

30th International Conference on Machine Learning

(ICML 2013)

Journal of Machine Learning Research Volume 28

**Atlanta, Georgia, USA
16-21 June 2013**

Part 1 of 3

Editors:

Sanjoy Dasgupta

David McAllester

ISBN: 978-1-62993-306-1

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© (2013) by the International Machine Learning Society (IMLS)
All rights reserved.

Printed by Curran Associates, Inc. (2013)

For permission requests, please contact the International Machine Learning Society (IMLS)
at the address below.

International Machine Learning Society (IMLS)
c/o Priscilla Rasmussen
209 North Eighth Street
Stroudsburg PA 18360

Phone: (570) 476-8006
Fax: (570) 476-0860

rasmussen@ptd.net

Additional copies of this publication are available from:

Curran Associates, Inc.
57 Morehouse Lane
Red Hook, NY 12571 USA
Phone: 845-758-0400
Fax: 845-758-2634
Email: curran@proceedings.com
Web: www.proceedings.com

TABLE OF CONTENTS

PART 1

An Optimal Policy for Target Localization with Application to Electron Microscopy	1
<i>Raphael Sznitman, Aurelien Lucchi, Peter Frazier, Bruno Jedynak, Pascal Fua</i>	
Domain Generalization via Invariant Feature Representation	10
<i>Krikamol Muandet, David Balduzzi, Bernhard Schölkopf</i>	
A Spectral Learning Approach to Range-Only SLAM	19
<i>Byron Boots, Geoff Gordon</i>	
Near-Optimal Bounds for Cross-Validation via Loss Stability	27
<i>Ravi Kumar, Daniel Lokshtanov, Sergei Vassilvitskii, Andrea Vattani</i>	
Sparsity-Based Generalization Bounds for Predictive Sparse Coding	36
<i>Nishant Mehta, Alexander Gray</i>	
Sparse Uncorrelated Linear Discriminant Analysis	45
<i>Xiaowei Zhang, Delin Chu</i>	
Block-Coordinate Frank-Wolfe Optimization for Structural SVMs	53
<i>Simon Lacoste-Julien, Martin Jaggi, Mark Schmidt, Patrick Pletscher</i>	
Fast Probabilistic Optimization from Noisy Gradients	62
<i>Philipp Hennig</i>	
Stochastic Gradient Descent for Non-smooth Optimization: Convergence Results and Optimal Averaging Schemes	71
<i>Ohad Shamir, Tong Zhang</i>	
Stochastic Alternating Direction Method of Multipliers	80
<i>Hua Ouyang, Niao He, Long Tran, Alexander Gray</i>	
Noisy Sparse Subspace Clustering	89
<i>Yu-Xiang Wang, Huan Xu</i>	
Parallel Markov Chain Monte Carlo for Nonparametric Mixture Models	98
<i>Sinead Williamson, Avinava Dubey, Eric Xing</i>	
Risk Bounds and Learning Algorithms for the Regression Approach to Structured Output Prediction	107
<i>Sébastien Giguère, François Laviolette, Mario Marchand, Khadija Sylla</i>	
Making a Science of Model Search: Hyperparameter Optimization in Hundreds of Dimensions for Vision Architectures	115
<i>James Bergstra, Daniel Yamins, David Cox</i>	
Gibbs Max-Margin Topic Models with Fast Sampling Algorithms	124
<i>Jun Zhu, Ning Chen, Hugh Perkins, Bo Zhang</i>	
Cost-Sensitive Tree of Classifiers	133
<i>Zhixiang Xu, Matt Kusner, Kilian Weinberger, Minmin Chen</i>	
Learning Hash Functions Using Column Generation	142
<i>Xi Li, Guosheng Lin, Chunhua Shen, Anton Van den Hengel, Anthony Dick</i>	
Combinatorial Multi-Armed Bandit: General Framework and Applications	151
<i>Wei Chen, Yajun Wang, Yang Yuan</i>	
Near-optimal Batch Mode Active Learning and Adaptive Submodular Optimization	160
<i>Yuxin Chen, Andreas Krause</i>	
Convex Formulations of Radius-margin based Support Vector Machines	169
<i>Huyen Do, Alexandros Kalousis</i>	
Modelling Sparse Dynamical Systems with Compressed Predictive State Representations	178
<i>William L. Hamilton, Mahdi Milani Fard, Joelle Pineau</i>	
A Machine Learning Framework for Programming by Example	187
<i>Aditya Menon, Omer Tamuz, Sumit Gulwani, Butler Lampson, Adam Kalai</i>	
Discriminatively Activated Sparselets	196
<i>Ross Girshick, Hyun Oh Song, Trevor Darrell</i>	
The Pairwise Piecewise-Linear Embedding for Efficient Non-Linear Classification	205
<i>Ofir Pele, Ben Taskar, Amir Globerson, Michael Werman</i>	
Fixed-Point Model For Structured Labeling	214
<i>Quannan Li, Jingdong Wang, David Wipf, Zhuowen Tu</i>	
Connecting the Dots with Landmarks: Discriminatively Learning Domain-Invariant Features for Unsupervised Domain Adaptation	222
<i>Boqing Gong, Kristen Grauman, Fei Sha</i>	
Fast Conical Hull Algorithms for Near-separable Non-negative Matrix Factorization	231
<i>Abhishek Kumar, Vikas Sindhwani, Prabhanjan Kambadur</i>	
Principal Component Analysis on Non-Gaussian Dependent Data	240
<i>Fang Han, Han Liu</i>	
Learning Linear Bayesian Networks with Latent Variables	249
<i>Animashree Anandkumar, Daniel Hsu, Adel Javanmard, Sham Kakade</i>	
Multiple Identifications in Multi-Armed Bandits	258
<i>Sébastien Bubeck, Tengyao Wang, Nitin Viswanathan</i>	
Learning Optimally Sparse Support Vector Machines	266
<i>Andrew Cotter, Shai Shalev-Shwartz, Nati Srebro</i>	

Dynamic Probabilistic Models for Latent Feature Propagation in Social Networks	275
<i>Creighton Heaukulani, Zoubin Ghahramani</i>	
Efficient Sparse Group Feature Selection via Nonconvex Optimization	284
<i>Shuo Xiang, Xiaoshen Tong, Jieping Ye</i>	
Domain Adaptation for Sequence Labeling Tasks with a Probabilistic Language Adaptation Model	293
<i>Min Xiao, Yuhong Guo</i>	
Maximum Variance Correction with Application to A* Search	302
<i>Wenlin Chen, Kilian Weinberger, Yixin Chen</i>	
Adaptive Sparsity in Gaussian Graphical Models	311
<i>Eleanor Wong, Suyash Awate, P. Thomas Fletcher</i>	
Average Reward Optimization Objective In Partially Observable Domains	320
<i>Yuri Grinberg, Doina Precup</i>	
Feature Selection in High-Dimensional Classification	329
<i>Mladen Kolar, Han Liu</i>	
Human Boosting	338
<i>Harsh Pareek, Pradeep Ravikumar</i>	
Efficient Dimensionality Reduction for Canonical Correlation Analysis	347
<i>Haim Avron, Christos Boutsidis, Sivan Toledo, Anastasios Zouzias</i>	
Parsing Epileptic Events using a Markov Switching Process Model for Correlated Time Series	356
<i>Drausin Wulsin, Emily Fox, Brian Litt</i>	
Optimal Rates for Stochastic Convex Optimization Under Tsybakov Noise Condition	365
<i>Aaditya Ramdas, Aarti Singh</i>	
A Randomized Mirror Descent Algorithm for Large Scale Multiple Kernel Learning	374
<i>Arash Afkanpour, András György, Csaba Szepesvari, Michael Bowling</i>	
Noisy and Missing Data Regression: Distribution-Oblivious Support Recovery	383
<i>Yudong Chen, Constantine Caramanis</i>	
Dual Averaging and Proximal Gradient Descent for Online Alternating Direction Multiplier Method	392
<i>Taiji Suzuki</i>	
A New Frontier of Kernel Design for Structured Data	401
<i>Kilho Shin</i>	
Learning with Marginalized Corrupted Features	410
<i>Laurens Van der Maaten, Minmin Chen, Stephen Tyree, Kilian Weinberger</i>	
Approximation Properties of DBNs with Binary Hidden Units and Real-valued Visible Units	419
<i>Oswin Krause, Asja Fischer, Tobias Glasmachers, Christian Igel</i>	
Revisiting Frank-Wolfe: Projection-Free Sparse Convex Optimization	427
<i>Martin Jaggi</i>	
General Functional Matrix Factorization Using Gradient Boosting	436
<i>Tianqi Chen, Hang Li, Qiang Yang, Yong Yu</i>	
Iterative Learning and Denoising in Convolutional Neural Associative Memories	445
<i>Amin Karbasi, Amir Hesam Salavati, Amin Shokrollahi</i>	
Scaling Multidimensional Gaussian Processes using Projected Additive Approximations	454
<i>Elad Gilboa, Yunus Saatçi, John Cunningham</i>	
Active Learning for Multi-Objective Optimization	462
<i>Marcela Zuluaga, Guillaume Sergent, Andreas Krause, Markus Püschel</i>	
A Generalized Kernel Approach to Structured Output Learning	471
<i>Hachem Kadri, Mohammad Ghavamzadeh, Philippe Preux</i>	
Efficient Active Learning of Halfspaces: an Aggressive Approach	480
<i>Alon Gonen, Sivan Sabato, Shai Shalev-Shwartz</i>	
Enhanced Statistical Rankings Via Targeted Data Collection	489
<i>Braxton Osting, Christoph Brune, Stanley Osher</i>	
Online Feature Selection for Model-based Reinforcement Learning	498
<i>Trung Nguyen, Zhuoru Li, Tomi Silander, Tze Yun Leong</i>	
ELLA: An Efficient Lifelong Learning Algorithm	507
<i>Paul Ruvolo, Eric Eaton</i>	
A Structural SVM Based Approach for Optimizing Partial AUC	516
<i>Harikrishna Narasimhan, Shivani Agarwal</i>	
Convex Relaxations for Learning Bounded-Treewidth Decomposable Graphs	525
<i>K. S. Sesh Kumar, Francis Bach</i>	
Adaptive Task Assignment for Crowdsourced Classification	534
<i>Chien-Ju Ho, Shahin Jabbari, Jennifer Wortman Vaughan</i>	
Optimal Regret Bounds for Selecting the State Representation in Reinforcement Learning	543
<i>Odalric-Ambrym Maillard, Phuong Nguyen, Ronald Ortner, Daniil Ryabko</i>	
Better Mixing via Deep Representations	552
<i>Yoshua Bengio, Gregoire Mesnil, Yann Dauphin, Salah Rifai</i>	
Online Latent Dirichlet Allocation with Infinite Vocabulary	561
<i>Ke Zhai, Jordan Boyd-Graber</i>	
Characterizing the Representer Theorem	570
<i>Yaoliang Yu, Hao Cheng, Dale Schuurmans, Csaba Szepesvari</i>	
Dynamical Models and Tracking Regret in Online Convex Programming	579
<i>Eric Hall, Rebecca Willett</i>	

Large-Scale Bandit Problems and KWIK Learning	588
<i>Jacob Abernethy, Kareem Amin, Michael Kearns, Moez Draief</i>	
Vanishing Component Analysis	597
<i>Roi Livni, David Lehaví, Sagi Schein, Hila Nachliely, Shai Shalev-Shwartz, Amir Globerson</i>	
Learning an Internal Dynamics Model from Control Demonstration	606
<i>Matthew Golub, Steven Chase, Byron Yu</i>	
Robust Structural Metric Learning	615
<i>Daryl Lim, Gert Lanckriet, Brian McFee</i>	
Constrained Fractional Set Programs and Their Application in Local Clustering and Community Detection	624
<i>Thomas Bühler, Shyam Sundar Rangapuram, Simon Setzer, Matthias Hein</i>	
Efficient Semi-supervised and Active Learning of Disjunctions	633
<i>Nina Balcan, Christopher Berind, Steven Ehrlich, Yingyu Liang</i>	
Convex Adversarial Collective Classification	642
<i>MohamadAli Torkamani, Daniel Lowd</i>	
Rounding Methods for Discrete Linear Classification	651
<i>Yann Chevaleyre, Frédéric Kriche, Jean-Daniel Zucker</i>	
Mixture of Mutually Exciting Processes for Viral Diffusion	660
<i>Shuang-Hong Yang, Hongyuan Zha</i>	
Gaussian Process Vine Copulas for Multivariate Dependence	669
<i>David Lopez-Paz, Jose Miguel Hernández-Lobato, Ghahramani Zoubin</i>	
Stochastic Simultaneous Optimistic Optimization	678
<i>Michal Valko, Alexandra Carpentier, Rémi Munos</i>	
Toward Optimal Stratification for Stratified Monte-Carlo Integration	687
<i>Alexandra Carpentier, Rémi Munos</i>	
A General Iterative Shrinkage and Thresholding Algorithm for Non-convex Regularized Optimization Problems	696
<i>Pinghua Gong, Changshui Zhang, Zhaosong Lu, Jianhua Huang, Jieping Ye</i>	
Thurstonian Boltzmann Machines: Learning from Multiple Inequalities	705
<i>Truyen Tran, Dinh Phung, Svetha Venkatesh</i>	
A Variational Approximation for Topic Modeling of Hierarchical Corpora	714
<i>Do-kyum Kim, Geoffrey Voelker, Lawrence Saul</i>	
Forecastable Component Analysis	723
<i>Georg Goerg</i>	
Ellipsoidal Multiple Instance Learning	732
<i>Gabriel Krümmenacher, Cheng Soon Ong, Joachim Buhmann</i>	
Local Low-Rank Matrix Approximation	741
<i>Joonseok Lee, Seungyeon Kim, Guy Lebanon, Yoram Singer</i>	
Generic Exploration and K-armed Voting Bandits	750
<i>Tanguy Urvoy, Fabrice Clerot, Raphael Féraud, Sami Naamane</i>	
A Unifying Framework for Vector-valued Manifold Regularization and Multi-view Learning	759
<i>Minh Hà Quang, Loris Bazzani, Vittorio Murino</i>	
Learning Connections in Financial Time Series	768
<i>Gartheeban Ganeshapillai, John Guttag, Andrew Lo</i>	
Fast Dropout Training	777
<i>Sida Wang, Christopher Manning</i>	
Scalable Optimization of Neighbor Embedding for Visualization	786
<i>Zhirong Yang, Jaakko Peltonen, Samuel Kaski</i>	
Precision-recall Space to Correct External Indices for Biclustering	795
<i>Blaise Hanczar, Mohamed Nadif</i>	
Monochromatic Bi-Clustering	804
<i>Sharon Wulff, Ruth Urner, Shai Ben-David</i>	
Gated Autoencoders with Tied Input Weights	813
<i>Droniou Alain, Sigaud Olivier</i>	
Strict Monotonicity of Sum of Squares Error and Normalized Cut in the Lattice of Clusterings	822
<i>Nicola Rebagliati</i>	
Transition Matrix Estimation in High Dimensional Time Series	831
<i>Fang Han, Han Liu</i>	
Label Partitioning For Sublinear Ranking	840
<i>Jason Weston, Ameesh Makadia, Hector Yee</i>	

PART 2

Subproblem-Tree Calibration: A Unified Approach to Max-Product Message Passing	849
<i>Huayan Wang, Daphne Koller</i>	
Collaborative Hyperparameter Tuning	858
<i>Rémi Bardenet, Mátys Brendel, Balázs Kégl, Michèle Sebag</i>	
SADA: A General Framework to Support Robust Causation Discovery	867
<i>Ruichu Cai, Zhenjie Zhang, Zhifeng Hao</i>	
Learning and Selecting Features Jointly with Point-wise Gated Boltzmann Machines	876
<i>Kihyuk Sohn, Guanyu Zhou, Chansoo Lee, Honglak Lee</i>	

Sequential Bayesian Search	885
<i>Zheng Wen, Branislav Kveton, Brian Eriksson, Sandilya Bhamidipati</i>	
Sparse Projections Onto the Simplex	894
<i>Anastasios Kyrillidis, Stephen Becker, Volkan Cevher, Christoph Koch</i>	
Modeling Musical Influence with Topic Models	903
<i>Uri Shalit, Daphna Weinshall, Gal Chechik</i>	
Subtle Topic Models and Discovering Subtly Manifested Software Concerns Automatically	912
<i>Mrinal Das, Suparna Bhattacharya, Chiranjib Bhattacharyya, Gopinath Kanchi</i>	
Exploring the Mind: Integrating Questionnaires and fMRI	921
<i>Esther Salazar, Ryan Bogdan, Adam Gorka, Ahmad Hariri, Lawrence Carin</i>	
A Proximal Newton Framework for Composite Minimization: Graph Learning Without Cholesky Decompositions and Matrix Inversions	930
<i>Quoc Tran Dinh, Anastasios Kyrillidis, Volkan Cevher</i>	
A Practical Algorithm for Topic Modeling with Provable Guarantees	939
<i>Sanjeev Arora, Rong Ge, Yonatan Halpern, David Mimno, Ankur Moitra, David Sontag, Yichen Wu, Michael Zhu</i>	
Distributed Training of Large-scale Logistic Models	948
<i>Siddharth Gopal, Yiming Yang</i>	
An Adaptive Learning Rate for Stochastic Variational Inference	957
<i>Rajesh Ranganath, Chong Wang, Blei David, Eric Xing</i>	
Margins, Shrinkage, and Boosting	966
<i>Matus Telgarsky</i>	
Canonical Correlation Analysis based on Hilbert-Schmidt Independence Criterion and Centered Kernel Target Alignment	975
<i>Billy Chang, Uwe Kruger, Rafal Kustra, Junping Zhang</i>	
Large-Scale Learning with Less RAM via Randomization	984
<i>Daniel Golovin, D. Sculley, Brendan McMahan, Michael Young</i>	
Taming the Curse of Dimensionality: Discrete Integration by Hashing and Optimization	993
<i>Stefano Ermon, Carla Gomes, Ashish Sabharwal, Bart Selman</i>	
Sparse Coding for Multitask and Transfer Learning	1002
<i>Andreas Maurer, Massi Pontil, Bernardino Romera-Paredes</i>	
Direct Modeling of Complex Invariances for Visual Object Features	1011
<i>Ka Yu Hui</i>	
Hierarchically-coupled Hidden Markov Models for Learning Kinetic Rates from Single-molecule Data	1020
<i>Jan-Willem van de Meent, Jonathan Bronson, Frank Wood, Ruben Gonzalez Jr., Chris Wiggins</i>	
Activated Learning with Uniform Classification Noise	1029
<i>Liu Yang, Steve Hanneke</i>	
Guided Policy Search	1038
<i>Sergey Levine, Vladlen Koltun</i>	
Squared-loss Mutual Information Regularization: A Novel Information-theoretic Approach to Semi-supervised Learning	1047
<i>Gang Niu, Wittawat Jitkrittum, Bo Dai, Hirotaka Hachiya, Masashi Sugiyama</i>	
Gossip-based Distributed Stochastic Bandit Algorithms	1056
<i>Balazs Szorenyi, Robert Busa-Fekete, Istvan Hegedus, Robert Ormandi, Mark Jelasity, Balazs Kegl</i>	
The Sample-Complexity of General Reinforcement Learning	1065
<i>Tor Lattimore, Marcus Hutter, Peter Sunehag</i>	
Hierarchical Regularization Cascade for Joint Learning	1074
<i>Alon Zweig, Daphna Weinshall</i>	
Multi-Class Classification with Maximum Margin Multiple Kernel	1083
<i>Corinna Cortes, Mehryar Mohri, Afshin Rostamizadeh</i>	
Bayesian Games for Adversarial Regression Problems	1092
<i>Michael Großhans, Christoph Sawade, Michael Brückner, Tobias Scheffer</i>	
Optimistic Knowledge Gradient Policy for Optimal Budget Allocation in Crowdsourcing	1101
<i>Xi Chen, Qihang Lin, Dengyong Zhou</i>	
Markov Network Estimation From Multi-attribute Data	1110
<i>Mladen Kolar, Han Liu, Eric Xing</i>	
MILEAGE: Multiple Instance LEARNING with Global Embedding	1119
<i>Dan Zhang, Jingrui He, Luo Si, Richard Lawrence</i>	
Guaranteed Sparse Recovery under Linear Transformation	1128
<i>Ji Liu, Lei Yuan, Jieping Ye</i>	
Learning Invariant Features by Harnessing the Aperture Problem	1137
<i>Roland Memisevic, Georgios Exarchakis</i>	
Efficient Ranking from Pairwise Comparisons	1146
<i>Fabian Wauthier, Michael Jordan, Nebojsa Jojic</i>	
Differentially Private Learning with Kernels	1155
<i>Prateek Jain, Abhradeep Thakurta</i>	
Thompson Sampling for Contextual Bandits with Linear Payoffs	1164
<i>Shipra Agrawal, Navin Goyal</i>	
Learning Multiple Behaviors from Unlabeled Demonstrations in a Latent Controller Space	1173
<i>Javier Almingol, Lui Montesano, Manuel Lopes</i>	

Inference Algorithms for Pattern-based CRFs on Sequence Data	1182
<i>Vladimir Kolmogorov, Rustem Takhanov</i>	
One-Bit Compressed Sensing: Provable Support and Vector Recovery	1191
<i>Sivakant Gopi, Praneeth Netrapalli, Prateek Jain, Aditya Nori</i>	
Tensor Analyzers	1200
<i>Yichuan Tang, Ruslan Salakhutdinov, Geoffrey Hinton</i>	
Learning Sparse Penalties for Change-point Detection using Max Margin Interval Regression	1209
<i>Toby Hocking, Guillem Rigaill, Jean-Philippe Vert, Francis Bach</i>	
Learning from Human-Generated Lists	1218
<i>Kwang-Sung Jun, Jerry Zhu, Burr Settles, Timothy Rogers</i>	
A Fast and Exact Energy Minimization Algorithm for Cycle MRFs	1227
<i>Huayan Wang, Koller Daphne</i>	
Stochastic k-Neighborhood Selection for Supervised and Unsupervised Learning	1236
<i>Daniel Tarlow, Kevin Swersky, Laurent Charlin, Ilya Sutskever, Rich Zemel</i>	
An Efficient Posterior Regularized Latent Variable Model for Interactive Sound Source Separation	1245
<i>Nicholas Bryan, Gautham Mysore</i>	
Estimating Unknown Sparsity in Compressed Sensing	1254
<i>Miles Lopes</i>	
MAD-Bayes: MAP-based Asymptotic Derivations from Bayes	1263
<i>Tamara Broderick, Brian Kulis, Michael Jordan</i>	
The Most Generative Maximum Margin Bayesian Networks	1272
<i>Robert Peharz, Sebastian Tschiatschek, Franz Pernkopf</i>	
Fastfood - Computing Hilbert Space Expansions in Loglinear Time	1281
<i>Quoc Le, Tamas Sarlos, Alexander Smola</i>	
Joint Transfer and Batch-mode Active Learning	1290
<i>Rita Chattopadhyay, Wei Fan, Ian Davidson, Sethuraman Panchanathan, Jieping Ye</i>	
#####Message Passing with L1 Penalized KL Minimization	1299
<i>Yuan Qi, Yandong Guo</i>	
Mean Reversion with a Variance Threshold	1308
<i>Marco Cuturi, Alexandre D'Aspremont</i>	
Top-down Particle Filtering for Bayesian Decision Trees	1317
<i>Balaji Lakshminarayanan, Daniel Roy, Yee Whye Teh</i>	
Smooth Sparse Coding via Marginal Regression for Learning Sparse Representations	1326
<i>Krishnakumar Balasubramanian, Kai Yu, Guy Lebanon</i>	
Robust and Discriminative Self-Taught Learning	1335
<i>Hua Wang, Feiping Nie, Heng Huang</i>	
Safe Policy Iteration	1344
<i>Matteo Pirota, Marcello Restelli, Alessio Pecorino, Daniele Calandriello</i>	
Unfolding Latent Tree Structures using 4th Order Tensors	1353
<i>Mariya Ishteva, Haesun Park, Le Song</i>	
Learning Fair Representations	1362
<i>Rich Zemel, Yu Wu, Kevin Swersky, Toni Pitassi, Cynthia Dwork</i>	
Hierarchical Tensor Decomposition of Latent Tree Graphical Models	1371
<i>Le Song, Mariya Ishteva, Ankur Parikh, Eric Xing, Haesun Park</i>	
No More Pesky Learning Rates	1380
<i>Tom Schaul, Sixin Zhang, Yann LeCun</i>	
Multi-View Clustering and Feature Learning via Structured Sparsity	1389
<i>Hua Wang, Feiping Nie, Heng Huang</i>	
Planning by Prioritized Sweeping with Small Backups	1398
<i>Harm Van Seijen, Rich Sutton</i>	
Solving Continuous POMDPs: Value Iteration with Incremental Learning of an Efficient Space Representation	1407
<i>Sebastian Brechtel, Tobias Gindele, Rüdiger Dillmann</i>	
Learning Heteroscedastic Models by Convex Programming under Group Sparsity	1416
<i>Arnak Dalalyan, Mohamed Hebiri, Katia Meziani, Joseph Salmon</i>	
Covariate Shift in Hilbert Space: A Solution via Sorrogate Kernels	1425
<i>Kai Zhang, Vincent Zheng, Qiaojun Wang, James Kwok, Qiang Yang, Ivan Marsic</i>	
A Local Algorithm for Finding Well-Connected Clusters	1433
<i>Zeyuan Allen Zhu, Silvio Lattanzi, Vahab Mirrokni</i>	
Efficient Multi-label Classification with Many Labels	1442
<i>Wei Bi, James Kwok</i>	
Spectral Compressed Sensing via Structured Matrix Completion	1451
<i>Yuxin Chen, Yuejie Chi</i>	
Multi-Task Learning with Gaussian Matrix Generalized Inverse Gaussian Model	1460
<i>Ming Yang, Yingming Li, Zhongfei Zhang</i>	
Simple Sparsification Improves Sparse Denoising Autoencoders in Denoising Highly Corrupted Images	1469
<i>Kyunghyun Cho</i>	
On the Generalization Ability of Online Learning Algorithms for Pairwise Loss Functions	1478
<i>Purushottam Kar, Bharath Sriperumbudur, Prateek Jain, Harish Karnick</i>	
Non-Linear Stationary Subspace Analysis with Application to Video Classification	1487
<i>Mahsa Baktashmotlagh, Mehrtash Harandi, Abbas Bigdeli, Brian Lovell, Mathieu Salzmann</i>	

Two-Sided Exponential Concentration Bounds for Bayes Error Rate and Shannon Entropy	1496
<i>Jean Honorio, Tommi Jaakkola</i>	
That Was Fast! Speeding Up NN Search of High Dimensional Distributions	1505
<i>Emanuele Coviello, Adeel Mumtaz, Antoni Chan, Gert Lanckriet</i>	
Entropic Affinities: Properties and Efficient Numerical Computation	1514
<i>Max Vladymyrov, Miguel Carreira-Perpinan</i>	
Local Deep Kernel Learning for Efficient Non-linear SVM Prediction	1523
<i>Cijo Jose, Prasoon Goyal, Parv Aggrwal, Manik Varma</i>	
Temporal Difference Methods for the Variance of the Reward To Go	1532
<i>Aviv Tamar, Dotan Di Castro, Shie Mannor</i>	
αSVM for Learning with Label Proportions	1541
<i>Felix Yu, Dong Liu, Sanjiv Kumar, Jebara Tony, Shih-Fu Chang</i>	
Parameter Learning and Convergent Inference for Dense Random Fields	1550
<i>Philipp Kraehenbuehl, Vladlen Koltun</i>	
Loss-Proportional Subsampling for Subsequent ERM	1559
<i>Paul Mineiro, Nikos Karampatziakis</i>	
Scalable Simple Random Sampling and Stratified Sampling	1568
<i>Xiangrui Meng</i>	
Riemannian Similarity Learning	1577
<i>Li Cheng</i>	
On Compact Codes for Spatially Pooled Features	1586
<i>Yangqing Jia, Oriol Vinyals, Trevor Darrell</i>	
Dynamic Covariance Models for Multivariate Financial Time Series	1595
<i>Yue Wu, Jose Miguel Hernandez-Lobato, Ghahramani Zoubin</i>	
Revisiting the Nystrom Method for Improved Large-scale Machine Learning	1604
<i>Alex Gittens, Michael Mahoney</i>	
Infinite Positive Semidefinite Tensor Factorization for Source Separation of Mixture Signals	1613
<i>Kazuyoshi Yoshii, Ryota Tomioka, Daichi Mochihashi, Masataka Goto</i>	
A Unified Robust Regression Model for Lasso-like Algorithms	1622
<i>Wenzhuo Yang, Huan Xu</i>	
Quickly Boosting Decision Trees – Pruning Underachieving Features Early	1631
<i>Ron Appel, Thomas Fuchs, Piotr Dollar, Pietro Perona</i>	
On the Statistical Consistency of Algorithms for Binary Classification under Class Imbalance	1640
<i>Aditya Menon, Harikrishna Narasimhan, Shivani Agarwal, Sanjay Chawla</i>	
Topic Model Diagnostics: Assessing Domain Relevance via Topical Alignment	1649
<i>Jason Chuang, Sonal Gupta, Christopher Manning, Jeffrey Heer</i>	
Online Kernel Learning with a Near Optimal Sparsity Bound	1658
<i>Lijun Zhang, Jinfeng Yi, Rong Jin, Ming Lin, Xiaofei He</i>	
Spectral Learning of Hidden Markov Models from Dynamic and Static Data	1667
<i>Tzu-Kuo Huang, Jeff Schneider</i>	
Analogy-preserving Semantic Embedding for Visual Object Categorization	1676
<i>Sung Ju Hwang, Kristen Grauman, Fei Sha</i>	
Algebraic Classifiers: A Generic Approach to Fast Cross-validation, Online Training, and Parallel Training	1685
<i>Michael Izbicki</i>	
Factorial Multi-Task Learning : A Bayesian Nonparametric Approach	1694
<i>Sunil Gupta, Dinh Phung, Svetha Venkatesh</i>	

PART 3

Modeling Information Propagation with Survival Theory	1703
<i>Manuel Gomez-Rodriguez, Jure Leskovec, Bernhard Scholkopf</i>	
Better Rates for Any Adversarial Deterministic MDP	1712
<i>Ofar Dekel, Elad Hazan</i>	
ABC Reinforcement Learning	1721
<i>Christos Dimitrakakis, Nikolaos Tziortziotis</i>	
Sharp Generalization Error Bounds for Randomly-projected Classifiers	1730
<i>Robert Durrant, Ata Kaban</i>	
On Learning Parametric-output HMMs	1739
<i>Aryeh Kontorovich, Boaz Nadler, Roi Weiss</i>	
LDA Topic Model with Soft Assignment of Descriptors to Words	1748
<i>Daphna Weinshall, Gal Levi, Dmitri Hanukaev</i>	
On Autoencoder Scoring	1757
<i>Hanna Kamyshanska, Roland Memisevic</i>	
Infinite Markov-Switching Maximum Entropy Discrimination Machines	1766
<i>Sotirios Chatzis</i>	
A PAC-Bayesian Approach for Domain Adaptation with Specialization to Linear Classifiers	1775
<i>Pascal Germain, Amaury Habrard, Francois Laviolette, Emilie Morvant</i>	
Sparse PCA through Low-rank Approximations	1784
<i>Dimitris Papailiopoulos, Alexandros Dimakis, Stavros Korokythakis</i>	

Computation-Risk Tradeoffs for Covariance-Thresholded Regression	1793
<i>Dinah Shender, John Lafferty</i>	
Exact Rule Learning via Boolean Compressed Sensing	1802
<i>Dmitry Malioutov, Kush Varshney</i>	
Robust Sparse Regression under Adversarial Corruption	1811
<i>Yudong Chen, Constantine Caramanis, Shie Mannor</i>	
Optimization with First-Order Surrogate Functions	1820
<i>Julien Mairal</i>	
Learning Spatio-Temporal Structure from RGB-D Videos for Human Activity Detection and Anticipation	1829
<i>Hema Koppula, Ashutosh Saxena</i>	
Consistency versus Realizable H-Consistency for Multiclass Classification	1838
<i>Phil Long, Rocco Servedio</i>	
Feature Multi-Selection among Subjective Features	1847
<i>Sivan Sabato, Adam Kalai</i>	
Domain Adaptation under Target and Conditional Shift	1856
<i>Kun Zhang, Bernhard Scholkopf, Krikamol Muandet, Zhikun Wang</i>	
Collective Stability in Structured Prediction: Generalization from One Example	1865
<i>Ben London, Bert Huang, Ben Taskar, Lise Getoor</i>	
Stable Coactive Learning via Perturbation	1874
<i>Karthik Raman, Thorsten Joachims, Pannaga Shivaswamy, Tobias Schnabel</i>	
Max-Margin Multiple-Instance Dictionary Learning	1883
<i>Xinggang Wang, Baoyuan Wang, Xiang Bai, Wenyu Liu, Zhuowen Tu</i>	
Fast Semidifferential-based Submodular Function Optimization	1892
<i>Rishabh Iyer, Stefanie Jegelka, Jeff Bilmes</i>	
Kernelized Bayesian Matrix Factorization	1901
<i>Mehmet Gönen, Suleiman Khan, Samuel Kaski</i>	
Learning the Structure of Sum-Product Networks	1910
<i>Robert Gens, Domingos Pedro</i>	
Quantile Regression for Large-scale Applications	1918
<i>Jiyan Yang, Xiangrui Meng, Michael Mahoney</i>	
Robust Regression on MapReduce	1925
<i>Xiangrui Meng, Michael Mahoney</i>	
Infinitesimal Annealing for Training Semi-Supervised Support Vector Machines	1934
<i>Kohei Ogawa, Motoki Imamura, Ichiro Takeuchi, Masashi Sugiyama</i>	
One-Pass AUC Optimization	1943
<i>Wei Gao, Rong Jin, Shenghuo Zhu, Zhi-Hua Zhou</i>	
Learning Convex QP Relaxations for Structured Prediction	1952
<i>Jeremy Jancsary, Sebastian Nowozin, Carsten Rother</i>	
Concurrent Reinforcement Learning from Customer Interactions	1961
<i>David Silver, Leonard Newnham, David Barker, Suzanne Weller, Jason McFall</i>	
Saving Evaluation Time for the Decision Function in Boosting: Representation and Reordering Base Learner	1970
<i>Peng Sun, Jie Zhou</i>	
Stability and Hypothesis Transfer Learning	1979
<i>Ilya Kuzborskij, Francesco Orabona</i>	
Fast Dual Variational Inference for Non-Conjugate Latent Gaussian Models	1988
<i>Mohammad Emtiyaz Khan, Aleksandr Aravkin, Michael Friedlander, Matthias Seeger</i>	
Modeling Temporal Evolution and Multiscale Structure in Networks	1997
<i>Tue Herlau, Morten Mrup, Mikkel Schmidt</i>	
Dependent Normalized Random Measures	2006
<i>Changyou Chen, Vinayak Rao, Wray Buntine, Yee Whye Teh</i>	
Fast Max-Margin Matrix Factorization with Data Augmentation	2015
<i>Minjie Xu, Jun Zhu, Bo Zhang</i>	
Natural Image Bases to Represent Neuroimaging Data	2024
<i>Ashish Gupta, Murat Ayhan, Anthony Maida</i>	
Breaking the Small Cluster Barrier of Graph Clustering	2032
<i>Nir Ailon, Yudong Chen, Huan Xu</i>	
Approximate Inference in Collective Graphical Models	2041
<i>Daniel Sheldon, Tao Sun, Akshat Kumar, Tom Dietterich</i>	
Scaling the Indian Buffet Process via Submodular Maximization	2050
<i>Colorado Reed, Ghahramani Zoubin</i>	
Mini-Batch Primal and Dual Methods for SVMs	2059
<i>Martin Takac, Avleen Bijral, Peter Richtarik, Nati Srebro</i>	
The Lasso, Persistence, and Cross-validation	2068
<i>Darren Homrighausen, Daniel McDonald</i>	
Spectral Experts for Estimating Mixtures of Linear Regressions	2077
<i>Arun Tejasvi Chaganty, Percy Liang</i>	
Distribution to Distribution Regression	2086
<i>Junier Oliva, Barnabas Poczos, Jeff Schneider</i>	
Regularization of Neural Networks using DropConnect	2095
<i>Li Wan, Matthew Zeiler, Sixin Zhang, Yann Le Cun, Rob Fergus</i>	

Gaussian Process Kernels for Pattern Discovery and Extrapolation	2104
<i>Andrew Wilson, Ryan Adams</i>	
Anytime Representation Learning	2113
<i>Zhixiang Xu, Matt Kusner, Gao Huang, Kilian Weinberger</i>	
Algorithms for Direct 0–1 Loss Optimization in Binary Classification	2122
<i>Tan Nguyen, Scott Sanner</i>	
Top-k Selection based on Adaptive Sampling of Noisy Preferences	2131
<i>Robert Busa-Fekete, Balazs Szorenyi, Weiwei Cheng, Paul Weng, Eyke Huellermeier</i>	
The Extended Parameter Filter	2140
<i>Yusuf Bugra Erol, Lei Li, Bharath Ramsundar, Russell Stuart</i>	
Exploiting Ontology Structures and Unlabeled Data for Learning	2149
<i>Nina Balcan, Avrim Blum, Yishay Mansour</i>	
O(logT) Projections for Stochastic Optimization of Smooth and Strongly Convex Functions	2158
<i>Lijun Zhang, Tianbao Yang, Rong Jin, Xiaofei He</i>	
Optimizing the F-Measure in Multi-Label Classification: Plug-in Rule Approach versus Structured Loss Minimization	2167
<i>Krzysztof Dembczynski, Arkadiusz Jachnik, Wojciech Kotlowski, Willem Waegeman, Eyke Huellermeier</i>	
On the Importance of Initialization and Momentum in Deep Learning	2176
<i>Ilya Sutskever, James Martens, George Dahl, Geoffrey Hinton</i>	
A Non-IID Framework for Collaborative Filtering with Restricted Boltzmann Machines	2185
<i>Kostadin Georgiev, Preslav Nakov</i>	
Intersecting Singularities for Multi-structured Estimation	2194
<i>Emile Richard, Francis Bach, Jean-Philippe Vert</i>	
Structure Discovery in Nonparametric Regression through Compositional Kernel Search	2203
<i>David Duvenaud, James Lloyd, Roger Grosse, Joshua Tenenbaum, Zoubin Ghahramani</i>	
Copy or Coincidence? A Model for Detecting Social Influence and Duplication Events	2212
<i>Lisa Friedland, David Jensen, Michael Lavine</i>	
Smooth Operators	2221
<i>Steffen Grunewalder, Gretton Arthur, John Shawe-Taylor</i>	
The Cross-Entropy Method Optimizes for Quantiles	2230
<i>Sergiu Goschin, Ari Weinstein, Michael Littman</i>	
Topic Discovery through Data Dependent and Random Projections	2239
<i>Weicong Ding, Mohammad Hossein Rohban, Prakash Ishwar, Venkatesh Saligrama</i>	
Bayesian Learning of Recursively Factored Environments	2248
<i>Marc Bellemare, Joel Veness, Michael Bowling</i>	
Selective Sampling Algorithms for Cost-sensitive Multiclass Prediction	2257
<i>Alekh Agarwal</i>	
The Bigraphical Lasso	2266
<i>Alfredo Kalaitzis, John Lafferty, Neil Lawrence, Shuheng Zhou</i>	
Almost Optimal Exploration in Multi-Armed Bandits	2275
<i>Zohar Karnin, Tomer Koren, Oren Somekh</i>	
Deep Canonical Correlation Analysis	2284
<i>Galen Andrew, Raman Arora, Jeff Bilmes, Karen Livescu</i>	
Consistency of Online Random Forests	2293
<i>Misha Denil, David Matheson, Nando De Freitas</i>	
Sparse Gaussian Conditional Random Fields: Algorithms, Theory, and Application to Energy Forecasting	2302
<i>Matt Wytoczek, J. Zico Kolter</i>	
Fast Image Tagging	2311
<i>Minmin Chen, Alice Zheng, Kilian Weinberger</i>	
Expensive Function Optimization with Stochastic Binary Outcomes	2320
<i>Matthew Tesch, Jeff Schneider, Howie Choset</i>	
Multiple-source Cross-validation	2329
<i>Krzysztof Geras, Charles Sutton</i>	
Learning Triggering Kernels for Multi-dimensional Hawkes Processes	2338
<i>Ke Zhou, Hongyuan Zha, Le Song</i>	
On the Difficulty of Training Recurrent Neural Networks	2347
<i>Razvan Pascanu, Tomas Mikolov, Yoshua Bengio</i>	
Maxout Networks	2356
<i>Ian Goodfellow, David Warde-Farley, Mehdi Mirza, Aaron Courville, Yoshua Bengio</i>	
Predictable Dual-View Hashing	2365
<i>Mohammad Rastegari, Jonghyun Choi, Shobeir Fakhraei, Daume Hal, Larry Davis</i>	
Deep Learning with COTS HPC Systems	2374
<i>Adam Coates, Brody Huval, Tao Wang, David Wu, Bryan Catanzaro, Ng Andrew</i>	
Nonparametric Mixture of Gaussian Processes with Constraints	2383
<i>James Ross, Jennifer Dy</i>	
Scale Invariant Conditional Dependence Measures	2392
<i>Sashank J Reddi, Barnabas Poczos</i>	
Learning Policies for Contextual Submodular Prediction	2401
<i>Stephane Ross, Jiaji Zhou, Yisong Yue, Debadeepta Dey, Drew Bagnell</i>	

Manifold Preserving Hierarchical Topic Models for Quantization and Approximation	2410
<i>Minje Kim, Paris Smaragdis</i>	
Safe Screening of Non-Support Vectors in Pathwise SVM Computation	2419
<i>Kohei Ogawa, Yoshiki Suzuki, Ichiro Takeuchi</i>	
Cost-sensitive Multiclass Classification Risk Bounds	2428
<i>Bernardo Avila Pires, Csaba Szepesvari, Mohammad Ghavamzadeh</i>	
Semi-supervised Clustering by Input Pattern Assisted Pairwise Similarity Matrix Completion	2450
<i>Jinfeng Yi, Lijun Zhang, Rong Jin, Qi Qian, Anil Jain</i>	
Learning the Beta-Divergence in Tweedie Compound Poisson Matrix Factorization Models	2459
<i>Umut Simsekli, Ali Taylan Cemgil, Yusuf Kenan Yilmaz</i>	
Fast Algorithms for Sparse Principal Component Analysis Based on Rayleigh Quotient Iteration	2468
<i>Volodymyr Kuleshov</i>	
Nested Chinese Restaurant Franchise Process: Applications to User Tracking and Document Modeling	2476
<i>Amr Ahmed, Liangjie Hong, Alexander Smola</i>	
Tree-Independent Dual-Tree Algorithms	2485
<i>Ryan Curtin, William March, Parikshit Ram, David Anderson, Alexander Gray, Charles Isbell Jr.</i>	
Multilinear Multitask Learning	2494
<i>Bernardino Romera-Paredes, Hane Aung, Nadia Bianchi-Berthouze, Massimiliano Pontil</i>	
Online Learning under Delayed Feedback	2503
<i>Pooria Joulani, Andras Gyorgy, Csaba Szepesvari</i>	
Adaptive Hamiltonian and Riemann Manifold Monte Carlo	2512
<i>Ziyu Wang, Shakir Mohamed, Nando De Freitas</i>	
Coco-Q: Learning in Stochastic Games with Side Payments	2521
<i>Eric Sodomka, Elizabeth Hilliard, Michael Littman, Amy Greenwald</i>	
On A Nonlinear Generalization of Sparse Coding and Dictionary Learning	2530
<i>Jeffrey Ho, Yuchen Xie, Baba Vemuri</i>	
Estimation of Causal Peer Influence Effects	2539
<i>Panos Toulis, Edward Kao</i>	
Author Index	