidence from nonequilibrium simulations , MICHAEL HABECK, Max Planck Society	1754
Minimal Exploration in Structured Stochastic Bandits , RICHARD COMBES, and STEFAN MAGUREANU, KTH, ALEXANDRE PROUTIERE	1764
Learned D-AMP: Principled Neural Network based Compressive Image Recovery , CHRISTOPHER METZLER, and ALI MOUSAVI, RICHARD BARANIUK, Rice University	1773
Deliberation Networks: Sequence Generation Beyond One-Pass Decoding , YINGCE XIA, University of Science and Technology of China, and FEI TIAN, Microsoft Research, LIJUN WU, Sun Yat-sen University, JIANXIN LIN, USTC, TAO QIN, Microsoft Research, NENGHAI YU, University of Science and Technology of China, TIE-YAN LIU, Microsoft Research	1785
Adaptive Clustering through Semidefinite Programming , MARTIN ROYER, Université Paris-Saclay	1796

Log-normality and Skewness of Estimated State/Action Values in Reinforcement Learning , LIANGPENG ZHANG, University of Science and Technology of China, and KE TANG, XIN YAO, Southern University of Science and Technology China	1805
Repeated Inverse Reinforcement Learning , KAREEM AMIN, Google Research, and NAN JIANG, University of Michigan, SATINDER SINGH, UMICH ..	1816
The Numerics of GANs , LARS MESCHEDER, MPI Tübingen, and SEBASTIAN NOWOZIN, Microsoft Research, ANDREAS GEIGER, MPI Tübingen	1826
Practical Bayesian Optimization for Model Fitting with Bayesian Adaptive Direct Search , LUIGI ACERBI, and WEI MA, New York University	1837
Learning Chordal Markov Networks via Branch and Bound , KARI RANTANEN, and ANTTI HYTTINEN, MATTI JÄRVISALO, University of Helsinki	1848
Revenue Optimization with Approximate Bid Predictions , ANDRES MUNOZ, and SERGEI VASSILVITSKII, Google	1859
Solving Most Systems of Random Quadratic Equations , GANG WANG, and GEORGIOS GIANNAKIS, YOUSEF SAAD, University of Minnesota, JIE CHEN, Beijing Institute of Technology	1868
Unsupervised Learning of Disentangled and Interpretable Representations from Sequential Data , WEI-NING HSU, Massachusetts Institute of Technology, and ZHANG YU, JAMES GLASS, Massachusetts Institute of Technology	1879
Lookahead Bayesian Optimization with Inequality Constraints , REMI LAM, and KAREN WILLCOX, Massachusetts Institute of Technology	1891
Hierarchical Methods of Moments , MATTEO RUFFINI, UPC, and GUILLAUME RABUSSEAU, McGill University, BORJA BALLE	1902
Interpretable and Globally Optimal Prediction for Textual Grounding using Image Concepts , RAYMOND YEH, University of Illinois at Urbana-Champaign, and JINJUN XIONG, WEN-MEI HWU, MINH DO, ALEXANDER SCHWING, University of Illinois at Urbana-Champaign	1913
Revisit Fuzzy Neural Network: Demystifying Batch Normalization and ReLU with Generalized Hamming Network , LIXIN FAN, NOKIA	1924
Speeding Up Latent Variable Gaussian Graphical Model Estimation via Nonconvex Optimization , PAN XU, University of Virginia, and JIAN MA, Carnegie Mellon University, QUANQUAN GU, University of Virginia	1934
Batch Renormalization: Towards Reducing Minibatch Dependence in Batch-Normalized Models , SERGEY IOFFE, Google ..	1946
Generating steganographic images via adversarial training , JAMIE HAYES, University College London, and GEORGE DANEZIS	1955
Near-linear time approximation algorithms for optimal transport via Sinkhorn iteration , JASON ALTSCHULER, and JONATHAN WEED, PHILIPPE RIGOLLET, Massachusetts Institute of Technology	1965
PixelGAN Autoencoders , ALIREZA MAKHZANI, University of Toronto, and BRENDAN FREY, U. Toronto	1976

Consistent Multitask Learning with Nonlinear Output Relations, CARLO CILIBERTO, University College London, and ALESSANDRO RUDI, Istituto Italiano di Tecnologia, LORENZO ROSASCO, Massachusetts Institute of Technology, MASSIMILIANO PONTIL	1987
Alternating minimization for dictionary learning with random initialization, NILADRI CHATTERJI, and PETER BARTLETT, UC Berkeley	1998
Learning ReLUs via Gradient Descent, MAHDI SOLTANOLKOTABI, University of Southern California	2008
Stabilizing Training of Generative Adversarial Networks through Regularization, KEVIN ROTH, Eidgenössische Technische Hochschule Zürich, and AURELIEN LUCCHI, SEBASTIAN NOWOZIN, Microsoft Research, THOMAS HOFMANN	2019
Expectation Propagation with Stochastic Kinetic Model in Complex Interaction Systems, LE FANG, SUNY University at Buffalo, and FAN YANG, WEN DONG, University at Buffalo, TONG GUAN, CHUNMING QIAO, University at Buffalo	2030
Data-Efficient Reinforcement Learning in Continuous State-Action Gaussian-POMDPs, ROWAN MCALLISTER, and CARL RASMUSSEN, University of Cambridge	2041
Compatible Reward Inverse Reinforcement Learning, ALBERTO METELLI, Politecnico di Milano, and MATTEO PIROTTA, INRIA Sequel, MARCELLO RESTELLI	2051
First-Order Adaptive Sample Size Methods to Reduce Complexity of Empirical Risk Minimization, ARYAN MOKHTARI, and ALEJANDRO RIBEIRO, University of Pennsylvania	2061
Hiding Images in Plain Sight: Deep Steganography, SHUMEET BALUJA, Google Inc.	2070
Neural Program Meta-Induction, JACOB DEVLIN, Microsoft Research, and RUDY BUNEL, University of oxford, RISHABH SINGH, Microsoft, MATTHEW HAUSKNECHT, Microsoft Research, PUSHMEET KOHLI	2081
Bayesian Dyadic Trees and Histograms for Regression, STEPHANIE PAS, Leiden University, and VERONIKA ROCKOVA, University of Chicago	2090
A graph-theoretic approach to multitasking, NOGA ALON, Tel Aviv University, and DANIEL REICHMAN, IGOR SHINKAR, UC Berkeley, TAL WAGNER, Massachusetts Institute of Technology, SEBASTIAN MUSSLICK, JONATHAN COHEN, Princeton University, TOM GRIFFITHS, UC Berkeley, BISWADIP DEY, KAYHAN OZCIMDER, Princeton University	2101
Consistent Robust Regression, KUSH BHATIA, University of California Berkeley, and PRATEEK JAIN, PARAMESWARAN KAMALARUBAN, The Australian National Univer, PURUSHOTTAM KAR, IIT Kanpur	2111
Natural Value Approximators: Learning when to Trust Past Estimates, ZHONGWEN XU, and JOSEPH MODAYIL, DeepMind, HADO HASSELT, ANDRE BARRETO, DAVID SILVER, TOM SCHAUL	2121
Bandits Dueling on Partially Ordered Sets, JULIEN AUDIFFREN, CMLA ENS CACHAN, and LIVA RALAIVOLA, IUF Université Aix-Marseille CNRS	2130

Elementary Symmetric Polynomials for Optimal Experimental Design , ZELDA MARIET, and SUVRIT SRA, Massachusetts Institute of Technology	2140
Emergence of Language with Multi-agent Games: Learning to Communicate with Sequences of Symbols , SERHII HAVRYLOV, ILLC University of Amsterdam, and IVAN TITOV, University of Edinburgh / University of Amsterdam	2150
Training Deep Networks without Learning Rates Through Coin Betting , FRANCESCO ORABONA, Stony Brook University, and TATIANA TOMMASI, UNC Chapel Hill	2161
Pixels to Graphs by Associative Embedding , ALEJANDRO NEWELL, and JIA DENG, University of Michigan	2172
Runtime Neural Pruning , JI LIN, and YONGMING RAO, JIWEN LU, JIE ZHOU, Tsinghua University	2182
Eigenvalue Decay Implies Polynomial-Time Learnability for Neural Networks , SURBHI GOEL, and ADAM KLIVANS, University of Texas at Austin	2193
MMD GAN: Towards Deeper Understanding of Moment Matching Network , CHUN-LIANG LI, and WEI-CHENG CHANG, Carnegie Mellon University, YU CHENG, IBM Research, YIMING YANG, BARNABAS POCZOS, Carnegie Mellon University	2204
The Reversible Residual Network: Backpropagation Without Storing Activations , AIDAN GOMEZ, and MENGYE REN, University of Toronto, RAQUEL URTASUN, SAC, ROGER GROSSE, University of Toronto ...	2215
Fast Rates for Bandit Optimization with Upper-Confidence Frank-Wolfe , QUENTIN BERTHET, and VIANNEY PERCHET, ENS Paris-Saclay & Criteo Research	2226
Zap Q-Learning , ADITHYA DEVRAJ, and SEAN MEYN, University of Florida	2236
Expectation Propagation for t-Exponential Family Using q-Algebra , FUTOSHI FUTAMI, The University of Tokyo, and ISSEI SATO, The University of Tokyo/RIKEN, MASASHI SUGIYAMA, RIKEN/The University of Tokyo	2246
Few-Shot Learning Through an Information Retrieval Lens , ELENI TRIANTAFILLOU, and RICHARD ZEMEL, University of Toronto, RAQUEL URTASUN, SAC	2256
Formal Guarantees on the Robustness of a Classifier against Adversarial Manipulation , MATTHIAS HEIN, and MAKSYM ANDRIUSHCHENKO, Saarland University	2267
Associative Embedding: End-to-End Learning for Joint Detection and Grouping , ALEJANDRO NEWELL, University of Michigan, and ZHIAO HUANG, IIIS Tsinghua University, JIA DENG, University of Michigan	2278
Practical Locally Private Heavy Hitters , RAEF BASSILY, University of California San Diego, and KOBBY NISSIM, Georgetown University, URI STEMMER, Harvard University, ABHRADEEP THAKURTA, University of California Santa Cruz	2289

Large-Scale Quadratically Constrained Quadratic Program via Low-Discrepancy Sequences, KINJAL BASU, LinkedIn, and ANKAN SAHA, SHAUNAK CHATTERJEE	2298
Inhomogeneous Hypergraph Clustering with Applications, PAN LI, and OLGICA MILENKOVIC, University of Illinois at Urbana-Champaign	2309
Differentiable Learning of Logical Rules for Knowledge Base Reasoning, FAN YANG, and ZHILIN YANG, WILLIAM COHEN, Carnegie Mellon University	2320
Deep Multi-task Gaussian Processes for Survival Analysis with Competing Risks, AHMED ALAA, University of California Los Angeles, and MIHAELA SCHAAR, University of Oxford	2330
Masked Autoregressive Flow for Density Estimation, GEORGE PAPAMAKARIOS, University of Edinburgh, and IAIN MURRAY, THEO PAVLAKOU	2339
Non-convex Finite-Sum Optimization Via SCSG Methods, LIHUA LEI, and CHENG JU, JIANBO CHEN, UC Berkeley, MICHAEL JORDAN, University of California at Berkeley	2349
Beyond normality: Learning sparse probabilistic graphical models in the non-Gaussian setting, REBECCA MORRISON, and RICARDO BAPTISTA, YOUSSEF MARZOUK, Massachusetts Institute of Technology	2360
An inner-loop free solution to inverse problems using deep neural networks, KAI FAN, and QI WEI, Duke University, LAWRENCE CARIN, duke, KATHERINE HELLER	2371
OnACID: Online Analysis of Calcium Imaging Data in Real Time, ANDREA GIOVANNUCCI, Flatiron Institute, and JOHANNES FRIEDRICH, Columbia University, MATT KAUFMAN, ANNE CHURCHLAND, Cold Spring Harbor Laboratory, DMITRI CHKLOVSKII, Simons Foundation, LIAM PANINSKI, EFTYCHIOS PNEVMATIKAKIS, Flatiron Institute	2382
Collaborative PAC Learning, AVRIM BLUM, Toyota Technological Institute at Chicago, and NIKA HAGHTALAB, ARIEL PROCACCIA, Carnegie Mellon University, MINGDA QIAO, Tsinghua University	2393
Fast Black-box Variational Inference through Stochastic Trust-Region Optimization, JEFFREY REGIER, UC Berkeley, and MICHAEL JORDAN, University of California at Berkeley, JON MCAULIFFE, UC Berkeley	2403
Scalable Demand-Aware Recommendation, JINFENG YI, and CHO-JUI HSIEH, University of California, Davis, KUSH VARSHNEY, LIJUN ZHANG, YAO LI, University of California Davis	2413
SGD Learns the Conjugate Kernel Class of the Network, AMIT DANIELY	2423
Noise-Tolerant Interactive Learning Using Pairwise Comparisons, YICHONG XU, Carnegie Mellon UniversityC, and HONGYANG ZHANG, AARTI SINGH, ARTUR DUBRAWSKI, KYLE MILLER, Carnegie Mellon University	2432
Analyzing Hidden Representations in End-to-End Automatic Speech Recognition Systems, YONATAN BELINKOV, and JAMES GLASS, Massachusetts Institute of Technology	2442

Generative Local Metric Learning for Kernel Regression, YUNG-KYUN NOH, and MASASHI SUGIYAMA, RIKEN/The University of Tokyo, KEE-EUNG KIM, FRANK PARK, Seoul National University, DANIEL LEE, UPenn	2453
Information Theoretic Properties of Markov Random Fields, and their Algorithmic Applications, LINUS HAMILTON, and FREDERIC KOEHLER, ANKUR MOITRA, Massachusetts Institute of Technology	2464
Fitting Low-Rank Tensors in Constant Time, KOHEI HAYASHI, AIST / RIKEN, and YUICHI YOSHIDA, National Institute of Informat	2474
Deep Supervised Discrete Hashing, QI LI, CASIA, and ZHENAN SUN, RAN HE, TIENIU TAN	2483
Using Options and Covariance Testing for Long Horizon Off-Policy Policy Evaluation, ZHAOHAN GUO, and PHILIP THOMAS, Carnegie Mellon University, EMMA BRUNSKILL, Stanford University	2493
How regularization affects the critical points in linear networks, AMIRHOSSEIN TAGHVAEI, and JIN KIM, PRASHANT MEHTA, University of Illinois at Urbana-Champaign	2503
Fisher GAN, YOUSSEF MROUEH, and TOM SERCU, International Business Machines (IBM)	2514
Information-theoretic analysis of generalization capability of learning algorithms, AOLIN XU, University of Illinois, and MAXIM RAGINSKY, University of Illinois at Urbana-Champaign	2525
Sparse Approximate Conic Hulls, GREG BUSKIRK, and BEN RAICHEL, NICHOLAS RUOZZI, UT Dallas	2535
Rigorous Dynamics and Consistent Estimation in Arbitrarily Conditioned Linear Systems , ALLIE ALLIE, Alyson, and MOJTABA SAHRAEE-ARDAKAN, University of California Los Angeles, SUNDEEP RANGAN, New York University, PHILIP SCHNITER, Ohio State University	2546
Toward Goal-Driven Neural Network Models for the Rodent Whisker-Trigeminal System, CHENGXU ZHUANG, Stanford University, and JONAS KUBILIUS, Massachusetts Institute of Technology, MITRA HARTMANN, Northwestern University, DANIEL YAMINS, Stanford University ..	2556
Accuracy First: Selecting a Differential Privacy Level for Accuracy Constrained ERM, KATRINA LIGETT, and SETH NEEL, AARON ROTH, UPenn, BO WAGGONER, STEVEN WU, Microsoft	2567
EX2: Exploration with Exemplar Models for Deep Reinforcement Learning, JUSTIN FU, and JOHN CO-REYES, SERGEY LEVINE, UC Berkeley	2578
Multitask Spectral Learning of Weighted Automata, GUILLAUME RABUSSEAU, McGill University, and BORJA BALLE, JOELLE PINEAU	2589
Multi-way Interacting Regression via Factorization Machines, MIKHAIL YUROCHKIN, and XUANLONG NGUYEN, University of Michigan, NIKOLAOS VASILOGLOU, LogicBlox	2599
Predicting Organic Reaction Outcomes with Weisfeiler-Lehman Network, WENGONG JIN, MIT CSAIL, and CONNOR COLEY, MIT Department of Chemical Engineering, REGINA BARZILAY, MIT CSAIL, TOMMI JAAKKOLA, Massachusetts Institute of Technology	2608

Practical Data-Dependent Metric Compression with Provable Guarantees , PIOTR INDYK, and ILYA RAZENSHTEYN, TAL WAGNER, Massachusetts Institute of Technology	2618
REBAR: Low-variance, unbiased gradient estimates for discrete latent variable models , GEORGE TUCKER, Google Brain, and ANDRIY MNIH, DeepMind, CHRISTOPHER MADDISON, Oxford/DeepMind, DIETERICH LAWSON, Google Brain, JASCHA SOHL-DICKSTEIN	2628
Nonlinear random matrix theory for deep learning , JEFFREY PENNINGTON, and PRATIK WORAH, Google	2638
Parallel Streaming Wasserstein Barycenters , MATTHEW STAIB, and SEBASTIAN CLAICI, JUSTIN SOLOMON, STEFANIE JEGELKA, Massachusetts Institute of Technology	2648
ELF: An Extensive, Lightweight and Flexible Research Platform for Real-time Strategy Games , YUANDONG TIAN, and QUCHENG GONG, Facebook AI Research, WENLING SHANG, 435024885627, YUXIN WU, LARRY ZITNICK, Facebook AI Research	2660
Dual Discriminator Generative Adversarial Nets , TU NGUYEN, and TRUNG LE, HUNG VU, DINH PHUNG, Deakin University	2671
Dynamic Revenue Sharing , SANTIAGO BALSEIRO, Duke University, and MAX LIN, VAHAB MIRROKNI, Google, RENATO LEME, Google Research, SONG ZUO, IIS Tsinghua University	2682
Decomposition-Invariant Conditional Gradient for General Polytopes with Line Search , MOHAMMAD BASHIRI, and XINHUA ZHANG, University of Illinois Chicago	2691
VAIN: Attentional Multi-agent Predictive Modeling , YEDID HOSHEN, FAIR	2702
An Empirical Bayes Approach to Optimizing Machine Learning Algorithms , JAMES MCINERNEY, Columbia	2713
Differentially Private Empirical Risk Minimization Revisited: Faster and More General , DI WANG, SUNY at Buffalo, and MINWEI YE, University at Buffalo, JINHUI XU, SUNY at Buffalo	2723
Variational Inference via χ Upper Bound Minimization , ADJI DIENG, and DUSTIN TRAN, Columbia University, RAJESH RANGANATH, Princeton University, JOHN PAISLEY, DAVID BLEI, Columbia University	2733
On Quadratic Convergence of DC Proximal Newton Algorithm in Nonconvex Sparse Learning , XINGGUO LI, University of Minnesota, and LIN YANG, JASON GE, Princeton University, JARVIS HAUPT, University of Minnesota, TONG ZHANG, TUO ZHAO, Georgia Tech	2743
#Exploration: A Study of Count-Based Exploration for Deep Reinforcement Learning , HAORAN TANG, UC Berkeley, and REIN HOUTHOOFT, OpenAI, DAVIS FOOTE, ADAM STOOKE, UC Berkeley, XI CHEN, OpenAI UC Berkeley, YAN DUAN, JOHN SCHULMAN, OpenAI, FILIP DETURCK, PIETER ABBEEL, University of California at Berkeley	2754
An Empirical Study on The Properties of Random Bases for Kernel Methods , MAXIMILIAN ALBER, TU Berlin, and PIETER-JAN KINDERMANS, KRISTOF SCHÜTT, TU Berlin, KLAUS-ROBERT MÜLLER, FEI SHA, University of Southern California	2764

Bridging the Gap Between Value and Policy Based Reinforcement Learning, OFIR NACHUM, Google, and MOHAMMAD NOROUZI, KELVIN XU, Google, DALE SCHUURMANS, University of Alberta & Google Brain	2776
Premise Selection for Theorem Proving by Deep Graph Embedding, MINGZHE WANG, and YIHE TANG, JIAN WANG, JIA DENG, University of Michigan	2787
A Bayesian Data Augmentation Approach for Learning Deep Models, TOAN TRAN, and TRUNG PHAM, GUSTAVO CARNEIRO, LYLE PALMER, IAN REID, The University of Adelaide	2798
Principles of Riemannian Geometry in Neural Networks, MICHAEL HAUSER, and ASOK RAY, Pennsylvania State University	2808
Cold-Start Reinforcement Learning with Softmax Policy Gradient, NAN DING, and RADU SORICUT, Google	2818
Online Dynamic Programming, HOLAKOU RAHMANIAN, and MANFRED WARMUTH, University of California Santa Cruz	2828
Alternating Estimation for Structured High-Dimensional Multi-Response Models, SHENG CHEN, Univ of Minnesota Twin Cities, and ARINDAM BANERJEE, University of Minnesota	2839
Convolutional Gaussian Processes, MARK WILK, and CARL RASMUSSEN, University of Cambridge, JAMES HENSMAN	2850
Estimation of the covariance structure of heavy-tailed distributions, XIAOHAN WEI, Univ. of Southern California, and STANISLAV MINSKER, University of Southern California	2860
Mean Field Residual Networks: On the Edge of Chaos, GE YANG, Harvard University, and SAM SCHOENHOLZ, Google	N/A
Decomposable Submodular Function Minimization: Discrete and Continuous, ALINA ENE, Boston University, and HUY NGUYEN, LASZLO VEGH, London School of Economics	2871
Gauging Variational Inference, SUNGSOO AHN, KAIST, and MISHA CHERTKOV, JINWOO SHIN, KAIST	2882
Deep Recurrent Neural Network-Based Identification of Precursor microRNAs, SEUNGHYUN PARK, Korea University, and SEONWOO MIN, HYUN-SOO CHOI, SUNGROH YOON, Seoul National University	2892
Robust Estimation of Neural Signals in Calcium Imaging, HAKAN INAN, Stanford University, and MURAT ERDOGDU, Microsoft Research, MARK SCHNITZER, Stanford University	2902
State Aware Imitation Learning, YANNICK SCHROECKER, Georgia Institute of Technolog, and CHARLES ISBELL, Georgia Tech	2912
Beyond Parity: Fairness Objectives for Collaborative Filtering, SIRUI YAO, and BERT HUANG, Virginia Tech	2922
A PAC-Bayesian Analysis of Randomized Learning with Application to Stochastic Gradient Descent, BEN LONDON, Amazon	2932
Fully Decentralized Policies for Multi-Agent Systems: An Information Theoretic Approach, ROEL DOBBE, and DAVID FRIDOVICH-KEIL, CLAIRE TOMLIN, EECS UC Berkeley	2942

Model-Powered Conditional Independence Test , RAJAT SEN, University of Texas at Austin, and ANANDA SURESH, Google, KARTHIKEYAN SHANMUGAM, IBM Research, ALEXANDROS DIMAKIS, UTexas, SANJAY SHAKKOTTAI, University of Texas at Austin	2952
Deep Voice 2: Multi-Speaker Neural Text-to-Speech , ANDREW GIBIANSKY, and SERCAN ARIK, Baidu Research, GREGORY DIAMOS, Baidu, JOHN MILLER, KAINAN PENG, WEI PING, JONATHAN RAIMAN, YANQI ZHOU, Baidu Research	2963
Variance-based Regularization with Convex Objectives , HONGSEOK NAMKOONG, and JOHN DUCHI, Stanford University	2972
Deep Lattice Networks and Partial Monotonic Functions , SEUNGIL YOU, Google, and DAVID DING, KEVIN CANINI, JAN PFEIFER, Google, MAYA GUPTA	2982
Continual Learning with Deep Generative Replay , HANUL SHIN, MIT and SK T-Brain, and JUNG LEE, JAEHONG KIM, JIWON KIM, SK T-Brain	2991
AIDE: An algorithm for measuring the accuracy of probabilistic inference algorithms , MARCO CUSUMANO-TOWNER, and VIKASH MANSINGHKA, Massachusetts Institute of Technology	3001
Learning Causal Structures Using Regression Invariance , AMIREMAD GHASSAMI, and SABER SALEHKALEYBAR, NEGAR KIYAVASH, University of Illinois at Urbana-Champaign, KUN ZHANG, Carnegie Mellon University	3012
Online Influence Maximization under Independent Cascade Model with Semi-Bandit Feedback , ZHENG WEN, and BRANISLAV KVETON, Adobe Research, MICHAL VALKO, Inria Lille - Nord Europe, SHARAN VASWANI, UBC	3023
Near Minimax Optimal Players for the Finite-Time 3-Expert Prediction Problem , YASIN ABBASI, and PETER BARTLETT, UC Berkeley, VICTOR GABILLON	3034
Reinforcement Learning under Model Mismatch , AURKO ROY, Google Brain, and HUAN XU, Georgia Tech, SEBASTIAN POKUTTA, GT	3044
Hierarchical Attentive Recurrent Tracking , ADAM KOSIOREK, and ALEX BEWLEY, INGMAR POSNER, University of Oxford	3054
Tomography of the London Underground: a Scalable Model for Origin-Destination Data , NICOLO COLOMBO, University College London, and RICARDO SILVA, SOONG KANG, University College London	3063
Rotting Bandits , NIR LEVINE, The Technion, and KOBY CRAMMER, SHIE MANNOR, Technion	3075
Unbiased estimates for linear regression via volume sampling , MICHAL DEREZINSKI, UC Santa Cruz, and MANFRED WARMUTH, University of California Santa Cruz	3085
Approximation Bounds for Hierarchical Clustering: Average Linkage, Bisecting K-means, and Local Search , BENJAMIN MOSELEY, Carnegie Mellon University, and JOSHUA WANG, Stanford University	3095

Adaptive Accelerated Gradient Converging Method under Hölderian Error Bound Condition, MINGRUI LIU, and TIANBAO YANG, University of Iowa	3105
Stein Variational Gradient Descent as Gradient Flow, QIANG LIU, Dartmouth College	3116
Partial Hard Thresholding: Towards A Principled Analysis of Support Recovery, JIE SHEN, Rutgers University, and PING LI	3125
Shallow Updates for Deep Reinforcement Learning, NIR LEVINE, and TOM ZAHAVY, The Technion, DANIEL MANKOWITZ, Technion, AVIV TAMAR, UC Berkeley, SHIE MANNOR, Technion	3136
LightGBM: A Highly Efficient Gradient Boosting Decision Tree, GUOLIN KE, Microsoft, and QI MENG, Peking University, THOMAS FINLEY, TAIFENG WANG, WEI CHEN, WEIDONG MA, QIWEI YE, TIE-YAN LIU, Microsoft Research	3147
Adversarial Ranking for Language Generation, KEVIN LIN, and DIANQI LI, University of Washington, XIAODONG HE, MING-TING SUN, University of Washington, ZHENGYOU ZHANG, Microsoft Research	3156
Regret Minimization in MDPs with Options without Prior Knowledge, RONAN FRUIT, Inria Lille, and MATTEO PIROTTA, INRIA Sequel, ALESSANDRO LAZARIC, INRIA, EMMA BRUNSKILL, Carnegie Mellon University	3167
Net-Trim: Convex Pruning of Deep Neural Networks with Performance Guarantee, ALIREZA AGHASI, IBM TJ Watson Research Center, and AFSHIN ABDI, Georgia Institute of Technology, NAM NGUYEN, IBM TJ Watson Research, JUSTIN ROMBERG, Georgia Institute of Technology	3178
Graph Matching via Multiplicative Update Algorithm, BO JIANG, Anhui University, and JIN TANG, CHRIS DING, YIHONG GONG, BIN LUO	3188
Dynamic Importance Sampling for Anytime Bounds of the Partition Function, QI LOU, and RINA DECHTER, ALEX IHLER, University of California Irvine	3197
Is the Bellman residual a bad proxy?, MATTHIEU GEIST, Université de Lorraine, and BILAL PIOT, OLIVIER PIETQUIN, DeepMind	3206
Generalization Properties of Learning with Random Features, ALESSANDRO RUDI, and LORENZO ROSASCO, Massachusetts Institute of Technology	3216
Differentially private Bayesian learning on distributed data, MIKKO HEIKKILÄ, and EEMIL LAGERSPETZ, University of Helsinki, SAMUEL KASKI, KANA SHIMIZU, Waseda University, SASU TARKOMA, ANTTI HONKELA, University of Helsinki	3227
Learning to Compose Domain-Specific Transformations for Data Augmentation, ALEXANDER RATNER, and HENRY EHRENBERG, ZESHAN HUSSAIN, JARED DUNNMON, CHRISTOPHER RE, Stanford University	3237
Wasserstein Learning of Deep Generative Point Process Models, SHUAI XIAO, Shanghai Jiao Tong University, and MEHRDAD FARAJTABAR, Georgia Tech, XIAOJING YE, IBM TJ Watson Research Center, JUNCHI YAN, IBM Research - China, XIAOKANG YANG, Shanghai Jiao Tong University, LE SONG, Georgia Tech, HONGYUAN ZHA, Georgia Institute of Technology	3248

Ensemble Sampling , XIUYUAN LU, and BENJAMIN VAN ROY, Stanford University	3259
Language Modeling with Recurrent Highway Hypernetworks , JOSEPH SUAREZ, Stanford University	3268
Adaptive SVRG Methods under Error Bound Conditions with Unknown Growth Parameter , YI XU, The University of Iowa, and QIHANG LIN, Univ Iowa faculty, TIANBAO YANG, University of Iowa	3278
Bayesian Compression for Deep Learning , CHRISTOS LOUIZOS, and KAREN ULLRICH, University of Amsterdam, MAX WELLING, University of Amsterdam and University of California Irvine and CIFAR	3289
Streaming Sparse Gaussian Process Approximations , THANG BUI, and CUONG NGUYEN, University of Cambridge, RICHARD TURNER	3300
VEEGAN: Reducing Mode Collapse in GANs using Implicit Variational Learning , AKASH SRIVASTAVA, and LAZAR VALKOZ, University of Edinburgh, CHRIS RUSSELL, The Alan Turing Institute, MICHAEL GUTMANN, University of Edinburgh, CHARLES SUTTON	3309
Sparse Embedded k-Means Clustering , WEIWEI LIU, UNSW, and XIAOBO SHEN, NJUST, IVOR TSANG, University of Technology Sydney	3320
Dynamic-Depth Context Tree Weighting , JOAO MESSIAS, Morpheus Labs, and SHIMON WHITESON, Oxford University	3329
A Regularized Framework for Sparse and Structured Neural Attention , VLAD NICULAE, Cornell University, and MATHIEU BLONDEL, NTT	3339
Multi-output Polynomial Networks and Factorization Machines , MATHIEU BLONDEL, NTT, and VLAD NICULAE, Cornell University, TAKUMA OTSUKA, NTT, NAONORI UEDA	3350
Clustering Billions of Reads for DNA Data Storage , CYRUS RASHTCHIAN, University of Washington, and KONSTANTIN MAKARYCHEV, Northwestern university, MIKLOS RACZ, Princeton University, SIENA ANG, DJORDJE JEVDJIC, SERGEY YEKHANIN, Microsoft, LUIS CEZE, University of Washington, KARIN STRAUSS, Microsoft	3361
Multi-Objective Non-parametric Sequential Prediction , GUY UZIEL, and RAN EL-YANIV, Technion	3373
A Universal Analysis of Large-Scale Regularized Least Squares Solutions , ASHKAN PANAHI, NC state university, and BABAK HASSIBI	3382
Deep Sets , MANZIL ZAHEER, and SATWIK KOTTUR, SIAMAK RAVANBAKSH, BARNABAS POZOS, RUSLAN SALAKHUTDINOV, ALEX SMOLA, Carnegie Mellon University	3392
ExtremeWeather: A large-scale climate dataset for semi-supervised detection, localization, and understanding of extreme weather events , EVAN RACAH, U. Montreal, and CHRISTOPHER BECKHAM, TEGAN MAHARAJ, MILA Polytechnic Montreal, SAMIRA KAHOU, Microsoft Maluuba, MR PRABHAT, CHRIS PAL	3403
Process-constrained batch Bayesian optimisation , PRATIBHA VELLANKI, and SANTU RANA, SUNIL GUPTA, DAVID RUBIN, ALESSANDRA SUTTI, THOMAS DORIN, MURRAY HEIGHT, Deakin University, PAUL SANDERS, SVETHA VENKATESH, Deakin University	3415

Bayesian Inference of Individualized Treatment Effects using Multi-task Gaussian Processes, AHMED ALAA, University of California Los Angeles, and MIHAELA SCHAAR, University of Oxford	3425
Spherical convolutions and their application in molecular modelling, WOUTER BOOMSMA, University of Copenhagen, and JES FRELLSEN, IT University of Copenhagen	3434
Efficient Optimization for Linear Dynamical Systems with Applications to Clustering and Sparse Coding, WENBING HUANG, Tsinghua University, and MEHRTASH HARANDI, Data61, TONG ZHANG, Australian National University, LIJIE FAN, FUCHUN SUN, Tsinghua University, JUNZHOU HUANG	3445
On Optimal Generalizability in Parametric Learning, AHMAD BEIRAMI, Harvard University, and MEISAM RAZAVIYAYN, University of Southern California, SHAHIN SHAHRAMPOUR, VAHID TAROKH, Harvard University	3456
Near Optimal Sketching of Low-Rank Tensor Regression, XINGGUO LI, and JARVIS HAUPT, University of Minnesota, DAVID WOODRUFF, Carnegie Mellon University	3467
Tractability in Structured Probability Spaces, ARTHUR CHOI, and YUJIA SHEN, ADNAN DARWICHE, University of California Los Angeles	3478
Model-based Bayesian inference of neural activity and connectivity from all-optical interrogation of a neural circuit, LAURENCE AITCHISON, University of Cambridge, and LLOYD RUSSELL, ADAM PACKER, University College London, JINYAO YAN, Janelia Research Campus, PHILIPPE CASTONGUAY, MICHAEL HAUSSER, University College London, SRINIVAS TURAGA, HHMI Janelia	3487
Gaussian process based nonlinear latent structure discovery in multivariate spike train data, ANQI WU, Princeton neuroscience institu, and NICHOLAS ROY, Princeton University, STEPHEN KEELEY, JONATHAN PILLOW, Princeton University	3497
Neural system identification for large populations separating “what” and “where”, DAVID KLINDT, University of Tübingen, and ALEXANDER ECKER, THOMAS EULER, Institute for Ophthalmic Research University of Tübingen, MATTHIAS BETHGE	3507
Certified Defenses for Data Poisoning Attacks, JACOB STEINHARDT, and PANG KOH, PERCY LIANG, Stanford University	3518
Eigen-Distortions of Hierarchical Representations, ALEXANDER BERARDINO, New York University, and VALERO LAPARRA, JOHANNES BALLÉ, EERO SIMONCELLI, New York University	3531
Limitations on Variance-Reduction and Acceleration Schemes for Finite Sums Optimization, YOSSI ARJEVANI, Weizmann Institute of Science	3541
Unsupervised Sequence Classification using Sequential Output Statistics, YU LIU, SUNY Buffalo, and JIANSHU CHEN, LI DENG	3551
Subset Selection under Noise, CHAO QIAN, University of Science and Technology of China, and JING-CHENG SHI, LAMDA, YANG YU, Nanjing University, KE TANG, University of Science and Technology of China, ZHI-HUA ZHOU, Nanjing University	3561

Collecting Telemetry Data Privately , BOLIN DING, and JANARDHAN KULKARNI, Microsoft Research, SERGEY YEKHANIN, Microsoft	3572
Concrete Dropout , YARIN GAL, University of Cambridge, and GEORGE HRON, ALEX KENDALL, University of Cambridge	3582
Adaptive Batch Size for Safe Policy Gradients , MATTEO PAPINI, Politecnico di Milano, and MATTEO PIROTTA, INRIA Sequel, MARCELLO RESTELLI	3592
A Disentangled Recognition and Nonlinear Dynamics Model for Unsupervised Learning , MARCO FRACCARO, Technical University of Denmark, and SIMON KAMRONN, DTU, ULRICH PAQUET, OLE WINTHER, Technical University of Denmark	3602
PASS-GLM: polynomial approximate sufficient statistics for scalable Bayesian GLM inference , JONATHAN HUGGINS, Massachusetts Institute of Technology, and RYAN ADAMS, Harvard University, TAMARA BRODERICK, Massachusetts Institute of Technology	3612
Bayesian GAN , YUNUS SAATCI, and ANDREW WILSON, Cornell University	3623
Off-policy evaluation for slate recommendation , ADITH SWAMINATHAN, Microsoft Research, and AKSHAY KRISHNAMURTHY, UMass, Amherst, ALEKH AGARWAL, MIRO DUDIK, JOHN LANGFORD, Microsoft Research, DAMIEN JOSE, IMED ZITOUNI, Microsoft	3633
A multi-agent reinforcement learning model of common-pool resource appropriation , JULIEN PEROLAT, DeepMind, and JOEL LEIBO, VINICIUS ZAMBALDI, CHARLES BEATTIE, KARL TUYLS, THORE GRAEPEL, DeepMind	3644
On the Optimization Landscape of Tensor Decompositions , RONG GE, and TENG YU MA, Princeton University	3654
High-Order Attention Models for Visual Question Answering , IDAN SCHWARTZ, Technion, and ALEXANDER SCHWING, University of Illinois at Urbana-Champaign, TAMIR HAZAN	3665
Sparse convolutional coding for neuronal assembly detection , SVEN PETER, University Heidelberg, and ELKE KIRSCHBAUM, IWR Heidelberg University, MARTIN BOTH, Institute for Physiology and Pathophysiology Heidelberg University, LEE CAMPBELL, BRANDON HARVEY, CONOR HEINS, Intramural Research Program National Institute on Drug Abuse, DANIEL DURSTEWITZ, CIMH Heidelberg University, FERRAN DIEGO, Bosch, FRED HAMPRECHT	3676
Quantifying how much sensory information in a neural code is relevant for behavior , GIUSEPPE PICA, Istituto Italiano di Tecnologia, and EUGENIO PIASINI, IIT, HOUMAN SAFAAI, CAROLINE RUNYAN, CHRISTOPHER HARVEY, Harvard Medical School, MATHEW DIAMOND, International School for Advanced Studies, CHRISTOPH KAYSER, University of Glasgow, TOMMASO FELLIN, STEFANO PANZERI, Istituto Italiano di Tecnologia	3687
Geometric Matrix Completion with Recurrent Multi-Graph Neural Networks , FEDERICO MONTI, University of Lugano, and MICHAEL BRONSTEIN, XAVIER BRESSON, NTU, Singapore	3698
Reducing Reparameterization Gradient Variance , ANDREW MILLER, Harvard University, and NICHOLAS FOTI, University of Washington, ALEXANDER D'AMOUR, UC Berkeley, RYAN ADAMS, Harvard University	3709

Visual Reference Resolution using Attention Memory for Visual Dialog , PAUL SEO, POSTECH, and ANDREAS LEHRMANN, Disney Research, BOHYUNG HAN, LEONID SIGAL	3720
Joint distribution optimal transportation for domain adaptation , NICOLAS COURTY, IRISA/UBS, and REMI FLAMARY, AMAURY HABRARD, University of Saint-Etienne Laboratoire Hubert Curien, ALAIN RAKOTOMAMONJY	3731
Multiresolution Kernel Approximation for Gaussian Process Regression , YI DING, and RISI KONDOR, JONATHAN ESKREIS-WINKLER, University of Chicago	3741
Collapsed variational Bayes for Markov jump processes , BOQIAN ZHANG, Purdue University, and JIANGWEI PAN, Facebook, VINAYAK RAO, Purdue University	3750
Universal consistency and minimax rates for online Mondrian Forests , JAOUAD MOURTADA, and STÉPHANE GAÏFFAS, ERWAN SCORNET, Ecole polytechnique	3759
Welfare Guarantees from Data , DARRELL HOY, University of Maryland, and DENIS NEKIPELOV, University of Virginia, VASILIS SYRGKANIS	3769
Diving into the shallows: a computational perspective on large-scale shallow learning , SIYUAN MA, and MIKHAIL BELKIN, The Ohio State University	3779
End-to-End Differentiable Proving , TIM ROCKTÄSCHEL, University College London, and SEBASTIAN RIEDEL	3789
Influence Maximization with ε-Almost Submodular Threshold Functions , QIANG LI, Institute of Computing Technol, and WEI CHEN, XIAOMING SUN, JIALIN ZHANG, Institute of Computing Technology Chinese Academy of Sciences	3802
InfoGAIL: Interpretable Imitation Learning from Visual Demonstrations , YUNZHU LI, Massachusetts Institute of Technology, and JIAMING SONG, STEFANO ERMON, Stanford University	3813
Variational Laws of Visual Attention for Dynamic Scenes , DARIO ZANCA, University of Florence University of Siena, and MARCO GORI	3824
Recursive Sampling for the Nystrom Method , CAMERON MUSCO, Massachusetts Institute of Tec, and CHRISTOPHER MUSCO, Mass Institute of Technology	3834
Interpolated Policy Gradient: Merging On-Policy and Off-Policy Gradient Estimation for Deep Reinforcement Learning , SHIXIANG GU, University of Cambridge, and TIMOTHY LILLICRAP, Google DeepMind, RICHARD TURNER, ZOUBIN GHAHRAMANI, University of Cambridge & Uber, BERNHARD SCHÖLKOPF, Max Planck Institute for Intelligent Systems, SERGEY LEVINE, UC Berkeley	3847
Dynamic Routing Between Capsules , SARA SABOUR, and NICHOLAS FROSST, Google, GEOFFREY HINTON, U. Toronto	3857
Incorporating Side Information by Adaptive Convolution , DI KANG, and DEBARUN DHAR, ANTONI CHAN, City University of Hong Kong ..	3868

Conic Scan-and-Cover algorithms for nonparametric topic modeling, MIKHAIL YUROCHKIN, and ARITRA GUHA, XUANLONG NGUYEN, University of Michigan	3879
FALKON: An Optimal Large Scale Kernel Method, ALESSANDRO RUDI, Massachusetts Institute of Technology, and LUIGI CARRATINO, Università di Genova, LORENZO ROSASCO, Massachusetts Institute of Technology	3889
Structured Generative Adversarial Networks, ZHIJIE DENG, Tsinghua University, and HAO ZHANG, XIAODAN LIANG, LUONA YANG, Carnegie Mellon University, SHIZHEN XU, JUN ZHU, Tsinghua University, ERIC XING, SAC	3900
Conservative Contextual Linear Bandits, ABBAS KAZEROUNI, Stanford University, and MOHAMMAD GHAVAMZADEH, YASIN YADKORI, Adobe Research, BENJAMIN VAN ROY, Stanford University	3911
Variational Memory Addressing in Generative Models, JORG BORNSCHEIN, and ANDRIY MNIH, DANIEL ZORAN, DeepMind, DANILO REZENDE	3921
On Tensor Train Rank Minimization : Statistical Efficiency and Scalable Algorithm, MASA AKI IMAIZUMI, Institute of Statistical Mathematics, and TAKANORI MAEHARA, RIKEN Center for Advanced Intelligence Project, KOHEI HAYASHI, AIST / RIKEN	3931
Scalable Levy Process Priors for Spectral Kernel Learning, PHILLIP JANG, and ANDREW LOEB, Cornell University, MATTHEW DAVIDOW, Cornell, ANDREW WILSON, Cornell University	3941
Deep Hyperspherical Learning, WEIYANG LIU, Georgia Tech, and YAN-MING ZHANG, Institute of Automation Chinese Academy of Sciences, XINGGUO LI, University of Minnesota, ZHIDING YU, Carnegie Mellon University, BO DAI, TUO ZHAO, LE SONG, Georgia Tech	3951
Learning Deep Structured Multi-Scale Features using Attention-Gated CRFs for Contour Prediction, DAN XU, University of Trento, and WANLI OUYANG, XAVIER ALAMEDA-PINEDA, INRIA, ELISA RICCI, XIAOGANG WANG, NICU SEBE	3962
On-the-fly Operation Batching in Dynamic Computation Graphs, GRAHAM NEUBIG, Carnegie Mellon University, and YOAV GOLDBERG, CHRIS DYER, Carnegie Mellon University	3972
Nonlinear Acceleration of Stochastic Algorithms, DAMIEN SCIEUR, INRIA - ENS, and FRANCIS BACH, SAC, ALEXANDRE D'ASPREMONT, ENS CNRS	3983
Optimized Pre-Processing for Discrimination Prevention, FLAVIO CALMON, Harvard, and DENNIS WEI, BHANUKIRAN VINZAMURI, International Business Machines (IBM), KARTHIKEYAN RAMAMURTHY, KUSH VARSHNEY	3993
YASS: Yet Another Spike Sorter, JIN LEE, Columbia University, and DAVID CARLSON, Duke University, HOOSHMAND RAZAGHI, WEICHI YAO, Columbia University, GEORGES GOETZ, Stanford University, ESPEN HAGEN, ELEANOR BATTY, Columbia University, EJ CHICHILNISKY, Stanford University, GAUTE EINEVOLL, LIAM PANINSKI	4003
Independence clustering (without a matrix), DANIIL RYABKO, INRIA	4014

Fast amortized inference of neural activity from calcium imaging data with variational autoencoders, ARTUR SPEISER, research center caesar an associate of the Max Planck Society, and JINYAO YAN, Janelia Research Campus, EVAN ARCHER, LARS BUESING, Google DeepMind, SRINIVAS TURAGA, HHMI Janelia, JAKOB MACKE	4025
Adaptive Active Hypothesis Testing under Limited Information, FABIO CECCHI, Eindhoven University of Technology, and NIDHI HEGDE, Nokia Bell Labs	4036
Streaming Weak Submodularity: Interpreting Neural Networks on the Fly, ETHAN ELENBERG, UT Austin, and ALEXANDROS DIMAKIS, UTexas, MORAN FELDMAN, Open University of Israel, AMIN KARBASI, Yale University	4045
Successor Features for Transfer in Reinforcement Learning, ANDRE BARRETO, and WILL DABNEY, REMI MUNOS, DeepMind, JONATHAN HUNT, TOM SCHAUL, DAVID SILVER, HADO HASSELT	4056
Counterfactual Fairness, MATT KUSNER, and JOSHUA LOFTUS, CHRIS RUSSELL, The Alan Turing Institute, RICARDO SILVA	4067
Prototypical Networks for Few-shot Learning, SNELL JAKE, Toronto, and KEVIN SWERSKY, Google Brain, RICHARD ZEMEL, University of Toronto	4078
Triple Generative Adversarial Nets, CHONGXUAN LI, and KUN XU, JUNE ZHU, BO ZHANG, Tsinghua University	4089
Efficient Sublinear-Regret Algorithms for Online Sparse Linear Regression with Limited Observation, SHINJI ITO, NEC Corporation, and DAISUKE HATANO, HANNA SUMITA, AKIHIRO YABE, TAKURO FUKUNAGA, NII, NAONORI KAKIMURA, KEN-ICHI KAWARABAYASHI	4100
Mapping distinct timescales of functional interactions among brain networks, MALI SUNDARESAN, Indian Institute of Science Bangalore, and ARSHED NABEEL, DEVARAJAN SRIDHARAN	4110
Multi-Armed Bandits with Metric Movement Costs, TOMER KOREN, Google, and ROI LIVNI, Princeton, YISHAY MANSOUR, Tel Aviv University	4120
Learning A Structured Optimal Bipartite Graph for Co-Clustering, FEIPING NIE, Northwestern Polytechnical University, and XIAOQIAN WANG, University of Texas at Arlington, CHENG DENG, School of Electronic Engineering Xidian University China, HENG HUANG, University of Pittsburgh	4130
Learning Low-Dimensional Metrics, BLAKE MASON, University of Wisconsin - Madi, and LALIT JAIN, University of Michigan, ROBERT NOWAK, University of Wisconsin-Madison	4140
The Marginal Value of Adaptive Gradient Methods in Machine Learning, ASHIA WILSON, Berkeley, and REBECCA ROELOFS, MITCHELL STERN, UC Berkeley, NATI SREBRO, TTI Chicago, BEN RECHT, UC Berkeley	4149
Aggressive Sampling for Multi-class to Binary Reduction with Applications to Text Classification, BIKASH JOSHI, UGA, and MASSIH-REZA AMINI, IOANNIS PARTALAS, Expedia LPS Geneva, FRANCK IUTZELER, YURY MAXIMOV, Los Alamos National Laboratory and Skolkovo Institute of Science and Technology	4160

Deconvolutional Paragraph Representation Learning , YIZHE ZHANG, and DINGHAN SHEN, GUOYIN WANG, Duke University, ZHE GAN, duke, RICARDO HENAO, Duke University, LAWRENCE CARIN, duke	4170
Random Permutation Online Isotonic Regression , WOJCIECH KOTLOWSKI, and WOUTER KOOLEN, ALAN MALEK, Massachusetts Institute of Technology	4181
A Unified Game-Theoretic Approach to Multiagent Reinforcement Learning , MARC LANCTOT, DeepMind, and VINICIUS ZAMBALDI, AUDRUNAS GRUSLYS, ANGELIKI LAZARIDOU, DeepMind, KARL TUYLS, JULIEN PEROLAT, DAVID SILVER, THORE GRAEPEL, DeepMind	4191
Inverse Filtering for Hidden Markov Models , ROBERT MATTLILA, and CRISTIAN ROJAS, KTH Royal Inst. of Technology, VIKRAM KRISHNAMURTHY, Cornell University, BO WAHLBERG, KTH Royal Inst. of Technology	4205
Non-parametric Structured Output Networks , ANDREAS LEHRMANN, Disney Research, and LEONID SIGAL	4215
Learning Active Learning from Data , KSENIA KONYUSHKOVA, École polytechnique fédérale de Lausanne, and RAPHAEL SZNITMAN, PASCAL FUA ..	4226
VAE Learning via Stein Variational Gradient Descent , YUNCHEN PU, Duke University, and ZHE GAN, duke, RICARDO HENAO, CHUNYUAN LI, SHAOBO HAN, Duke University, LAWRENCE CARIN, duke	4237
Reconstructing perceived faces from brain activations with deep adversarial neural decoding , YAMUR GÜÇLÜTÜRK, Radboud University, and UMUT GÜÇLÜ, Donders Institute, KATJA SEELIGER, SANDER BOSCH, ROB LIER, MARCEL GERVEN, Radboud University	4247
Efficient Use of Limited-Memory Accelerators for Linear Learning on Heterogeneous Systems , CELESTINE DÜNNER, and THOMAS PARNELL, IBM Research, MARTIN JAGGI	4259
Temporal Coherency based Criteria for Predicting Video Frames using Deep Multi-stage Generative Adversarial Networks , PRATEEP BHATTACHARJEE, and SUKHENDU DAS, IIT Madras	4269
Sobolev Training for Neural Networks , WOJCIECH CZARNECKI, and SIMON OSINDERO, DeepMind, MAX JADERBERG, GRZEGORZ SWIRSZCZ, DeepMind, RAZVAN PASCANU	4279
Multi-Information Source Optimization , MATTHIAS POLOCZEK, Cornell, and JIALEI WANG, International Business Machines (IBM), PETER FRAZIER, Cornell	4289
Deep Reinforcement Learning from Human Preferences , PAUL CHRISTIANO, OpenAI, and JAN LEIKE, DeepMind, TOM BROWN, Google Brain, MILJAN MARTIC, SHANE LEGG, DeepMind, DARIO AMODEI, OpenAI	4300
On the Fine-Grained Complexity of Empirical Risk Minimization: Kernel Methods and Neural Networks , ARTURS BACKURS, and PIOTR INDYK, LUDWIG SCHMIDT, Massachusetts Institute of Technology	4309
Policy Gradient With Value Function Approximation For Collective Multiagent Planning , DUC NGUYEN, and AKSHAT KUMAR, Singapore Management University, HOONG LAU, Singapore Management University	4320

Adversarial Symmetric Variational Autoencoder , YUNCHEN PU, and WEIYAO WANG, RICARDO HENAO, Duke University, LIQUN CHEN, ZHE GAN, duke, CHUNYUAN LI, Duke University, LAWRENCE CARIN, duke	4331
Unified representation of tractography and diffusion-weighted MRI data using sparse multidimensional arrays , CESAR CAIAFA, Indiana University / CONICET, and OLAF SPORNS, Department of Psychological and Brain Sciences - Indiana University, ANDREW SAYKIN, Indiana University, FRANCO PESTILLI, Department of Psychological and Brain Sciences - Indiana University	4341
A Minimax Optimal Algorithm for Crowdsourcing , THOMAS BONALD, Telecom Paristech, and RICHARD COMBES	4353
Estimating Accuracy from Unlabeled Data: A Probabilistic Logic Approach , EMMANOUIL PLATANIOS, Carnegie Mellon University, and HOIFUNG POON, Microsoft Research, TOM MITCHELL, Carnegie Mellon University, ERIC HORVITZ, Microsoft	4362
A Decomposition of Forecast Error in Prediction Markets , MIRO DUDIK, and SÉBASTIEN LAHAIE, Microsoft Research, RYAN ROGERS, JENNIFER VAUGHAN, Microsoft Research	4372
Safe Adaptive Importance Sampling , SEBASTIAN STICH, EPF Lausanne, and ANANT RAJ, Max-Planck Institute, MARTIN JAGGI	4382
Variational Walkback: Learning a Transition Operator as a Stochastic Recurrent Net , ANIRUDH GOYAL, MILA, and NAN KE, Polytechnique Montreal, SURYA GANGULI, Stanford University, YOSHUA BENGIO, U. Montreal	4393
Polynomial Codes: an Optimal Design for High-Dimensional Coded Matrix Multiplication , QIAN YU, University of Southern California, and MOHAMMAD MADDAH-ALI, Nokia Bell Labs, SALMAN AVESTIMEHR, University of Southern California	4404
Unsupervised Learning of Disentangled Representations from Video , EMILY DENTON, and VIGHNESH BIRODKAR, New York University ..	4415
Federated Multi-Task Learning , VIRGINIA SMITH, Massachusetts Institute of Technology, and CHAO-KAI CHIANG, MAZIAR SANJABI, AMEET TALWALKAR, University of California Los Angeles	4425
Is Input Sparsity Time Possible for Kernel Low-Rank Approximation? , CAMERON MUSCO, Massachusetts Institute of Tec, and DAVID WOODRUFF, Carnegie Mellon University	4436
The Expxorcist: Nonparametric Graphical Models Via Conditional Exponential Densities , ARUN SUGGALA, Carnegie Mellon University, and MLADEN KOLAR, University of Chicago, PRADEEP RAVIKUMAR, Carnegie Mellon University	4447
Improved Graph Laplacian via Geometric Self-Consistency , DOMINIQUE PERRAULT-JONCAS, Google, and MARINA MEILA, SAC, JAMES MCQUEEN, University of Washington	4458
Dual Path Networks , YUNPENG CHEN, National University of Singapore, and JIANAN LI, HUAXIN XIAO, NUDT, XIAOJIE JIN, National University of Singapore, SHUICHENG YAN, Qihoo 360 AI Institute, JIASHI FENG, National University of Singapore	4468

Faster and Non-ergodic $O(1/K)$ Stochastic Alternating Direction Method of Multipliers , CONG FANG, and FENG CHENG, ZHOUCHE LIN, Peking University	4477
A Probabilistic Framework for Nonlinearities in Stochastic Neural Networks , QINLIANG SU, and XUEJUN LIAO, Duke University, LAWRENCE CARIN, duke	4487
Distral: Robust multitask reinforcement learning , YEE TEH, and VICTOR BAPST, WOJCIECH CZARNECKI, DeepMind, JOHN QUAN, JAMES KIRKPATRICK, DeepMind, RAIJA HADSELL, NICOLAS HEESS, RAZVAN PASCANU	4497
Online Learning of Optimal Bidding Strategy in Repeated Multi-Commodity Auctions , MUKADDER BALTAOGLU, and LANG TONG, QING ZHAO, Cornell University	4508
Trimmed Density Ratio Estimation , SONG LIU, University of Bristol, and AKIKO TAKEDA, Institute of Statistical Mathematics / RIKEN, TAJI SUZUKI, University of Tokyo/JST-PRESTO/RIKEN AIP, KENJI FUKUMIZU ..	4519
Training recurrent networks to generate hypotheses about how the brain solves hard navigation problems , INGMAR KANITSCHIEDER, and ILA FIETE, University of Texas at Austin	4530
Visual Interaction Networks: Learning a Physics Simulator from Video , NICHOLAS WATTERS, and DANIEL ZORAN, DeepMind, THEOPHANE WEBER, PETER BATTAGLIA, DeepMind, RAZVAN PASCANU, ANDREA TACCHETTI, Massachusetts Institute of Technology	4540
Reconstruct & Crush Network , ERINC MERDIVAN, Austrian Institute of Tech., and MOHAMMAD LOGHMANI, TU Wien, MATTHIEU GEIST, Université de Lorraine	4549
Streaming Robust Submodular Maximization: A Partitioned Thresholding Approach , SLOBODAN MITROVIC, École polytechnique fédérale de Lausanne, and ILIJA BOGUNOVIC, Ecole polytechnique fédérale d, ASHKAN NOROUZI-FARD, JAKUB TARNAWSKI, École polytechnique fédérale de Lausanne, VOLKAN CEVHER	4558
Simple strategies for recovering inner products from coarsely quantized random projections , PING LI, and MARTIN SLAWSKI	4568
Discovering Potential Correlations via Hypercontractivity , HYEJI KIM, and WEIHAO GAO, University of Illinois at Urbana-Champaign, SREERAM KANNAN, University of Washington, SEWOONG OH, PRAMOD VISWANATH, University of Illinois at Urbana-Champaign	4578
Doubly Stochastic Variational Inference for Deep Gaussian Processes , HUGH SALIMBENI, and MARC DEISENROTH, Imperial College London	4589
Ranking Data with Continuous Labels through Oriented Recursive Partitions , STEPHAN CLEMENCON, Telecom ParisTech, and MASTANE ACHAB, Télécom ParisTech	4601
Scalable Model Selection for Belief Networks , ZHAO SONG, Duke University, and YUSUKE MURAOKA, RYOHEI FUJIMAKI, LAWRENCE CARIN, duke	4610

Targeting EEG/LFP Synchrony with Neural Nets, YITONG LI, and MICHAEL MURIAS, SAMANTHA MAJOR, GERALDINE DAWSON, KAFUI DZIRASA, Duke University, LAWRENCE CARIN, duke, DAVID CARLSON, Duke University	4621
Near-Optimal Edge Evaluation in Explicit Generalized Binomial Graphs, SANJIBAN CHOUDHURY, and SHERVIN JAVDANI, SIDDHARTHA SRINIVASA, SEBASTIAN SCHERER, Carnegie Mellon University	4632
Non-Stationary Spectral Kernels, SAMI REMES, Aalto University, and MARKUS HEINONEN, SAMUEL KASKI	4643
Overcoming Catastrophic Forgetting by Incremental Moment Matching, SANG-WOO LEE, and JIN-HWA KIM, JAEHYUN JUN, Seoul National University, JUNG-WOO HA, Clova NAVER Corp. & LINE, BYOUNG-TAK ZHANG, Seoul National University	4653
Balancing information exposure in social networks, KIRAN GARIMELLA, Aalto University, and ARISTIDES GIONIS, NIKOS PAROTSIDIS, University of Rome Tor Vergata, NIKOLAJ TATTI, Aalto University	4664
SafetyNets: Verifiable Execution of Deep Neural Networks on an Untrusted Cloud, ZAHRA GHODSI, and TIANYU GU, New York University, SIDDHARTH GARG, New York Univ	4673
Query Complexity of Clustering with Side Information, ARYA MAZUMDAR, and BARNA SAHA, UMass, Amherst	4683
QMDP-Net: Deep Learning for Planning under Partial Observability, PETER KARKUS, and DAVID HSU, WEE SUN LEE, National University of Singapore	4695
Robust Optimization for Non-Convex Objectives, ROBERT CHEN, Harvard University, and BRENDAN LUCIER, Microsoft Research, YARON SINGER, Harvard University, VASILIS SYRGKANIS	4706
Thy Friend is My Friend: Iterative Collaborative Filtering for Sparse Matrix Estimation, CHRISTIAN BORGS, and JENNIFER CHAYES, Microsoft Research, CHRISTINA LEE, DEVAVRAT SHAH, Massachusetts Institute of Technology	4716
Adaptive Classification for Prediction Under a Budget, NAN FENG, and VENKATESH SALIGRAMA, Boston University	4728
Convergence rates of a partition based Bayesian multivariate density estimation method, LINXI LIU, Columbia University, and DANGNA LI, STANFORD UNIVERSITY, WING WONG, Stanford University	4739
Affine-Invariant Online Optimization and the Low-rank Experts Problem, TOMER KOREN, Google, and ROI LIVNI, Princeton	4748
Beyond Worst-case: A Probabilistic Analysis of Affine Policies in Dynamic Optimization , OMAR HOUSNI, and VINEET GOYAL, Columbia University	4757
A Unified Approach to Interpreting Model Predictions, SCOTT LUNDBERG, and SU-IN LEE, University of Washington	4766
Stochastic Approximation for Canonical Correlation Analysis, RAMAN ARORA, Johns Hopkins University, and TEODOR MARINOV, JHU, POORYA MIANJY, Johns Hopkins University, NATI SREBRO, TTI Chicago	4776

Resurrecting the sigmoid in deep learning through dynamical isometry: theory and practice , JEFFREY PENNINGTON, and SAM SCHOENHOLZ, Google, SURYA GANGULI, Stanford University	4786
Sample and Computationally Efficient Learning Algorithms under S-Concave Distributions , MARIA-FLORINA BALCAN, SAC, and HONGYANG ZHANG, Carnegie Mellon University	4797
Scalable Variational Inference for Dynamical Systems , NICO GORBACH, and STEFAN BAUER, JOACHIM BUHMANN, Eidgenössische Technische Hochschule Zürich	4807
Context Selection for Embedding Models , LIPING LIU, and FRANCISCO RUIZ, Columbia University, SUSAN ATHEY, Stanford University, DAVID BLEI, Columbia University	4817
Working hard to know your neighbor’s margins: Local descriptor learning loss , ANASTASIIA MISHCHUK, Szkocka Research Group Ukraine, and DMYTRO MISHKIN, Czech Technical University, FILIP RADENOVIC, VRG CTU Prague, JIRI MATAS, Czech Technical University	4827
Accelerated Stochastic Greedy Coordinate Descent by Soft Thresholding Projection onto Simplex , CHAOBING SONG, and SHAOBO CUI, Tsinghua University, YONG JIANG, SHU-TAO XIA, Tsinghua University	4839
Multi-Task Learning for Contextual Bandits , ANIKET DESHMUKH, University of Michigan, and URUN DOGAN, CLAYTON SCOTT, University of Michigan	4849
Learning to Prune Deep Neural Networks via Layer-wise Optimal Brain Surgeon , XIN DONG, Nanyang Technological Univ, and SHANGYU CHEN, Nanyang Technological Unvi, SINNO PAN, NTU Singapore	4858
Accelerated First-order Methods for Geodesically Convex Optimization on Riemannian Manifolds , YUANYUAN LIU, and FANHUA SHANG, JAMES CHENG, HONG CHENG, The Chinese University of Hong Kong, LICHENG JIAO, Xidian University	4869
Selective Classification for Deep Neural Networks , YONATAN GEIFMAN, and RAN EL-YANIV, Technion	4879
Minimax Estimation of Bandable Precision Matrices , ADDISON HU, Yale University, and SAHAND NEGAHBAN, Yale	4889
Monte-Carlo Tree Search by Best Arm Identification , EMILIE KAUFMAN, Lille 1/CNRS, and WOUTER KOOLEN	4898
Group Additive Structure Identification for Kernel Nonparametric Regression , CHAO PAN, JP Morgan Chase, and MICHAEL ZHU, Purdue University	4908
Fast, Sample-Efficient Algorithms for Structured Phase Retrieval , GAURI JAGATAP, and CHINMAY HEGDE, Iowa State University	4918
Hash Embeddings for Efficient Word Representations , DAN SVENSTRUP, DTU compute, and JONAS HANSEN, Findzebra, OLE WINTHER, Technical University of Denmark	4929
Online Learning for Multivariate Hawkes Processes , YINGXIANG YANG, and JALAL ETESAMI, NIAO HE, NEGAR KIYAVASH, University of Illinois at Urbana-Champaign	4938

Maximum Margin Interval Trees , ALEXANDRE DROUIN, Université Laval, and TOBY HOCKING, McGill Genome Center McGill University, FRANCOIS LAVIOLETTE	4948
DropoutNet: Addressing Cold Start in Recommender Systems , MAKSIMS VOLKOV, Layer6.ai, and GUANG YU, TOMI POUTANEN	4958
A simple neural network module for relational reasoning , ADAM SANTORO, DeepMind, and DAVID RAPOSO, DAVID BARRETT, MATEUSZ MALINOWSKI, DeepMind, RAZVAN PASCANU, PETER BATTAGLIA, DeepMind, TIMOTHY LILLICRAP, Google DeepMind	4968
Q-LDA: Uncovering Latent Patterns in Text-based Sequential Decision Processes , JIANSHU CHEN, and CHONG WANG, LIN XIAO, JI HE, University of Washington, LIHONG LI, LI DENG	4978
Online Reinforcement Learning in Stochastic Games , CHEN-YU WEI, and YI-TE HONG, CHI-JEN LU, Academia Sinica	4988
Position-based Multiple-play Bandit Problem with Unknown Position Bias , JUNPEI KOMIYAMA, The University of Tokyo, and JUNYA HONDA, The University of Tokyo / RIKEN, AKIKO TAKEDA, Institute of Statistical Mathematics / RIKEN	4999
Active Exploration for Learning Symbolic Representations , GARRETT ANDERSEN, Prowler.io, and GEORGE KONIDARIS, Brown University	5010
Clone MCMC: Parallel High-Dimensional Gaussian Gibbs Sampling , ANDREI-CRISTIAN BARBOS, University of Bordeaux, and FRANCOIS CARON, JEAN-FRANÇOIS GIOVANNELLI, University of Bordeaux, ARNAUD DOUCET	5021
Fair Clustering Through Fairlets , FLAVIO CHIERICHETTI, and RAVI KUMAR, Google, SILVIO LATTANZI, SERGEI VASSILVITSKII, Google	5030
Polynomial time algorithms for dual volume sampling , CHENGTAO LI, and STEFANIE JEGELKA, SUVRIT SRA, Massachusetts Institute of Technology	5039
Hindsight Experience Replay , MARCIN ANDRYCHOWICZ, and FILIP WOLSKI, ALEX RAY, JONAS SCHNEIDER, RACHEL FONG, PETER WELINDER, BOB MCGREW, JOSH TOBIN, OpenAI, PIETER ABBEEL, OpenAI UC Berkeley, WOJCIECH ZAREMBA, OpenAI	5049
Stochastic and Adversarial Online Learning without Hyperparameters , ASHOK CUTKOSKY, and KWABENA BOAHEN, Stanford University	5060
Teaching Machines to Describe Images with Natural Language Feedback , HUAN LING, and SANJA FIDLER, University of Toronto	5069
Perturbative Black Box Variational Inference , ROBERT BAMLER, and CHENG ZHANG, Disney Research, MANFRED OPPER, TU Berlin, STEPHAN MANDT	5080
GibbsNet: Iterative Adversarial Inference for Deep Graphical Models , ALEX LAMB, Université de Montreal, and DEVON HJELM, University of Montreal, YAROSLAV GANIN, MILA, JOSEPH COHEN, Montreal Institute for Learning Algorithms, AARON COURVILLE, SAC, YOSHUA BENGIO, U. Montreal	5090

PointNet++: Deep Hierarchical Feature Learning on Point Sets in a Metric Space , CHARLES QI, and LI YI, HAO SU, LEONIDAS GUIBAS, Stanford University	5100
Regularizing Deep Neural Networks by Noise: Its Interpretation and Optimization , HYEONWOO NOH, and TACKGEUN YOU, JONGHWAN MUN, POSTECH, BOHYUNG HAN	5110
Learning Graph Representations with Embedding Propagation , ALBERTO GARCIA-DURAN, NEC Europe, and MATHIAS NIEPERT	5120
Efficient Modeling of Latent Information in Supervised Learning using Gaussian Processes , ZHENWEN DAI, Amazon.com, and MAURICIO ALVAREZ, The University of Sheffield, NEIL LAWRENCE, Amazon.com	5132
A-NICE-MC: Adversarial Training for MCMC , JIAMING SONG, and SHENGJIA ZHAO, STEFANO ERMON, Stanford University	5141
Excess Risk Bounds for the Bayes Risk using Variational Inference in Latent Gaussian Models , RISHIT SHETH, and RONI KHARDON, Tufts University	5152
Real-Time Bidding with Side Information , ARTHUR FLAJOLET, and PATRICK JAILLET, Massachusetts Institute of Technology	5163
Saliency-based Sequential Image Attention with Multiset Prediction , SEAN WELLECK, and JIALIN MAO, KYUNGHYUN CHO, ZHENG ZHANG, New York University	5174
Variational Inference for Gaussian Process Models with Linear Complexity , CHING-AN CHENG, and BYRON BOOTS, Georgia Tech	5185
K-Medoids For K-Means Seeding , JAMES NEWLING, Idiap Research Institute, and FRANÇOIS FLEURET	5196
Identifying Outlier Arms in Multi-Armed Bandit , HONGLEI ZHUANG, University of Illinois, and CHI WANG, Microsoft Research, YIFAN WANG, Tsinghua University	5205
Online Learning with Transductive Regret , SCOTT YANG, and MEHRYAR MOHRI, New York University	5215
Riemannian approach to batch normalization , MINHYUNG CHO, Gracenote, and JAEHYUNG LEE, Gracenote Inc.	5226
Self-supervised Learning of Motion Capture , HSIAO-YU TUNG, Carnegie Mellon University, and HSIAO-WEI TUNG, University of Pittsburgh, ERSIN YUMER, Adobe Research, KATERINA FRAGKIADAKI, Carnegie Mellon University	5237
Triangle Generative Adversarial Networks , ZHE GAN, and LIQUN CHEN, duke, WEIYAO WANG, YUNCHEN PU, YIZHE ZHANG, Duke University, HAO LIU, Nanjing University, CHUNYUAN LI, Duke University, LAWRENCE CARIN, duke	5248
PRUNE: Preserving Proximity and Global Ranking for Network Embedding , YI-AN LAI, National Taiwan University, and CHIN-CHI HSU, Academia Sinica, WEN CHEN, Student, MI-YEN YEH, Academia Sinica, SHOU-DE LIN	5258
Bayesian Optimization with Gradients , JIAN WU, Cornell University, and MATTHIAS POLOCZEK, Cornell, ANDREW WILSON, Cornell University, PETER FRAZIER, Cornell	5268

Second-order Optimization for Deep Reinforcement Learning using Kronecker-factored Approximation, YUHUI WU, University of Toronto, and ELMAN MANSIMOV, New York University, ROGER GROSSE, SHUN LIAO, JIMMY BA, University of Toronto	5280
Renyi Differential Privacy Mechanisms for Posterior Sampling, JOSEPH GEUMLEK, University of California San Diego, and SHUANG SONG, UC San Diego, KAMALIKA CHAUDHURI, University of California San Diego	5290
Online Learning with a Hint, OFER DEKEL, and ARTHUR FLAJOLET, Massachusetts Institute of Technology, NIKA HAGHTALAB, Carnegie Mellon University, PATRICK JAILLET, Massachusetts Institute of Technology	5300
Identification of Gaussian Process State Space Models, STEFANOS ELEFThERiADIS, and TOM NICHOLSON, PROWLER.io, MARC DEISENROTH, Imperial College London, JAMES HENSMAN, PROWLER.io	5310
Robust Imitation of Diverse Behaviors, ZIYU WANG, and JOSH MEREL, DeepMind, SCOTT REED, Google DeepMind, NANDO FREITAS, Google DeepMind / U. Oxford, GREG WAYNE, NICOLAS HEESS	5321
Can Decentralized Algorithms Outperform Centralized Algorithms? A Case Study for Decentralized Parallel Stochastic Gradient Descent, XIANGRU LIAN, University of Rochester, and CE ZHANG, Eidgenössische Technische Hochschule Zürich, HUAN ZHANG, CHO-JUI HSIEH, University of California, Davis, WEI ZHANG, IBM research, JI LIU	5331
Local Aggregative Games, VIKAS GARG, and TOMMI JAAKKOLA, Massachusetts Institute of Technology	5342
A Sample Complexity Measure with Applications to Learning Optimal Auctions, VASILIS SYRGKANIS	5353
Thinking Fast and Slow with Deep Learning and Tree Search, THOMAS ANTHONY, and ZHENG TIAN, DAVID BARBER, University College London	5361
EEG-GRAPH: A Factor-Graph-Based Model for Capturing Spatial, Temporal, and Observational Relationships in Electroencephalograms, YOGATHEESAN VARATHARAJAH, University of Illinois, and MIN CHONG, KRISHNAKANT SABOO, University of Illinois at Urbana-Champaign, BRENT BERRY, BENJAMIN BRINKMANN, GREGORY WORRELL, RAVISHANKAR IYER, University of Illinois at Urbana-Champaign	5372
Improving the Expected Improvement Algorithm, CHAO QIN, Northwestern University, and DIEGO KLABJAN, Northwestern university, DANIEL RUSSO	5382
Hybrid Reward Architecture for Reinforcement Learning, HARM SEIJEN, Microsoft Research, and MEHDI FATEMI, ROMAIN LAROCHE, JOSH ROMOFF, TAVIAN BARNES, Microsoft, JEFFREY TSANG	5393
Approximate Supermodularity Bounds for Experimental Design, LUIZ CHAMON, and ALEJANDRO RIBEIRO, University of Pennsylvania	5404
Maximizing Subset Accuracy with Recurrent Neural Networks in Multi-label Classification, JINSEOK NAM, and ENELDO MENCÍA, TU Darmstadt, HYUNWOO KIM, University of Wisconsin-Madison, JOHANNES FÜRNKRANZ, TU Darmstadt	5414

AdaGAN: Boosting Generative Models , ILYA TOLSTIKHIN, MPI IS, and SYLVAIN GELLY, Google Brain, OLIVIER BOUSQUET, Google, CARL-JOHANN SIMON-GABRIEL, MPI Tuebingen, BERNHARD SCHÖLKOPF, Max Planck Institute for Intelligent Systems	5425
Straggler Mitigation in Distributed Optimization Through Data Encoding , CAN KARAKUS, University of California Los Angeles, and YIFAN SUN, SUHAS DIGGAVI, WOTAO YIN, University of California Los Angeles	5435
Multi-View Decision Processes: The Helper-AI Problem , CHRISTOS DIMITRAKAKIS, Chalmers / Lille / Harvard, and DAVID PARKES, GORAN RADANOVIC, PAUL TYLKIN, Harvard	5444
A Greedy Approach for Budgeted Maximum Inner Product Search , HSIANG-FU YU, Amazon, and CHO-JUI HSIEH, University of California, Davis, QI LEI, UT Austin, INDERJIT DHILLON, SAC	5454
SVD-Softmax: Fast Softmax Approximation on Large Vocabulary Neural Networks , KYUHONG SHIM, and MINJAE LEE, IKSOO CHOI, YOONHO BOO, WON YONG SUNG, Seoul National University	5464
Plan, Attend, Generate: Planning for Sequence-to-Sequence Models , ÇALAR GÜLÇEHRE, U. Montreal, and FRANCIS DUTIL, MILA Université de Montréal, ADAM TRISCHLER, Microsoft, YOSHUA BENGIO, U. Montreal	5475
Task-based End-to-end Model Learning in Stochastic Optimization , PRIYA DONTI, and ZICO KOLTER, BRANDON AMOS, Carnegie Mellon University	5485
ALICE: Towards Understanding Adversarial Learning for Joint Distribution Matching , CHUNYUAN LI, Duke University, and HAO LIU, Nanjing University, CHANGYOU CHEN, YUNCHEN PU, Duke University, LIQUN CHEN, duke, RICARDO HENAO, Duke University, LAWRENCE CARIN, duke ...	5496
Finite sample analysis of the GTD Policy Evaluation Algorithms in Markov Setting , YUE WANG, School of Science BJTU, and WEI CHEN, Microsoft Research, YUTING LIU, Beijing Jiaotong University, ZHI-MING MA, TIE-YAN LIU, Microsoft Research	5505
On the Complexity of Learning Neural Networks , LE SONG, and SANTOSH VEMPALA, JOHN WILMES, BO XIE, Georgia Tech	5515
Hierarchical Implicit Models and Likelihood-Free Variational Inference , DUSTIN TRAN, Columbia University, and RAJESH RANGANATH, Princeton University, DAVID BLEI, Columbia University	5524
Semi-supervised Learning with GANs: Manifold Invariance with Improved Inference , ABHISHEK KUMAR, and PRASANNA SATTIGERI, IBM Research, P.T. FLETCHER	5535
Approximation and Convergence Properties of Generative Adversarial Learning , SHUANG LIU, University of California San Diego, and OLIVIER BOUSQUET, Google, KAMALIKA CHAUDHURI, University of California San Diego	5546
From Bayesian Sparsity to Gated Recurrent Nets , HAO HE, Massachusetts Institute of Technology, and BO XIN, Microsoft, SATOSHI IKEHATA, National Institute of Informatics, DAVID WIPF	5555

Min-Max Propagation , CHRISTOPHER SRINIVASA, RBC Research, and INMAR GIVONI, University of Toronto, SIAMAK RAVANBAKHS, Carnegie Mellon University, BRENDAN FREY, U. Toronto	5566
What Uncertainties Do We Need in Bayesian Deep Learning for Computer Vision? , ALEX KENDALL, and YARIN GAL, University of Cambridge	5575
Gradient descent GAN optimization is locally stable , VAISHNAVH NAGARAJAN, and ZICO KOLTER, Carnegie Mellon University	5586
Toward Robustness against Label Noise in Training Deep Discriminative Neural Networks , ARASH VAHDAT, D-Wave	5597
Dualing GANs , YUJIA LI, DeepMind, and ALEXANDER SCHWING, University of Illinois at Urbana-Champaign, KUAN-CHIEH WANG, RICHARD ZEMEL, University of Toronto	5607
Deep Learning for Precipitation Nowcasting: A Benchmark and A New Model , XINGJIAN SHI, and ZHIHAN GAO, LEONARD LAUSEN, HAO WANG, HKUST, DIT-YAN YEUNG, WAI-KIN WONG, WANG-CHUN WOO, HKO	5618
Do Deep Neural Networks Suffer from Crowding? , ANNA VOLOKITIN, Eidgenössische Technische Hochschule Zürich, and GEMMA ROIG, TOMASO POGGIO, Massachusetts Institute of Technology	5629
Learning from Complementary Labels , TAKASHI ISHIDA, The University of Tokyo, and GANG NIU, The University of Tokyo / RIKEN, WEIHUA HU, The University of Tokyo/ RIKEN, MASASHI SUGIYAMA, RIKEN/The University of Tokyo	5640
Online control of the false discovery rate with decaying memory , AADITYA RAMDAS, and FAN YANG, MARTIN WAINWRIGHT, UC Berkeley, MICHAEL JORDAN, University of California at Berkeley	5651
Learning from uncertain curves: The 2-Wasserstein metric for Gaussian processes , ANTON MALLASTO, University of Copenhagen, and AASA FERAGEN	5661
Discriminative State Space Models , VITALY KUZNETSOV, and MEHRYAR MOHRI, New York University	5672
On Fairness and Calibration , GEOFF PLEISS, and MANISH RAGHAVAN, FELIX WU, JON KLEINBERG, KILIAN WEINBERGER, Cornell University	5681
Imagination-Augmented Agents for Deep Reinforcement Learning , SEBASTIEN RACANIERE, DeepMind, and THEOPHANE WEBER, DAVID REICHERT, Brown, LARS BUESING, Google DeepMind, ARTHUR GUEZ, DANILO REZENDE, ADRIÀ BADIA, DeepMind, ORIOL VINYALS, NICOLAS HEES, YUJIA LI, DeepMind, RAZVAN PASCANU, PETER BATTAGLIA, DEMIS HASSABIS, DeepMind, DAVID SILVER, DAAN WIERSTRA, DeepMind ..	5691
Extracting low-dimensional dynamics from multiple large-scale neural population recordings by learning to predict correlations , MARCEL NONNENMACHER, research center caesar an associate of the Max Planck Society, and SRINIVAS TURAGA, HHMI Janelia, JAKOB MACKE	5703
Unifying PAC and Regret: Uniform PAC Bounds for Episodic Reinforcement Learning , CHRISTOPH DANN, Carnegie Mellon University, and TOR LATTIMORE, EMMA BRUNSKILL, Stanford University	5714

Gradients of Generative Models for Improved Discriminative Analysis of Tandem Mass Spectra, JOHN HALLORAN, University of California Davi, and DAVID ROCKE, University of California Davis	5725
Asynchronous Parallel Coordinate Minimization for MAP Inference, OFER MESHI, Google, and ALEXANDER SCHWING, University of Illinois at Urbana-Champaign	5735
Multiscale Quantization for Fast Similarity Search, XIANG WU, and RUIQI GUO, ANANDA SURESH, Google, SANJIV KUMAR, DAN HOLTSMANN-RICE, DAVID SIMCHA, Google, FELIX YU	5746
Diverse and Accurate Image Description Using a Variational Auto-Encoder with an Additive Gaussian Encoding Space, LIWEI WANG, University of Illinois, and ALEXANDER SCHWING, SVETLANA LAZEBNIK, University of Illinois at Urbana-Champaign	5757
Improved Training of Wasserstein GANs, ISHAAN GULRAJANI, and FARUK AHMED, MILA, MARTIN ARJOVSKY, New York University, VINCENT DUMOULIN, Université de Montréal, AARON COURVILLE, SAC	5768
Learning Populations of Parameters, KEVIN TIAN, and WEIHAO KONG, GREGORY VALIANT, Stanford University	5779
Clustering with Noisy Queries, ARYA MAZUMDAR, and BARNA SAHA, UMass, Amherst	5789
Higher-Order Total Variation Classes on Grids: Minimax Theory and Trend Filtering Methods, VEERANJANEYULU SADHANALA, and YU-XIANG WANG, Carnegie Mellon University, JAMES SHARPBACK, University of California, Davis, RYAN TIBSHIRANI	5801
Training Quantized Nets: A Deeper Understanding, HAO LI, University of Maryland at Coll, and SOHAM DE, University of Maryland at College Park, ZHENG XU, University of Maryland College Park, CHRISTOPH STUDER, Cornell University, HANAN SAMET, University of Maryland at College Park, TOM GOLDSTEIN, University of Maryland	5812
Permutation-based Causal Inference Algorithms with Interventions, YUHAO WANG, Massachusetts Institute of Technology, and LIAM SOLUS, KTH Royal Institute of Technology, KARREN YANG, CAROLINE UHLER, Massachusetts Institute of Technology	5823
Time-dependent spatially varying graphical models, with application to brain fMRI data analysis, KRISTJAN GREENEWALD, University of Michigan, and SEYOUNG PARK, Yale University, SHUHENG ZHOU, ALEXANDER GIESSING, University of Michigan	5833
Gradient Methods for Submodular Maximization, HAMED HASSANI, UPenn, and MAHDI SOLTANOLKOTABI, University of Southern California, AMIN KARBASI, Yale University	5842
Smooth Primal-Dual Coordinate Descent Algorithms for Nonsmooth Convex Optimization, AHMET ALACAOGLU, École polytechnique fédérale de Lausanne, and QUOC TRAN-DINH, OLIVIER FERCOQ, VOLKAN CEVHER	5853
The Importance of Communities for Learning to Influence, ERIC BALKANSKI, Harvard, and NICOLE IMMORLICA, Microsoft Research, YARON SINGER, Harvard University	5863

Multiplicative Weights Update with Constant Step-Size in Congestion Games: Convergence, Limit Cycles and Chaos, GERASIMOS PALAIOPANOS, SUTD, and IOANNIS PANAGEAS, MIT & SUTD, GEORGIOS PILIOURAS, SUTD	5873
Learning Neural Representations of Human Cognition across Many fMRI Studies, ARTHUR MENSCH, Inria, and JULIEN MAIRAL, DANILO BZDOK, RWTH Aachen University, BERTRAND THIRION, GAEL VAROQUAUX	5884
A KL-LUCB algorithm for Large-Scale Crowdsourcing, ERVIN TANCZOS, University of Wisconsin, and ROBERT NOWAK, University of Wisconsin-Madison, BOB MANKOFF, Former Cartoon Editor of The New Yorker	5895
Collaborative Deep Learning in Fixed Topology Networks, ZHANHONG JIANG, and ADITYA BALU, CHINMAY HEGDE, SOUMIK SARKAR, Iowa State University	5905
Fast-Slow Recurrent Neural Networks, ASIER MUJIKI, ETH Zürich, and FLORIAN MEIER, ANGELIKA STEGER, Eidgenössische Technische Hochschule Zürich	5916
Learning Disentangled Representations with Semi-Supervised Deep Generative Models, SIDDHARTH NARAYANASWAMY, University of Oxford, and BROOKS PAIGE, Alan Turing Institute, JAN-WILLEM MEENT, Northeastern University, ALBAN DESMAISON, University of Oxford, NOAH GOODMAN, Stanford University, PUSHMEET KOHLI, DeepMind, FRANK WOOD, PHILIP TORR, University of Oxford	5926
Self-Supervised Intrinsic Image Decomposition, MICHAEL JANNER, and JIAJUN WU, Massachusetts Institute of Technology, TEJAS KUKARNI, DeepMind, ILKER YILDIRIM, JOSH TENENBAUM, Massachusetts Institute of Technology	5937
Exploring Generalization in Deep Learning, BEHNAMEH NEYSHABUR, TTI-Chicago, and SRINADH BHOJANAPALLI, DAVID MCALLESTER, NATI SREBRO, TTI Chicago	5948
A framework for Multi-A(rmed)/B(audit) Testing with Online FDR Control, FAN YANG, and AADITYA RAMDAS, KEVIN JAMIESON, MARTIN WAINWRIGHT, UC Berkeley	5958
Fader Networks: Manipulating Images by Sliding Attributes, GUILLAUME LAMPLE, Facebook AI Research, and NEIL ZEGHIDOUR, facebook, NICOLAS USUNIER, ANTOINE BORDES, SAC, LUDOVIC DENOYER, MARC'AURELIO RANZATO, SAC	5968
Action Centered Contextual Bandits, KRISTJAN GREENEWALD, and AMBUJ TEWARI, SUSAN MURPHY, PREDAG KLASNJA, University of Michigan	5978
Estimating Mutual Information for Discrete-Continuous Mixtures, WEIHAO GAO, University of Illinois at Urbana-Champaign, and SREERAM KANNAN, University of Washington, SEWOONG OH, PRAMOD VISWANATH, University of Illinois at Urbana-Champaign	5987
Attention is All you Need, ASHISH VASWANI, and NOAM SHAZEER, NIKI PARMAR, JAKOB USZKOREIT, LILION JONES, Google, AIDAN GOMEZ, University of Toronto, LUKASZ KAISER, Google, ILLIA POLOSUKHIN	5999

Recurrent Ladder Networks , ISABEAU PRÉMONT-SCHWARZ, and ALEXANDER ILIN, TELE HAO, The Curious AI Company, ANTTI RASMUS, RINU BONEY, HARRI VALPOLA, The Curious AI Company	6010
Parameter-Free Online Learning via Model Selection , DYLAN FOSTER, Cornell University, and SATYEN KALE, Google, MEHRYAR MOHRI, New York University, KARTHIK SRIDHARAN, Cornell University	6021
Bregman Divergence for Stochastic Variance Reduction: Saddle-Point and Adversarial Prediction , ZHAN SHI, and XINHUA ZHANG, University of Illinois Chicago, YAOLIANG YU	6032
Unbounded cache model for online language modeling with open vocabulary , EDOUARD GRAVE, and MOUSTAPHA CISSE, Facebook AI Research, ARMAND JOULIN	6043
Predictive State Recurrent Neural Networks , CARLTON DOWNEY, Carnegie Mellon University, and AHMED HEFNNY, MLD Carnegie Mellon University, BYRON BOOTS, Georgia Tech, GEOFFREY GORDON, BOYUE LI, Carnegie Mellon University	6054
Early stopping for kernel boosting algorithms: A general analysis with localized complexities , YUTING WEI, UC BERKELEY, and FAN YANG, MARTIN WAINWRIGHT, UC Berkeley	6066
SVCCA: Singular Vector Canonical Correlation Analysis for Deep Learning Dynamics and Interpretability , MAITHRA RAGHU, Cornell, and JUSTIN GILMER, Google, JASON YOSINSKI, Uber, JASCHA SOHL-DICKSTEIN	6077
Convolutional Phase Retrieval , QING QU, and YUQIAN ZHANG, Columbia University, YONINA ELДАР, Israel Institute of Technology, JOHN WRIGHT, Columbia University	6087
Estimating High-dimensional Non-Gaussian Multiple Index Models via Stein’s Lemma , ZHUORAN YANG, Princeton University, and KRISHNAKUMAR BALASUBRAMANIAN, ZHAORAN WANG, Princeton Phd student, HAN LIU, Princeton University	6098
Gaussian Quadrature for Kernel Features , TRI DAO, and CHRISTOPHER SA, CHRISTOPHER RE, Stanford University	6108
Value Prediction Network , JUNHYUK OH, University of Michigan, and SATINDER SINGH, UMICH, HONGLAK LEE	6119
A Learning Error Analysis for Structured Prediction with Approximate Inference , YUANBIN WU, and MAN LAN, SHILIANG SUN, East China Normal University, QI ZHANG, XUANJING HUANG, Fudan University	6130
Efficient Second-Order Online Kernel Learning with Adaptive Embedding , DANIELE CALANDRIELLO, INRIA Lille - Nord Europe, and ALESSANDRO LAZARIC, INRIA, MICHAL VALKO, Inria Lille - Nord Europe ..	6141
Implicit Regularization in Matrix Factorization , SURIYA GUNASEKAR, TTIC, and BLAKE WOODWORTH, SRINADH BHOJANAPALLI, TTI Chicago, BEHNAM NEYSHABUR, TTI-Chicago, NATI SREBRO, TTI Chicago	6152
Optimal Shrinkage of Singular Values Under Random Data Contamination , DANNY BARASH, Hebrew University, and MATAN GAVISH	6161

Countering Feedback Delays in Multi-Agent Learning , ZHENGYUAN ZHOU, Stanford University, and PANAYOTIS MERTIKOPOULOS, NICHOLAS BAMBOS, PETER GLYNN, Stanford University, CLAIRE TOMLIN, EECS UC Berkeley	6172
Asynchronous Coordinate Descent under More Realistic Assumptions , TAO SUN, National university of defense technology, and ROBERT HANNAH, WOTAO YIN, University of California Los Angeles	6183
Linear Convergence of a Frank-Wolfe Type Algorithm over Trace-Norm Balls , ZEYUAN ALLEN-ZHU, Microsoft Research, and ELAD HAZAN, WEI HU, YUANZHI LI, Princeton University	6192
Hierarchical Clustering Beyond the Worst-Case , VINCENT COHEN-ADDAD, University of Copenhagen, and VARUN KANADE, FREDERIK MALLMANN-TRENN, École normale supérieure	6202
Invariance and Stability of Deep Convolutional Representations , ALBERTO BIETTI, Inria, and JULIEN MAIRAL	6211
Statistical Cost Sharing , ERIC BALKANSKI, Harvard, and UMAR SYED, SERGEI VASSILVITSKII, Google	6222
The Expressive Power of Neural Networks: A View from the Width , LU ZHOU, Peking Univ, and HONGMING PU, FEICHENG WANG, ZHIQIANG HU, Peking University, LIWEI WANG	6232
Spectrally-normalized margin bounds for neural networks , PETER BARTLETT, UC Berkeley, and DYLAN FOSTER, Cornell University, MATUS TELGARSKY, University of Illinois at Urbana-Champaign	6241
Robust and Efficient Transfer Learning with Hidden Parameter Markov Decision Processes , TAYLOR KILLIAN, and SAMUEL DAULTON, FINALE DOSHI-VELEZ, Harvard University, GEORGE KONIDARIS, Brown University	6251
Population Matching Discrepancy and Applications in Deep Learning , JIANFEI CHEN, and CHONGXUAN LI, YIZHONG RU, JUNE ZHU, Tsinghua University	6263
Scalable Planning with Tensorflow for Hybrid Nonlinear Domains , GA WU, and BUSER SAY, SCOTT SANNER, University of Toronto	6274
Boltzmann Exploration Done Right , NICOLÒ CESA-BIANCHI, and CLAUDIO GENTILE, SAC, GERGELY NEU, GÁBOR LUGOSI, Universitat Pompeu Fabra	6285
Learned in Translation: Contextualized Word Vectors , BRYAN MCCANN, and JAMES BRADBURY, Salesforce Research, CAIMING XIONG, Salesforce Research Metamind, RICHARD SOCHER	6295
Neural Discrete Representation Learning , AARON OORD, Google DeepMind, and ORIOL VINYALS, KORAY KAVUKCUOGLU	6307
Generalizing GANs: A Turing Perspective , RODERICH GROSS, and YUE GU, The University of Sheffield, WEI LI, University of York, MELVIN GAUCI, Harvard University	6317
Scalable Log Determinants for Gaussian Process Kernel Learning , KUN DONG, and DAVID ERIKSSON, Cornell University, HANNES NICKISCH, DAVID BINDEL, ANDREW WILSON, Cornell University	6328

Poincaré Embeddings for Learning Hierarchical Representations, MAXIMILLIAN NICKEL, and DOUWE KIELA, Facebook	6339
Learning Combinatorial Optimization Algorithms over Graphs, ELIAS KHALIL, and HANJUN DAI, YUYU ZHANG, BISTRA DILKINA, LE SONG, Georgia Tech	6349
Robust Conditional Probabilities, YOAV WALD, Hebrew University of Jerusalem, and AMIR GLOBERSON, Tel Aviv University Google	6360
Learning with Bandit Feedback in Potential Games, AMÉLIE HÉLIOU, Ecole Polytechnique, and JOHANNE COHEN, LRI-CNRS, PANAYOTIS MERTIKOPOULOS	6370
Multi-Agent Actor-Critic for Mixed Cooperative-Competitive Environments, RYAN LOWE, McGill University, and YI WU, AVIV TAMAR, UC Berkeley, JEAN HARB, McGill University, PIETER ABBEEL, OpenAI UC Berkeley, IGOR MORDATICH, OpenAI	6380
Communication-Efficient Distributed Learning of Discrete Distributions, ILIAS DIAKONIKOLAS, University of Southern California, and ELENA GRIGORESCU, Purdue University, JERRY LI, Massachusetts Institute of Technology, ABHIRAM NATARAJAN, Purdue University, KRZYSZTOF ONAK, IBM T.J. Watson Research Center, LUDWIG SCHMIDT, Massachusetts Institute of Technology	6392
Simple and Scalable Predictive Uncertainty Estimation using Deep Ensembles, BALAJI LAKSHMINARAYANAN, Google Deepmind, and ALEXANDER PRITZEL, CHARLES BLUNDELL	6403
When Worlds Collide: Integrating Different Counterfactual Assumptions in Fairness, CHRIS RUSSELL, and MATT KUSNER, JOSHUA LOFTUS, The Alan Turing Institute, RICARDO SILVA, ucl.ac.uk	6415
Matrix Norm Estimation from a Few Entries, ASHISH KHETAN, and SEWOONG OH, University of Illinois at Urbana-Champaign	6425
Neural Networks for Efficient Bayesian Decoding of Natural Images from Retinal Neurons, NIKHIL PARTHASARATHY, Stanford University, and ELEANOR BATTY, WILLIAM FALCON, THOMAS RUTTEN, MOHIT RAJPAL, Columbia University, EJ CHICHILNISKY, Stanford University, LIAM PANINSKI	6435
Causal Effect Inference with Deep Latent-Variable Models, CHRISTOS LOUIZOS, University of Amsterdam, and URI SHALIT, New York University, JORIS MOOIJ, University of Amsterdam, DAVID SONTAG, Massachusetts Institute of Technology, RICHARD ZEMEL, University of Toronto, MAX WELLING, University of Amsterdam and University of California Irvine and CIFAR	6447
Learning Identifiable Gaussian Bayesian Networks in Polynomial Time and Sample Complexity, ASISH GHOSHAL, and JEAN HONORIO, Purdue University	6458
Gradient Episodic Memory for Continual Learning, DAVID LOPEZ-PAZ, Facebook AI Research, and MARC'AURELIO RANZATO, SAC	6468
Effective Parallelisation for Machine Learning, MICHAEL KAMP, Fraunhofer IAIS, and MARIO BOLEY, Cluster of Excellence MMCI and Saarland University, OLANA MISSURA, Google Inc, THOMAS GAERTNER	6478

Semisupervised Clustering, AND-Queries and Locally Encodable Source Coding, ARYA MAZUMDAR, UMass, Amherst, and SOUMYABRATA PAL	6490
Clustering Stable Instances of Euclidean k-means., ARAVINDAN VIJAYARAGHAVAN, Northwestern university, and ABHRATANU DUTTA, ALEX WANG, Northwestern University	6501
Good Semi-supervised Learning That Requires a Bad GAN, ZIHANG DAI, and ZHILIN YANG, FAN YANG, WILLIAM COHEN, RUSLAN SALAKHUTDINOV, Carnegie Mellon University	6511
On Blackbox Backpropagation and Jacobian Sensing, KRZYSZTOF CHOROMANSKI, and VIKAS SINDHWANI	6522
Protein Interface Prediction using Graph Convolutional Networks, ALEX FOUT, and JONATHON BYRD, BASIR SHARIAT, ASA BEN-HUR, Colorado State University	6531
Solid Harmonic Wavelet Scattering: Predicting Quantum Molecular Energy from Invariant Descriptors of 3D Electronic Densities, MICHAEL EICKENBERG, Ecole normale superieure, and GEORGIOS EXARCHAKIS, École normale supérieure, MATTHEW HIRN, Michigan State University, STEPHANE MALLAT, Ecole normale superieure	6541
Towards Generalization and Simplicity in Continuous Control, ARAVIND RAJESWARAN, and KENDALL LOWREY, EMO TODOROV, SHAM KAKADE, University of Washington	6551
Random Projection Filter Bank for Time Series Data, AMIR-MASSOUD FARAHMAND, and SEPIDEH POURAZARM, Mitsubishi Electric Research Laboratories, DANIEL NIKOVSKI	6563
Filtering Variational Objectives, CHRIS MADDISON, University of Oxford, and DIETERICH LAWSON, GEORGE TUCKER, Google Brain, NICOLAS HEESS, MOHAMMAD NOROUZI, ANDRIY MNIH, DeepMind, ARNAUD DOUCET, YEE TEH, DeepMind	6574
On Frank-Wolfe and Equilibrium Computation, JACOB ABERNETHY, and JUN-KUN WANG, University of Michigan	6585
Modulating early visual processing by language, HARM VRIES, University of Montreal, and FLORIAN STRUB, University of Lille, JEREMIE MARY, HUGO LAROCHELLE, SAC, OLIVIER PIETQUIN, DeepMind, AARON COURVILLE, SAC	6595
Learning Mixture of Gaussians with Streaming Data, ADITI RAGHUNATHAN, Stanford University, and PRATEEK JAIN, RAVISHANKAR KRISHNAWAMY, Microsoft Research India	6606
Practical Hash Functions for Similarity Estimation and Dimensionality Reduction, SØREN DAHLGAARD, and MATHIAS KNUDSEN, MIKKEL THORUP, University of Copenhagen	6616
GANs Trained by a Two Time-Scale Update Rule Converge to a Local Nash Equilibrium, MARTIN HEUSEL, LIT AI Lab / University Linz, and HUBERT RAMSAUER, Johannes Kepler University Linz, THOMAS UNTERTHINER, BERNHARD NESSLER, Johannes Kepler University Linz, SEPP HOCHREITER, LIT AI Lab / University Linz	6627

The Scaling Limit of High-Dimensional Online Independent Component Analysis, CHUANG WANG, and YUE LU, Harvard University	6639
Approximation Algorithms for ℓ_0-Low Rank Approximation, KARL BRINGMANN, and PAVEL KOLEV, MPI-INF, DAVID WOODRUFF, Carnegie Mellon University	6649
The power of absolute discounting: all-dimensional distribution estimation, MOEIN FALAHATGAR, University of California San Diego, and MESROB OHANNESSIAN, ALON ORLITSKY, VENKATADHEERAJ PICHAPATI, University of California San Diego	6661
Few-Shot Adversarial Domain Adaptation, SAEID MOTIHAN, and QUINN JONES, SEYED IRANMANESH, GIANFRANCO DORETTO, West Virginia University	6671
Spectral Mixture Kernels for Multi-Output Gaussian Processes, GABRIEL PARRA, and FELIPE TOBAR, Universidad de Chile	6682
Neural Expectation Maximization, KLAUS GREFF, and SJOERD STEENKISTE, IDSIA, JÜRGEN SCHMIDHUBER	6692
Learning Linear Dynamical Systems via Spectral Filtering, ELAD HAZAN, and KARAN SINGH, Princeton University, CYRIL ZHANG, Princeton University CS	6703
Z-Forcing: Training Stochastic Recurrent Networks, ANIRUDH GOYAL, MILA, and ALESSANDRO SORDONI, Maluuba Microsoft, MARC-ALEXANDRE CÔTÉ, Microsoft Maluuba, NAN KE, Polytechnique Montreal, YOSHUA BENGIO, U. Montreal	6714
Learning Hierarchical Information Flow with Recurrent Neural Modules, DANIJAR HAFNER, and ALEXANDER IRPAN, JAMES DAVIDSON, Google, NICOLAS HEESS	6725
Neural Variational Inference and Learning in Undirected Graphical Models, VOLODYMYR KULESHOV, and STEFANO ERMON, Stanford University	6735
Subspace Clustering via Tangent Cones, AMIN JALALI, Univ. of Wisconsin-Madison, and REBECCA WILLETT, UW-Madison	6745
The Neural Hawkes Process: A Neurally Self-Modulating Multivariate Point Process, HONGYUAN MEI, and JASON EISNER, JOHNS HOPKINS UNIVERSITY	6755
Inverse Reward Design, DYLAN HADFIELD-MENELL, and SMITHA MILLI, UC Berkeley, PIETER ABBEEL, STUART RUSSELL, ANCA DRAGAN, University of California at Berkeley	6766
Structured Bayesian Pruning via Log-Normal Multiplicative Noise, KIRILL NEKLYUDOV, MIPT, and DMITRY MOLCHANOV, ARSENIY ASHUKHA, DMITRY VETROV	6776
Attend and Predict: Understanding Gene Regulation by Selective Attention on Chromatin, RITAMBHARA SINGH, University of Virginia, and JACK LANCHANTIN, UVA, ARSHDEEP SEKHON, YANJUN QI, University of Virginia	6786
Acceleration and Averaging in Stochastic Descent Dynamics, WALID KRICHENE, Google, and PETER BARTLETT, UC Berkeley	6797

Kernel functions based on triplet comparisons, MATTHÄUS KLEINDESSNER, and ULRIKE LUXBURG, University of Tübingen	6808
An Error Detection and Correction Framework for Connectomics, JONATHAN ZUNG, and IGNACIO TARTAVULL, Princeton University, KISUK LEE, Massachusetts Institute of Tec, SEBASTIAN SEUNG, Princeton University	6819
Style Transfer from Non-Parallel Text by Cross-Alignment, TIANXIAO SHEN, and TAO LEI, Massachusetts Institute of Technology, REGINA BARZILAY, MIT CSAIL, TOMMI JAAKKOLA, Massachusetts Institute of Technology	6831
Cross-Spectral Factor Analysis, NEIL GALLAGHER, Duke University, and KYLE ULRICH, AUSTIN TALBOT, KAFUI DZIRASA, Duke University, LAWRENCE CARIN, duke, DAVID CARLSON, Duke University	6843
Stochastic Submodular Maximization: The Case of Coverage Functions, MOHAMMAD KARIMI, and MARIO LUCIC, Eidgenössische Technische Hochschule Zürich, HAMED HASSANI, UPenn, ANDREAS KRAUSE	6854
Affinity Clustering: Hierarchical Clustering at Scale, MOHAMMADHOSSEIN BATENI, Google research, and SOHEIL BEHNEZHAD, MAHSA DERAKHSHAN, MOHAMMADTAGHI HAJIAGHAYI, University of Maryland, RAIMONDAS KIVERIS, Google research, SILVIO LATTANZI, VAHAB MIRROKNI, Google	6865
Unsupervised Transformation Learning via Convex Relaxations, TATSUNORI HASHIMOTO, and PERCY LIANG, JOHN DUCHI, Stanford University	6876
A Sharp Error Analysis for the Fused Lasso, with Application to Approximate Change-point Screening, KEVIN LIN, Carnegie Mellon University, and JAMES SHARPBACK, University of California, Davis, ALESSANDRO RINALDO, Carnegie Mellon University, RYAN TIBSHIRANI	6885
Linear Time Computation of Moments in Sum-Product Networks, HAN ZHAO, and GEOFFREY GORDON, Carnegie Mellon University	6895
A Meta-Learning Perspective on Cold-Start Recommendations for Items, MANASI VARTAK, Massachusetts Institute of Technology, and ARVIND THIAGARAJAN, CONRADO MIRANDA, JESHUA BRATMAN, HUGO LAROCHELLE, SAC	6905
Predicting Scene Parsing and Motion Dynamics in the Future, XIAOJIE JIN, National University of Singapore, and HUAXIN XIAO, NUDT, XIAOHUI SHEN, Adobe, JIMEI YANG, ZHE LIN, Adobe, YUNPENG CHEN, National University of Singapore, ZEQUAN JIE, Tencent AI Lab, JIASHI FENG, National University of Singapore, SHUICHENG YAN, Qihoo 360 AI Institute	6916
Sticking the Landing: Simple, Lower-Variance Gradient Estimators for Variational Inference, GEOFFREY ROEDER, and YUHUI WU, DAVID DUVENAUD, University of Toronto	6926
Efficient Approximation Algorithms for Strings Kernel Based Sequence Classification, MUHAMMAD FARHAN, LUMS, and JUVARIA TARIQ, Emory Univeristy, ARIF ZAMAN, LUMS, MUDASSIR SHABBIR, ITU, IMDAD KHAN, LUMS	6936
Kernel Feature Selection via Conditional Covariance Minimization, JIANBO CHEN, and MITCHELL STERN, MARTIN WAINWRIGHT, UC Berkeley, MICHAEL JORDAN, University of California at Berkeley	6947

Convergence of Gradient EM on Multi-component Mixture of Gaussians, BOWEI YAN, and MINGZHANG YIN, University of Texas at Austin, PURNAMRITA SARKAR, UTexas	6957
Real Time Image Saliency for Black Box Classifiers, PIOTR DABKOWSKI, and YARIN GAL, University of Cambridge	6968
Houdini: Fooling Deep Structured Visual and Speech Recognition Models with Adversarial Examples, MOUSTAPHA CISSE, Facebook AI Research, and YOSSI ADI, Bar-Ilan University, NATALIA NEVEROVA, Facebook AI Research, JOSEPH KESHET, Bar-Ilan University	6978
Efficient and Flexible Inference for Stochastic Systems, STEFAN BAUER, Eidgenössische Technische Hochschule Zürich, and NICO GORBACH, DJORDJE MILADINOVIC, JOACHIM BUHMANN, Eidgenössische Technische Hochschule Zürich	6989
When Cyclic Coordinate Descent Outperforms Randomized Coordinate Descent, MERT GUERBUEZBALABAN, and ASUMAN OZDAGLAR, PABLO PARRILO, DENIZCAN VANLI, Massachusetts Institute of Technology	7000
Active Learning from Peers, KEERTHIRAM MURUGESAN, and JAIME CARBONELL, Carnegie Mellon University	7009
Experimental Design for Learning Causal Graphs with Latent Variables, MURAT KOCAOGLU, UT Austin, and KARTHIKEYAN SHANMUGAM, IBM Research, ELIAS BAREINBOIM, Purdue University	7019
Learning to Model the Tail, YU-XIONG WANG, and DEVA RAMANAN, MARTIAL HEBERT, Carnegie Mellon University	7030
Stochastic Mirror Descent in Variationally Coherent Optimization Problems, ZHENGYUAN ZHOU, Stanford University, and PANAYOTIS MERTIKOPOULOS, NICHOLAS BAMBOS, STEPHEN BOYD, PETER GLYNN, Stanford University	7041
On Separability of Loss Functions, and Revisiting Discriminative Vs Generative Models, ADARSH PRASAD, Carnegie Mellon University, and ALEXANDRU NICULESCU-MIZIL, PRADEEP RAVIKUMAR, Carnegie Mellon University	7051
Maxing and Ranking with Few Assumptions, MOEIN FALAHATGAR, and YI HAO, ALON ORLITSKY, VENKATADHEERAJ PICHAPATI, University of California San Diego, VAISHAKH RAVINDRAKUMAR, UC San Diego	7061
On clustering network-valued data, SOUMENDU MUKHERJEE, UC Berkeley, and PURNAMRITA SARKAR, LIZHEN LIN, UTexas	7072
A General Framework for Robust Interactive Learning, EHSAN EMAMJOMEH-ZADEH, and DAVID KEMPE, U. of Southern California	7083
Multi-view Matrix Factorization for Linear Dynamical System Estimation, MAHDI KARAMI, University of Alberta, and MARTHA WHITE, Department of Psychological and Brain Sciences - Indiana University, DALE SCHUURMANS, University of Alberta & Google Brain, CSABA SZEPESVARI, University of Alberta	7093
Author Index	