

Table of Content

Preface	xv
Organization	xviii
Program Committees	xx
Sponsors	xxv
Invited Talks	xxvii
Best Paper Awards	xxxiv
Tutorials	xl
Workshops	xiiv
Invited Cross-Conference Talks	li
Hashing with Graphs	1
Wei Liu; Jun Wang; Sanjiv Kumar; Shih-Fu Chang	
Efficient Sparse Modeling with Automatic Feature Grouping	9
Wenliang Zhong; James Kwok	
Multi-Label Classification on Tree- and DAG-Structured Hierarchies	17
Wei Bi; James Kwok	
A Graph-based Framework for Multi-Task Multi-View Learning	25
Jingrui He; Rick Lawrence	
GoDec: Randomized Low-rank & Sparse Matrix Decomposition in Noisy Case	33
Tianyi Zhou; Dacheng Tao	
Unimodal Bandits	41
Jia Yuan Yu; Shie Mannor	
Learning Output Kernels with Block Coordinate Descent	49
Francesco Dinuzzo; Cheng Soon Ong; Peter Gehler; Gianluigi Pillonetto	
Vector-valued Manifold Regularization	57
Ha Quang Minh; Vikas Sindhwani	
On Information-Maximization Clustering: Tuning Parameter Selection and Analytic Solution	65
Masashi Sugiyama; Makoto Yamada; Manabu Kimura; Hirotaka Hachiya	
On tracking portfolios with certainty equivalents on a generalization of Markowitz model: the Fool, the Wise and the Adaptive	73
Richard Nock; Brice Magdalou; Eric Briys; Frank Nielsen	
Multiple Instance Learning with Manifold Bags	81
Boris Babenko; Nakul Verma; Piotr Dollar; Serge Belongie	
Eigenvalue Sensitive Feature Selection	89
Yi Jiang; Jiangtao Ren	
Large Scale Text Classification using Semi-supervised Multinomial Naive Bayes	97
Jiang Su; Jelber Sayyad Shirab; Stan Matwin	

Enhanced Gradient and Adaptive Learning Rate for Training Restricted Boltzmann Machines	105
KyungHyun Cho; Tapani Raiko; Alexander Ilin	
Dynamic Tree Block Coordinate Ascent	113
Daniel Tarlow; Dhruv Batra; Pushmeet Kohli; Vladimir Kolmogorov	
Implementing regularization implicitly via approximate eigenvector computation	121
Michael Mahoney; Lorenzo Orecchia	
Parsing Natural Scenes and Natural Language with Recursive Neural Networks	129
Richard Socher; Cliff Chiung-Yu Lin; Andrew Ng; Chris Manning	
Conjugate Markov Decision Processes	137
Philip Thomas; Andrew Barto	
Learning Mallows Models with Pairwise Preferences	145
Tyler Lu; Craig Boutilier	
Surrogate losses and regret bounds for cost-sensitive classification with example-dependent costs	153
Clayton Scott	
Efficient Rule Ensemble Learning using Hierarchical Kernels	161
Pratik Jawanpuria; Saketha Nath Jagarlapudi; Ganesh Ramakrishnan	
An Augmented Lagrangian Approach to Constrained MAP Inference	169
Andre Martins; Mario Figueiredo; Pedro Aguiar; Noah Smith; Eric Xing	
Mean-Variance Optimization in Markov Decision Processes	177
Shie Mannor; John Tsitsiklis	
Time Series Clustering: Complex is Simpler!	185
Lei Li; B. Aditya Prakash	
Max-margin Learning for Lower Linear Envelope Potentials in Binary Markov Random Fields	193
Stephen Gould	
Inference of Inversion Transduction Grammars	201
Alexander Clark	
BCDNPCL: Scalable Non-Parametric Kernel Learning Using Block Coordinate Descent	209
En-Liang Hu; Bo Wang; SongCan Chen	
Learning Discriminative Fisher Kernels	217
Laurens Van der Maaten	
Pruning nearest neighbor cluster trees	225
Samory Kpotufe; Ulrike von Luxburg	

Online AUC Maximization	233
Peilin Zhao; Steven Hoi; Rong Jin; Tianbao Yang	
Beat the Mean Bandit	241
Yisong Yue; Thorsten Joachims	
Ultra-Fast Optimization Algorithm for Sparse Multi Kernel Learning.....	249
Francesco Orabona; Luo Jie	
Estimating the Bayes Point Using Linear Knapsack Problems	257
Brian Potetz	
On optimization methods for deep learning	265
Quoc Le; Jiquan Ngiam; Adam Coates; Ahbik Lahiri; Bobby Prochnow; Andrew Ng	
Multiclass Classification with Bandit Feedback using Adaptive Regularization	273
Koby Crammer; Claudio Gentile	
On the Necessity of Irrelevant Variables	281
Dave Helmbold; Phil Long	
ABC-EP: Expectation Propagation for Likelihood-free Bayesian Computation	289
Simon Barthelmé; Nicolas Chopin	
A PAC-Bayes Sample-compression Approach to Kernel Methods	297
Pascal Germain; Alexandre Lacoste; Francois Laviolette; Mario Marchand; Sara Shanian	
Integrating Partial Model Knowledge in Model Free RL Algorithms.....	305
Aviv Tamar; Dotan Di Castro; Ron Meir	
Fast Newton-type Methods for Total Variation Regularization	313
Álvaro Barbero; Suvrit Sra	
Parallel Coordinate Descent for L1-Regularized Loss Minimization.....	321
Joseph Bradley; Aapo Kyrola; Daniel Bickson; Carlos Guestrin	
Large-Scale Convex Minimization with a Low-Rank Constraint.....	329
Shai Shalev-Shwartz; Alon Gonen; Ohad Shamir	
Approximate Dynamic Programming for Storage Problems	337
Lauren Hannah; David Dunson	
Online Submodular Minimization for Combinatorial Structures	345
Stefanie Jegelka; Jeff Bilmes	
Minimal Loss Hashing for Compact Binary Codes.....	353
Mohammad Norouzi; David Fleet	
The Hierarchical Beta Process for Convolutional Factor Analysis and Deep Learning.....	361
Bo Chen; Gungor Polatkan; Guillermo Sapiro; David Dunson; Lawrence Carin	
Simultaneous Learning and Covering with Adversarial Noise	369
Andrew Guillory; Jeff Bilmes	

Topic Modeling with Nonparametric Markov Tree.....	377
Haojun Chen; David Dunson; Lawrence Carin	
Relational Active Learning for Joint Collective Classification Models.....	385
Ankit Kuwadekar; Jennifer Neville	
A Co-training Approach for Multi-view Spectral Clustering.....	393
Abhishek Kumar; Hal Daume III	
Learning from Multiple Outlooks.....	401
Maayan Harel; Shie Mannor	
Adaptive Kernel Approximation for Large-Scale Non-Linear SVM Prediction.....	409
Michele Cossalter; Rong Yan; Lu Zheng	
Risk-Based Generalizations of f-divergences.....	417
Darío García-García; Ulrike von Luxburg; Raúl Santos-Rodríguez	
Learning Multi-View Neighborhood Preserving Projections.....	425
Novi Quadrianto; Christoph Lampert	
Better Algorithms for Selective Sampling.....	433
Francesco Orabona; Nicolò Cesa-Bianchi	
Minimax Learning Rates for Bipartite Ranking and Plug-in Rules.....	441
Sylvain Robbiano; Stéphan Cléménçon	
Task Space Retrieval Using Inverse Feedback Control.....	449
Nikolay Jetchev; Marc Toussaint	
Bayesian CCA via Group Sparsity.....	457
Seppo Virtanen; Arto Klami; Samuel Kaski	
PILCO: A Model-Based and Data-Efficient Approach to Policy Search.....	465
Marc Deisenroth; Carl Rasmussen	
Suboptimal Solution Path Algorithm for Support Vector Machine.....	473
Masayuki Karasuyama; Ichiro Takeuchi	
Incremental Basis Construction from Temporal Difference Error.....	481
Yi Sun; Faustino Gomez; Mark Ring; Jürgen Schmidhuber	
Predicting Legislative Roll Calls from Text.....	489
Sean Gerrish; David Blei	
On Bayesian PCA: Automatic Dimensionality Selection and Analytic Solution.....	497
Shinichi Nakajima; Masashi Sugiyama; Derin Babacan	
Learning Linear Functions with Quadratic and Linear Multiplicative Updates.....	505
Tom Bylander	
Domain Adaptation for Large-Scale Sentiment Classification: A Deep Learning Approach.....	513
Xavier Glorot; Antoine Bordes; Yoshua Bengio	

Learning with Whom to Share in Multi-task Feature Learning	521
Zhuoliang Kang; Kristen Grauman; Fei Sha	
Boosting on a Budget: Sampling for Feature-Efficient Prediction	529
Lev Reyzin	
Speeding-Up Hoeffding-Based Regression Trees With Options	537
Elena Ikonomovska; João Gama; Bernard Zenko; Saso Dzeroski	
Linear Regression under Fixed-Rank Constraints: A Riemannian Approach	545
Gilles Meyer; Silvère Bonnabel; Rodolphe Sepulchre	
Cauchy Graph Embedding	553
Dijun Luo; Chris Ding; Feiping Nie; Heng Huang	
Uncovering the Temporal Dynamics of Diffusion Networks	561
Manuel Gomez Rodriguez; David Balduzzi; Bernhard Schölkopf	
Multiclass Boosting with Hinge Loss based on Output Coding	569
Tianshi Gao; Daphne Koller	
Approximation Bounds for Inference using Cooperative Cuts	577
Stefanie Jegelka; Jeff Bilmes	
Brier Curves: a New Cost-Based Visualisation of Classifier Performance	585
Jose Hernandez-Orallo; Peter Flach; Cèsar Ferri	
Semi-supervised Penalized Output Kernel Regression for Link Prediction	593
Céline Brouard; Florence D'Alche-Buc; Marie Szafranski	
A New Bayesian Rating System for Team Competitions	601
Sergey Nikolenko; Alexander Sirotkin	
OptiML: An Implicitly Parallel Domain-Specific Language for Machine Learning	609
Arvind Sujeeth; HyoukJoong Lee; Kevin Brown; Tiark Rompf; Hassan Chafi; Michael Wu; Anand Atreya; Martin Odersky; Kunle Olukotun	
Infinite SVM: a Dirichlet Process Mixture of Large-margin Kernel Machines	617
Jun Zhu; Ning Chen; Eric Xing	
On the Integration of Topic Modeling and Dictionary Learning	625
Lingbo Li; Mingyuan Zhou; Guillermo Sapiro; Lawrence Carin	
Piecewise Bounds for Estimating Bernoulli-Logistic Latent Gaussian Models	633
Benjamin Marlin*, University of British Columbia; Mohammad Khan, University of British Columbia; Kevin Murphy, University of British Columbia	
Access to Unlabeled Data can Speed up Prediction Time	641
Ruth Urner; Shai Shalev-Shwartz; Shai Ben-David	
From PAC-Bayes Bounds to Quadratic Programs for Majority Votes	649
Jean-Francois Roy; Francois Laviolette; Mario Marchand	

A Coherent Interpretation of AUC as a Measure of Aggregated Classification Performance	657
Peter Flach; Jose Hernandez-Orallo; Cèsar Ferri	
Support Vector Machines as Probabilistic Models	665
Vojtech Franc; Alexander Zien; Bernhard Schölkopf	
Adaptively Learning the Crowd Kernel	673
Omer Tamuz; Ce Liu; Serge Belongie; Ohad Shamir; Adam Kalai	
Bayesian Learning via Stochastic Gradient Langevin Dynamics	681
Max Welling; Yee Whye Teh	
Multimodal Deep Learning	689
Jiquan Ngiam; Aditya Khosla; Mingyu Kim; Juhan Nam; Honglak Lee; Andrew Ng	
On the Robustness of Kernel Density M-Estimators	697
JooSeuk Kim; Clayton Scott	
Beam Search based MAP Estimates for the Indian Buffet Process	705
Piyush Rai; Hal Daume III	
Optimal Distributed Online Prediction	713
Ofer Dekel; Ran Gilad-Bachrach; Ohad Shamir; Lin Xiao	
Message Passing Algorithms for the Dirichlet Diffusion Tree	721
David Knowles; Jurgen Van Gael; Zoubin Ghahramani	
Convex Max-Product over Compact Sets for Protein Folding	729
Jian Peng; Tamir Hazan; David McAllester; Raquel Urtasun	
Structure Learning in Ergodic Factored MDPs without Knowledge of the Transition Function's In-Degree	737
Doran Chakraborty; Peter Stone	
Clusterpath: an Algorithm for Clustering using Convex Fusion Penalties	745
Toby Hocking; Jean-Philippe Vert; Francis Bach; Armand Joulin	
Tree preserving embedding	753
Albert Shieh; Tatsunori Hashimoto; Edo Airolidi	
Clustering by Left-Stochastic Matrix Factorization	761
Raman Arora; Maya Gupta; Amol Kapila; Maryam Fazel,	
The Infinite Regionalized Policy Representation	769
Miao Liu; Xuejun Liao; Lawrence Carin	
SampleRank: Training Factor Graphs with Atomic Gradients	777
Michael Wick; Khashayar Rohanimanesh; Kedar Bellare; Aron Culotta; Andrew McCallum	
Tree-Structured Infinite Sparse Factor Model	785
XianXing Zhang; David Dunson; Lawrence Carin	

Preserving Personalized Pagerank in Subgraphs	793
Andrea Vattani; Deepayan Chakrabarti; Maxim Gurevich	
Hierarchical Classification via Orthogonal Transfer	801
Lin Xiao; Dengyong Zhou; Mingrui Wu	
A Three-Way Model for Collective Learning on Multi-Relational Data	809
Maximilian Nickel; Volker Tresp; Hans-Peter Kriegel	
Variational Inference for Policy Search in changing situations	817
Gerhard Neumann	
Learning Scoring Functions with Order-Preserving Losses and Standardized Supervision	825
David Buffoni; Clément Calauzenes; Patrick Gallinari; Nicolas Usunier	
Contractive Auto-Encoders: Explicit Invariance During Feature Extraction	833
Salah Rifai; Pascal Vincent; Xavier Muller; Xavier Glorot; Yoshua Bengio	
Variational Heteroscedastic Gaussian Process Regression	841
Miguel Lazaro-Gredilla; Michalis Titsias	
Bounding the Partition Function using Holder's Inequality	849
Qiang Liu; Alexander Ihler	
Dynamic Egocentric Models for Citation Networks	857
Duy Vu; Arthur Asuncion; David Hunter; Padhraic Smyth	
The Constrained Weight Space SVM: Learning with Ranked Features	865
Kevin Small; Byron Wallace; Carla Brodley; Thomas Trikalinos	
Robust Matrix Completion and Corrupted Columns	873
Yudong Chen; Huan Xu; Constantine Caramanis; Sujay Sanghavi	
Online Discovery of Feature Dependencies	881
Alborz Geramifard; Finale Doshi; Joshua Redding; Nicholas Roy; Jonathan How	
Variational Inference for Stick-Breaking Beta Process Priors	889
John Paisley; Lawrence Carin; David Blei	
Apprenticeship Learning About Multiple Intentions	897
Monica Babes; Vukosi Marivate; Michael Littman; Kaushik Subramanian	
Minimum Probability Flow Learning	905
Jascha Sohl-Dickstein; Peter Battaglino; Michael DeWeese	
Infinite Dynamic Bayesian Networks	913
Finale Doshi; David Wingate; Josh Tenenbaum; Nicholas Roy	
The Importance of Encoding Versus Training with Sparse Coding and Vector Quantization	921
Adam Coates; Andrew Ng	

Fast Global Alignment Kernels	929
Marco Cuturi	
Learning attentional policies for tracking and recognition in video with deep networks	937
Loris Bazzani; Nando Freitas; Hugo Larochelle; Vittorio Murino; Jo-Anne Ting	
Large-Scale Learning of Embeddings with Reconstruction Sampling	945
Yann Dauphin; Xavier Glorot; Yoshua Bengio	
Automatic Feature Decomposition for Single View Co-training	953
Minmin Chen; Kilian Weinberger; Yixin Chen	
Mapping kernels for trees	961
Kilho Shin; Marco Cuturi; Tetsuji Kuboyama	
Stochastic Low-Rank Kernel Learning for Regression	969
Pierre Machart; Thomas Peel; Sandrine Anthoine; Liva Ralaivola; Hervé Glotin,	
Size-constrained Submodular Minimization through Minimum Norm Base	977
Kiyohito Nagano; Yoshinobu Kawahara; Kazuyuki Aihara	
Locally Linear Support Vector Machines	985
Lubor Ladicky; Philip Torr	
Functional Regularized Least Squares Classification with Operator-valued Kernels	993
Hachem Kadri; Asma Rabaoui; Philippe Preux; Emmanuel Duflos; Alain Rakotomamonjy	
Clustering Partially Observed Graphs via Convex Optimization	1001
Ali Jalali; Yudong Chen; Sujay Sanghavi; Huan Xu	
On the Use of Variational Inference for Learning Discrete Graphical Model	1009
Eunho Yang; Pradeep Ravikumar	
Generating Text with Recurrent Neural Networks	1017
Ilya Sutskever; James Martens; Geoffrey Hinton	
Probabilistic Matrix Addition	1025
Amrudin Agovic; Arindam Banerjee; Snigdhanu Chatterje	
Learning Recurrent Neural Networks with Hessian-Free Optimization	1033
James Martens; Ilya Sutskever	
Sparse Additive Generative Models of Text	1041
Jacob Eisenstein; Amr Ahmed; Eric Xing	
Classification-based Policy Iteration with a Critic	1049
Victor Gabillon; Alessandro Lazaric; Mohammad Ghavamzadeh; Bruno Scherrer	
Submodular meets Spectral: Greedy Algorithms for Subset Selection, Sparse Approximation and Dictionary Selection	1057
Abhimanyu Das; David Kempe	

A Spectral Algorithm for Latent Tree Graphical Models.....	1065
Ankur Parikh; Le Song; Eric Xing	
A Unified Probabilistic Model for Global and Local Unsupervised Feature Selection.....	1073
Yue Guan; Jennifer Dy; Michael Jordan	
Towards Making Unlabeled Data Never Hurt	1081
Yu-Feng Li; Zhi-Hua Zhou	
On Random Weights and Unsupervised Feature Learning.....	1089
Andrew Saxe; Pang Wei Koh; Zhenghao Chen; Maneesh Bhand; Bipin Suresh; Andrew Ng	
Doubly Robust Policy Evaluation and Learning	1097
Miroslav Dudik; John Langford; Lihong Li	
Learning Deep Energy Models.....	1105
Jiquan Ngiam; Zhenghao Chen; Pang Wei Koh; Andrew Ng	
Bipartite Ranking through Minimization of Univariate Loss.....	1113
Wojciech Kotlowski; Krzysztof Dembczynski; Eyke Huellermeier	
Manifold Identification of Dual Averaging Methods for Regularized Stochastic Online Learning.....	1121
Sangkyun Lee; Stephen Wright	
Noisy matrix decomposition via convex relaxation: Optimal rates in high dimensions.....	1129
Alekh Agarwal; Sahand Negahban; Martin Wainwright	
Bundle Selling by Online Estimation of Valuation Functions.....	1137
Daniel Vainsencher; Ofer Dekel; Shie Mannor	
Unsupervised Models of Images by Spike-and-Slab RBMs.....	1145
Aarron Courville; James Bergstra; Yoshua Bengio	
Approximating Correlated Equilibria using Relaxations on the Marginal Polytope.....	1153
Hetunandan Kamisetty; Eric Xing; Christopher Langmead	
Active Learning from Crowds.....	1161
Yan Yan; Romer Rosales; Glenn Fung; Jennifer Dy	
Computational Rationalization: The Inverse Equilibrium Problem.....	1169
Kevin Waugh; Brian Ziebart; Drew Bagnell	
Finite-Sample Analysis of Lasso-TD	1177
Mohammad Ghavamzadeh; Alessandro Lazaric; Remi Munos; Matthew Hoffman	
Generalized Value Functions for Large Action Sets.....	1185
Jason Pazis; Ron Parr	

k-DPPs: Fixed-Size Determinantal Point Processes	1193
Alex Kulesza; Ben Taskar	
On Autoencoders and Score Matching for Energy Based Models.....	1201
Kevin Swersky; Marc'Aurelio Ranzato; David Buchman; Benjamin Marlin; Nando Freitas	
Generalized Boosting Algorithms for Convex Optimization.....	1209
Alexander Grubb; Drew Bagnell	