

---

# **ICML 2009**

---



---

**Proceedings,  
Twenty-Sixth International  
Conference on Machine Learning**

---

*Edited by* Léon Bottou and Michael Littman

Copyright © 2009

These materials are owned by their respective copyright owners.  
All rights reserved.

ISBN 978-1-60558-516-1

Cover photo ©Tourisme Montréal, Stéphan Poulin.

Printed in Madison, Wisconsin, United States of America.

# Table of Contents

<b>1. Archipelago: Nonparametric Bayesian Semi-Supervised Learning</b> .....	<b>1</b>
Ryan Adams and Zoubin Ghahramani	
<b>2. Tractable Nonparametric Bayesian Inference in Poisson Processes with Gaussian Process Intensities</b> .....	<b>9</b>
Ryan Adams, Iain Murray, and David MacKay	
<b>3. Route Kernels for Trees</b> .....	<b>17</b>
Fabio Aioli, Giovanni Da San Martino, and Alessandro Sperduti	
<b>4. Incorporating Domain Knowledge into Topic Modeling via Dirichlet Forest Priors</b> .....	<b>25</b>
David Andrzejewski, Xiaojin Zhu, and Mark Craven	
<b>5. Grammatical Inference as a Principal Component Analysis Problem</b> .....	<b>33</b>
Raphaël Bailly, François Denis, and Liva Ralaivola	
<b>6. Curriculum Learning</b> .....	<b>41</b>
Yoshua Bengio, Jérôme Louradour, Ronan Collobert, and Jason Weston	
<b>7. Importance Weighted Active Learning</b> .....	<b>49</b>
Alina Beygelzimer, Sanjoy Dasgupta, and John Langford	
<b>8. Split Variational Inference</b> .....	<b>57</b>
Guillaume Bouchard and Onno Zoeter	
<b>9. Predictive Representations for Policy Gradient in POMDPs</b> .....	<b>65</b>
Abdeslam Boularias and Brahim Chaib-draa	
<b>10. Online Feature Elicitation in Interactive Optimization</b> .....	<b>73</b>
Craig Boutilier, Kevin Regan, and Paolo Viappiani	
<b>11. Spectral Clustering Based on the Graph <math>p</math>-Laplacian</b> .....	<b>81</b>
Thomas Bühler and Matthias Hein	
<b>12. Active Learning for Directed Exploration of Complex Systems</b> .....	<b>89</b>
Michael Burl and Esther Wang	
<b>13. Optimized Expected Information Gain for Nonlinear Dynamical Systems</b> .....	<b>97</b>
Alberto Giovanni Busetto, Cheng Soon Ong, and Joachim M. Buhmann	
<b>14. Probabilistic Dyadic Data Analysis with Local and Global Consistency</b> .....	<b>105</b>
Deng Cai, Xuanhui Wang, and Xiaofei He	
<b>15. Structure Learning of Bayesian Networks using Constraints</b> .....	<b>113</b>
Cassio de Campos, Zhi Zeng, and Qiang Ji	
<b>16. Robust Bounds for Classification via Selective Sampling</b> .....	<b>121</b>
Nicolò Cesa-Bianchi, Claudio Gentile, and Francesco Orabona	
<b>17. Multi-View Clustering via Canonical Correlation Analysis</b> .....	<b>129</b>
Kamalika Chaudhuri, Sham Kakade, Karen Livescu, and Karthik Sridharan	
<b>18. A Convex Formulation for Learning Shared Structures from Multiple Tasks</b> .....	<b>137</b>
Jianhui Chen, Lei Tang, Jun Liu, and Jieping Ye	

<b>19. Learning Kernels from Indefinite Similarities.....</b>	<b>145</b>
Yihua Chen, Maya Gupta, and Benjamin Recht	
<b>20. Matrix Updates for Perceptron Training of Continuous Density Hidden Markov Models.....</b>	<b>153</b>
Chih-Chieh Cheng, Fei Sha, and Lawrence Saul	
<b>21. Decision Tree and Instance-Based Learning for Label Ranking .....</b>	<b>161</b>
Weiwei Cheng, Jens Hühn, and Eyke Hüllermeier	
<b>22. Learning Dictionaries of Stable Autoregressive Models for Audio Scene Analysis .....</b>	<b>169</b>
Youngmin Cho and Lawrence Saul	
<b>23. Exploiting Sparse Markov <i>and</i> Covariance Structure in Multiresolution Models.....</b>	<b>177</b>
Myung Jin Choi, Venkat Chandrasekaran, and Alan Willsky	
<b>24. Nonparametric Estimation of the Precision-Recall Curve .....</b>	<b>185</b>
Stéphane Cléménçon and Nicolas Vayatis	
<b>25. EigenTransfer: A Unified Framework for Transfer Learning.....</b>	<b>193</b>
Wenyuan Dai, Ou Jin, Gui-Rong Xue, Qiang Yang, and Yong Yu	
<b>26. Fitting a Graph to Vector Data.....</b>	<b>201</b>
Samuel Daitch, Jonathan Kelner, and Daniel Spielman	
<b>27. Unsupervised Search-Based Structured Prediction .....</b>	<b>209</b>
Hal Daumé III	
<b>28. Deep Transfer via Second-Order Markov Logic.....</b>	<b>217</b>
Jesse Davis and Pedro Domingos	
<b>29. Analytic Moment-Based Gaussian Process Filtering.....</b>	<b>225</b>
Marc Peter Deisenroth, Marco F. Huber, and Uwe D. Hanebeck	
<b>30. Good Learners for Evil Teachers .....</b>	<b>233</b>
Ofar Dekel and Ohad Shamir	
<b>31. A Scalable Framework for Discovering Coherent Co-clusters in Noisy Data.....</b>	<b>241</b>
Meghana Deodhar, Joydeep Ghosh, Gunjan Gupta, Hyuk Cho, and Inderjit Dhillon	
<b>32. The Adaptive <math>k</math>-Meteorologists Problem and Its Application to Structure Learning and Feature Selection in Reinforcement Learning.....</b>	<b>249</b>
Carlos Diuk, Lihong Li, and Bethany Leffler	
<b>33. Proximal Regularization for Online and Batch Learning.....</b>	<b>257</b>
Chuong Do, Quoc Le, and Chuan-Sheng Foo	
<b>34. Large Margin Training for Hidden Markov Models with Partially Observed States.....</b>	<b>265</b>
Trinh-Minh-Tri Do and Thierry Artières	
<b>35. Accelerated Gibbs Sampling for the Indian Buffet Process .....</b>	<b>273</b>
Finale Doshi-Velez and Zoubin Ghahramani	
<b>36. Accounting for Burstiness in Topic Models .....</b>	<b>281</b>
Gabriel Doyle and Charles Elkan	
<b>37. Domain Adaptation from Multiple Sources via Auxiliary Classifiers.....</b>	<b>289</b>
Lixin Duan, Ivor W. Tsang, Dong Xu, and Tat-Seng Chua	

<b>38. Boosting with Structural Sparsity .....</b>	<b>297</b>
John Duchi and Yoram Singer	
<b>39. Learning to Segment from a Few Well-Selected Training Images .....</b>	<b>305</b>
Alireza Farhangfar, Russ Greiner, and Csaba Szepesvári	
<b>40. GAODE and HAODE: Two Proposals Based on AODE to Deal with Continuous Variables .....</b>	<b>313</b>
M. Julia Flores, José A. Gámez, Ana M. Martínez, and José M. Puerta	
<b>41. A Majorization-Minimization Algorithm for (Multiple) Hyperparameter Learning .....</b>	<b>321</b>
Chuan-Sheng Foo, Chuong Do, and Andrew Ng	
<b>42. Dynamic Mixed Membership Blockmodel for Evolving Networks .....</b>	<b>329</b>
Wenjie Fu, Le Song, and Eric Xing	
<b>43. Gradient Descent with Sparsification: An Iterative Algorithm for Sparse Recovery with Restricted Isometry Property .....</b>	<b>337</b>
Rahul Garg and Rohit Khandekar	
<b>44. Sequential Bayesian Prediction in the Presence of Changepoints .....</b>	<b>345</b>
Roman Garnett, Michael Osborne, and Stephen Roberts	
<b>45. PAC-Bayesian Learning of Linear Classifiers .....</b>	<b>353</b>
Pascal Germain, Alexandre Lacasse, François Laviolette, and Mario Marchand	
<b>46. Fast Evolutionary Maximum Margin Clustering .....</b>	<b>361</b>
Fabian Gieseke, Tapio Pahikkala, and Oliver Kramer	
<b>47. Dynamic Analysis of Multiagent Q-learning with <math>\epsilon</math>-greedy Exploration .....</b>	<b>369</b>
Eduardo Rodrigues Gomes and Ryszard Kowalczyk	
<b>48. Bayesian Inference for Plackett-Luce Ranking Models .....</b>	<b>377</b>
John Guiver and Edward Snelson	
<b>49. Bayesian Clustering for Email Campaign Detection .....</b>	<b>385</b>
Peter Haider and Tobias Scheffer	
<b>50. Efficient Learning Algorithms for Changing Environments .....</b>	<b>393</b>
Elad Hazan and C. Seshadhri	
<b>51. Hoeffding and Bernstein Races for Selecting Policies in Evolutionary Direct Policy Search .....</b>	<b>401</b>
Verena Heidrich-Meisner and Christian Igel	
<b>52. Partially Supervised Feature Selection with Regularized Linear Models .....</b>	<b>409</b>
Thibault Helleputte and Pierre Dupont	
<b>53. Learning with Structured Sparsity .....</b>	<b>417</b>
Junzhou Huang, Tong Zhang, and Dimitris Metaxas	
<b>54. Learning Linear Dynamical Systems without Sequence Information .....</b>	<b>425</b>
Tzu-Kuo Huang and Jeff Schneider	
<b>55. Group Lasso with Overlaps and Graph Lasso .....</b>	<b>433</b>
Laurent Jacob, Guillaume Obozinski, and Jean-Philippe Vert	
<b>56. Graph Construction and <math>b</math>-Matching for Semi-Supervised Learning .....</b>	<b>441</b>
Tony Jebara, Jun Wang, and Shih-Fu Chang	

<b>57. Trajectory Prediction: Learning to Map Situations to Robot Trajectories .....</b>	<b>449</b>
Nikolay Jetchev and Marc Toussaint	
<b>58. An Accelerated Gradient Method for Trace Norm Minimization.....</b>	<b>457</b>
Shuiwang Ji and Jieping Ye	
<b>59. A Novel Lexicalized HMM-Based Learning Framework for Web Opinion Mining .....</b>	<b>465</b>
Wei Jin and Hung Hay Ho	
<b>60. Orbit-Product Representation and Correction of Gaussian Belief Propagation .....</b>	<b>473</b>
Jason Johnson, Vladimir Chernyak, and Michael Chertkov	
<b>61. A Bayesian Approach to Protein Model Quality Assessment.....</b>	<b>481</b>
Hetunandan Kamisetty and Christopher James Langmead	
<b>62. Learning Prediction Suffix Trees with Winnow .....</b>	<b>489</b>
Nikos Karampatziakis and Dexter Kozen	
<b>63. Boosting Products of Base Classifiers .....</b>	<b>497</b>
Balázs Kégl and Róbert Busa-Fekete	
<b>64. Learning Markov Logic Network Structure via Hypergraph Lifting .....</b>	<b>505</b>
Stanley Kok and Pedro Domingos	
<b>65. Near-Bayesian Exploration in Polynomial Time.....</b>	<b>513</b>
J. Zico Kolter and Andrew Ng	
<b>66. Regularization and Feature Selection in Least-Squares Temporal Difference Learning.....</b>	<b>521</b>
J. Zico Kolter and Andrew Ng	
<b>67. The Graphlet Spectrum .....</b>	<b>529</b>
Risi Kondor, Nino Shervashidze, and Karsten Borgwardt	
<b>68. Rule Learning with Monotonicity Constraints .....</b>	<b>537</b>
Wojciech Kotlowski and Roman Slowinski	
<b>69. Multiple Indefinite Kernel Learning with Mixed Norm Regularization.....</b>	<b>545</b>
Matthieu Kowalski, Marie Szafranski, and Liva Ralaivola	
<b>70. On Sampling-Based Approximate Spectral Decomposition .....</b>	<b>553</b>
Sanjiv Kumar, Mehryar Mohri, and Ameet Talwalkar	
<b>71. Learning Spectral Graph Transformations for Link Prediction .....</b>	<b>561</b>
Jérôme Kunegis and Andreas Lommatzsch	
<b>72. Block-Wise Construction of Acyclic Relational Features with Monotone Irreducibility and Relevancy Properties.....</b>	<b>569</b>
Ondřej Kuželka and Filip Železný	
<b>73. Generalization Analysis of Listwise Learning-to-Rank Algorithms.....</b>	<b>577</b>
Yanyan Lan, Tie-Yan Liu, Zhiming Ma, and Hang Li	
<b>74. Approximate Inference for Planning in Stochastic Relational Worlds.....</b>	<b>585</b>
Tobias Lang and Marc Toussaint	
<b>75. Learning Nonlinear Dynamic Models .....</b>	<b>593</b>
John Langford, Ruslan Salakhutdinov, and Tong Zhang	



<b>76. Non-Linear Matrix Factorization with Gaussian Processes .....</b>	<b>601</b>
Neil D. Lawrence and Raquel Urtasun	
<b>77. Convolutional Deep Belief Networks for Scalable Unsupervised Learning of Hierarchical Representations.....</b>	<b>609</b>
Honglak Lee, Roger Grosse, Rajesh Ranganath, and Andrew Ng	
<b>78. Transfer Learning for Collaborative Filtering via a Rating-Matrix Generative Model .....</b>	<b>617</b>
Bin Li, Qiang Yang, and Xiangyang Xue	
<b>79. ABC-Boost: Adaptive Base Class Boost for Multi-Class Classification .....</b>	<b>625</b>
Ping Li	
<b>80. Semi-Supervised Learning Using Label Mean.....</b>	<b>633</b>
Yu-Feng Li, James Kwok, and Zhi-Hua Zhou	
<b>81. Learning from Measurements in Exponential Families .....</b>	<b>641</b>
Percy Liang, Dan Klein, and Michael Jordan	
<b>82. Blockwise Coordinate Descent Procedures for the Multi-Task Lasso, with Applications to Neural Semantic Basis Discovery .....</b>	<b>649</b>
Han Liu, Mark Palatucci, and Jian Zhang	
<b>83. Efficient Euclidean Projections in Linear Time .....</b>	<b>657</b>
Jun Liu and Jieping Ye	
<b>84. Topic-Link LDA: Joint Models of Topic and Author Community.....</b>	<b>665</b>
Yan Liu, Alexandru Niculescu-mizil, and Wojciech Gryc	
<b>85. Geometry-Aware Metric Learning.....</b>	<b>673</b>
Zhengdong Lu, Prateek Jain, and Inderjit Dhillon	
<b>86. Identifying Suspicious URLs: An Application of Large-Scale Online Learning .....</b>	<b>681</b>
Justin Ma, Lawrence K. Saul, Stefan Savage, and Geoffrey M. Voelker	
<b>87. Online Dictionary Learning for Sparse Coding .....</b>	<b>689</b>
Julien Mairal, Francis Bach, Jean Ponce, and Guillermo Sapiro	
<b>88. Proto-Predictive Representation of States with Simple Recurrent Temporal-Difference Networks .....</b>	<b>697</b>
Takaki Makino	
<b>89. Sparse Gaussian Graphical Models with Unknown Block Structure .....</b>	<b>705</b>
Benjamin M. Marlin and Kevin P. Murphy	
<b>90. Polyhedral Outer Approximations with Application to Natural Language Parsing .....</b>	<b>713</b>
André Martins, Noah Smith, and Eric Xing	
<b>91. Partial Order Embedding with Multiple Kernels.....</b>	<b>721</b>
Brian McFee and Gert Lanckriet	
<b>92. Bandit-Based Optimization on Graphs with Application to Library Performance Tuning.....</b>	<b>729</b>
Frédéric de Mesmay, Arpad Rimmel, Yevgen Voronenko, and Markus Püschel	
<b>93. Deep Learning from Temporal Coherence in Video .....</b>	<b>737</b>
Hossein Mobahi, Ronan Collobert, and Jason Weston	

<b>94. Regression by Dependence Minimization and its Application to Causal Inference .....</b>	<b>745</b>
Joris Mooij, Dominik Janzing, Jonas Peters, and Bernhard Schölkopf	
<b>95. Learning Complex Motions by Sequencing Simpler Motion Templates .....</b>	<b>753</b>
Gerhard Neumann, Wolfgang Maass, and Jan Peters	
<b>96. Convex Variational Bayesian Inference for Large Scale Generalized Linear Models .....</b>	<b>761</b>
Hannes Nickisch and Matthias Seeger	
<b>97. Solution Stability in Linear Programming Relaxations: Graph Partitioning and Unsupervised Learning.....</b>	<b>769</b>
Sebastian Nowozin and Stefanie Jegelka	
<b>98. Nonparametric Factor Analysis with Beta Process Priors .....</b>	<b>777</b>
John Paisley and Lawrence Carin	
<b>99. Unsupervised Hierarchical Modeling of Locomotion Styles .....</b>	<b>785</b>
Wei Pan and Lorenzo Torresani	
<b>100. Binary Action Search for Learning Continuous-Action Control Policies .....</b>	<b>793</b>
Jason Pavis and Michail Lagoudakis	
<b>101. Detecting the Direction of Causal Time Series .....</b>	<b>801</b>
Jonas Peters, Dominik Janzing, Arthur Gretton, and Bernhard Schölkopf	
<b>102. Constraint Relaxation in Approximate Linear Programs .....</b>	<b>809</b>
Marek Petrik and Shlomo Zilberstein	
<b>103. Multi-Class Image Segmentation using Conditional Random Fields and Global Classification .....</b>	<b>817</b>
Nils Plath, Marc Toussaint, and Shinichi Nakajima	
<b>104. Learning When to Stop Thinking and Do Something!.....</b>	<b>825</b>
Barnabás Póczos, Yasin Abbasi-Yadkori, Csaba Szepesvári, Russell Greiner, and Nathan Sturtevant	
<b>105. Independent Factor Topic Models.....</b>	<b>833</b>
Duangmanee Putthividhya, Hagai Attias, and Srikantan Nagarajan	
<b>106. An Efficient Sparse Metric Learning in High-Dimensional Space via <math>\ell_1</math>-Penalized Log-Determinant Regularization.....</b>	<b>841</b>
Guo-Jun Qi, Jinhui Tang, Zheng-Jun Zha, Tat-Seng Chua, and Hong-Jiang Zhang	
<b>107. Sparse Higher Order Conditional Random Fields for Improved Sequence Labeling.....</b>	<b>849</b>
Xian Qian, Xiaoqian Jiang, Zhang Qi, Xuanjing Huang, and Lide Wu	
<b>108. An Efficient Projection for <math>l_{\{1,\infty\}}</math> Regularization .....</b>	<b>857</b>
Ariadna Quattoni, Xavier Carreras, Michael Collins, and Trevor Darrell	
<b>109. Nearest Neighbors in High-Dimensional Data: The Emergence and Influence of Hubs .....</b>	<b>865</b>
Milos Radovanovic, Alexandros Nanopoulos, and Mirjana Ivanovic	
<b>110. Large-Scale Deep Unsupervised Learning using Graphics Processors .....</b>	<b>873</b>
Rajat Raina, Anand Madhavan, and Andrew Ng	
<b>111. The Bayesian Group-Lasso for Analyzing Contingency Tables.....</b>	<b>881</b>
Sudhir Raman, Thomas Fuchs, Peter Wild, Edgar Dahl, and Volker Roth	

<b>112. Supervised Learning from Multiple Experts: Whom to Trust when Everyone Lies a Bit .....</b>	<b>889</b>
Vikas Raykar, Shipeng Yu, Linda Zhao, Anna Jerebko, Charles Florin, Gerardo Valadez, Luca Bogoni, and Linda Moy	
<b>113. Surrogate Regret Bounds for Proper Losses .....</b>	<b>897</b>
Mark Reid and Robert Williamson	
<b>114. Learning Structurally Consistent Undirected Probabilistic Graphical Models .....</b>	<b>905</b>
Sushmita Roy, Terran Lane, and Margaret Werner-Washburne	
<b>115. Ranking Interesting Subgroups .....</b>	<b>913</b>
Stefan Rueding	
<b>116. Function Factorization using Warped Gaussian Processes .....</b>	<b>921</b>
Mikkel N. Schmidt	
<b>117. Stochastic Methods for <math>\ell_1</math> Regularized Loss Minimization .....</b>	<b>929</b>
Shai Shalev-Shwartz and Ambuj Tewari	
<b>118. Structure Preserving Embedding .....</b>	<b>937</b>
Blake Shaw and Tony Jebara	
<b>119. Monte-Carlo Simulation Balancing .....</b>	<b>945</b>
David Silver and Gerald Tesauro	
<b>120. Uncertainty Sampling and Transductive Experimental Design for Active Dual Supervision .....</b>	<b>953</b>
Vikas Sindhvani, Prem Melville, and Richard Lawrence	
<b>121. Hilbert Space Embeddings of Conditional Distributions with Applications to Dynamical Systems .....</b>	<b>961</b>
Le Song, Jonathan Huang, Alex Smola, and Kenji Fukumizu	
<b>122. Multi-Assignment Clustering for Boolean Data .....</b>	<b>969</b>
Andreas P. Streich, Mario Frank, David Basin, and Joachim M. Buhmann	
<b>123. A Least Squares Formulation for a Class of Generalized Eigenvalue Problems in Machine Learning .....</b>	<b>977</b>
Liang Sun, Shuiwang Ji, and Jieping Ye	
<b>124. A Simpler Unified Analysis of Budget Perceptrons .....</b>	<b>985</b>
Ilya Sutskever	
<b>125. Fast Gradient-Descent Methods for Temporal-Difference Learning with Linear Function Approximation .....</b>	<b>993</b>
Richard Sutton, Hamid Maei, Doina Precup, Shalabh Bhatnagar, David Silver, Csaba Szepesvari and Eric Wiewiora	
<b>126. Optimistic Initialization and Greediness Lead to Polynomial Time Learning in Factored MDPs .....</b>	<b>1001</b>
István Szita and András Lörincz	
<b>127. Discriminative <math>k</math> Metrics .....</b>	<b>1009</b>
Arthur Szlam and Guillermo Sapiro	
<b>128. Kernelized Value Function Approximation for Reinforcement Learning .....</b>	<b>1017</b>
Gavin Taylor and Ronald Parr	

<b>129. Factored Conditional Restricted Boltzmann Machines for Modeling Motion Style .....</b>	<b>1025</b>
Graham Taylor and Geoffrey Hinton	
<b>130. Using Fast Weights to Improve Persistent Contrastive Divergence .....</b>	<b>1033</b>
Tijmen Tieleman and Geoffrey Hinton	
<b>131. Structure Learning with Independent Non-Identically Distributed Data .....</b>	<b>1041</b>
Robert Tillman	
<b>132. Robot Trajectory Optimization using Approximate Inference .....</b>	<b>1049</b>
Marc Toussaint	
<b>133. Ranking with Ordered Weighted Pairwise Classification .....</b>	<b>1057</b>
Nicolas Usunier, David Buffoni, and Patrick Gallinari	
<b>134. More Generality in Efficient Multiple Kernel Learning .....</b>	<b>1065</b>
Manik Varma and Rakesh Babu	
<b>135. Information Theoretic Measures for Clusterings Comparison: Is a Correction for Chance Necessary? .....</b>	<b>1073</b>
Nguyen Xuan Vinh, Julien Epps, and James Bailey	
<b>136. Model-Free Reinforcement Learning as Mixture Learning .....</b>	<b>1081</b>
Nikos Vlassis and Marc Toussaint	
<b>137. BoltzRank: Learning to Maximize Expected Ranking Gain .....</b>	<b>1089</b>
Maksims Volkovs and Richard Zemel	
<b>138. K-means in Space: A Radiation Sensitivity Evaluation .....</b>	<b>1097</b>
Kiri Wagstaff and Benjamin Bornstein	
<b>139. Evaluation Methods for Topic Models .....</b>	<b>1105</b>
Hanna Wallach, Iain Murray, Ruslan Salakhutdinov, and David Mimno	
<b>140. Feature Hashing for Large Scale Multitask Learning .....</b>	<b>1113</b>
Kilian Weinberger, Anirban Dasgupta, John Langford, Alex Smola, and Josh Attenberg	
<b>141. Herding Dynamic Weights to Learn .....</b>	<b>1121</b>
Max Welling	
<b>142. A Stochastic Memoizer for Sequence Data .....</b>	<b>1129</b>
Frank Wood, Cédric Archambeau, Jan Gasthaus, Lancelot James, and Yee Whye Teh	
<b>143. Optimal Reverse Prediction: A Unified Perspective on Supervised, Unsupervised and Semi-supervised Learning .....</b>	<b>1137</b>
Linli Xu, Martha White, and Dale Schuurmans	
<b>144. Non-Monotonic Feature Selection .....</b>	<b>1145</b>
Zenglin Xu, Rong Jin, Jieping Ye, Michael R. Lyu, and Irwin King	
<b>145. Online Learning by Ellipsoid Method .....</b>	<b>1153</b>
Liu Yang, Rong Jin, and Jieping Ye	
<b>146. Stochastic Search using the Natural Gradient .....</b>	<b>1161</b>
Sun Yi, Daan Wierstra, Tom Schaul, and Juergen Schmidhuber	
<b>147. Learning Structural SVMs with Latent Variables .....</b>	<b>1169</b>
Chun-Nam Yu and Thorsten Joachims	

<b>148. Piecewise-Stationary Bandit Problems with Side Observations .....</b>	<b>1177</b>
Jia Yuan Yu and Shie Mannor	
<b>149. Large-scale Collaborative Prediction Using a Nonparametric Random Effects Model.....</b>	<b>1185</b>
Kai Yu, John Lafferty, Shenghuo Zhu, and Yihong Gong	
<b>150. Robust Feature Extraction via Information Theoretic Learning .....</b>	<b>1193</b>
Xiao-Tong Yuan and Bao-Gang Hu	
<b>151. Interactively Optimizing Information Retrieval Systems as a Dueling Bandits Problem.....</b>	<b>1201</b>
Yisong Yue and Thorsten Joachims	
<b>152. Compositional Noisy-Logical Learning.....</b>	<b>1209</b>
Alan Yuille and Songfeng Zheng	
<b>153. Discovering Options from Example Trajectories.....</b>	<b>1217</b>
Peng Zang, Peng Zhou, David Minnen, and Charles Isbell	
<b>154. Learning Instance Specific Distances Using Metric Propagation.....</b>	<b>1225</b>
De-Chuan Zhan, Ming Li, Yu-Feng Li, and Zhi-Hua Zhou	
<b>155. Prototype Vector Machine for Large Scale Semi-supervised Learning .....</b>	<b>1233</b>
Kai Zhang, James T. Kwok, and Bahram Parvin	
<b>156. Learning Non-Redundant Codebooks for Classifying Complex Objects.....</b>	<b>1241</b>
Wei Zhang, Akshat Surve, Xiaoli Fern, and Thomas Dietterich	
<b>157. Multi-Instance Learning by Treating Instances As Non-I.I.D. Samples.....</b>	<b>1249</b>
Zhi-Hua Zhou, Yu-Yin Sun, and Yu-Feng Li	
<b>158. MedLDA: Maximum Margin Supervised Topic Models for Regression and Classification .....</b>	<b>1257</b>
Jun Zhu, Amr Ahmed, and Eric Xing	
<b>159. On Primal and Dual Sparsity of Markov Networks .....</b>	<b>1265</b>
Jun Zhu and Eric Xing	
<b>160. SimpleNPKL: Simple Non-Parametric Kernel Learning .....</b>	<b>1273</b>
Jinfeng Zhuang, Ivor W. Tsang, and Steven C.H. Hoi	

---

## Preface

---

This volume contains the papers accepted to the 26<sup>th</sup> International Conference on Machine Learning (ICML 2009). ICML is the annual conference of the International Machine Learning Society (IMLS) and provides a venue for the presentation and discussion of current research in the field of machine learning. These proceedings can also be found online at <http://www.machinelearning.org>.

ICML 2009 was held June 14–18 on the downtown campus of McGill University in Montréal, Canada. It was co-located with COLT-2009, the 22<sup>nd</sup> Annual Conference on Computational Learning Theory, and UAI-2009, the 25<sup>th</sup> Conference on Uncertainty in Artificial Intelligence.

Including papers that were ultimately withdrawn, 595 papers were submitted to the conference. The review process was managed by two programme co-chairs, 39 Area Chairs, and 480 dedicated reviewers. A novel review process was instituted in 2009 to further encourage innovative papers on a variety of topics. First, we experimented with “inverse bidding”. In this scheme, Area Chairs described their area descriptors on a webpage and recruited 10 or so reviewers to help them handle papers on these topics. As part of the submission process, authors indicated preferences for Area Chairs to handle their papers. The goal here was to ensure each submission was considered by reviewers most appropriate to the paper’s intended contribution. Aside from a few special cases, papers were assigned to one of the top three Area Chairs selected by the authors.

The next innovation was to use an explicit multi-round reviewing process. Each submitted paper received two first-round reviews. As in recent years, authors had the opportunity to see and respond to these reviews. Papers that garnered at least one positive review in the first round received one or more additional reviews specifically selected to provide definitive feedback for deciding whether to accept the paper. Unfortunately, due to the early scheduling of the conference, there was no time for authors to respond to these additional reviews. Final decisions were made using the input from all reviewers, the author feedback, the Area Chair’s comments, any discussion between the reviewers to try to reach consensus, and, in some cases, close examination by the Programme co-chairs and a secondary Area Chair assigned to help decide borderline cases. Reviewing was blind to the identities of the authors. Conditional accepts were not used this year.

Apart from the length restrictions on papers and the compressed time frame, the review process for ICML resembles that of many journal publications. In total, 160 papers were accepted to ICML this year, including a small number of papers accepted as “food for thought” papers that presented intriguing ideas in spite of some apparent flaws. The overall acceptance rate of 27% matched that of last year’s conference.

ICML authors presented their papers both orally and in an evening poster session, allowing time for detailed discussions with any interested attendees of the conference. One oral session was dedicated to the memory of Paul Utgoff, a machine-learning pioneer and chair

of ICML 1993, who died this year. Each day of the main conference included an invited talk by a prominent researcher. We were very fortunate to be able to host Corinna Cortes, Google Research, NY, Emmanuel Dupoux, Centre National de la Recherche Scientifique, and Yoav Freund, University of California, San Diego. In addition to the technical talks, ICML 2009 also included nine tutorials held before the main conference, presented by top-notch researchers Alina Beygelzimer, John Langford, and Bianca Zadrozny; Eyal Even-Dar and Vahab Mirrokni; Volker Tresp and Kai Yu; Manfred K. Warmuth and S.V.N. Vishwanathan; Yael Niv; Paul Bennett, Misha Bilenko, and Kevyn Collins-Thompson; Sanjoy Dasgupta and John Langford; Jure Leskovec; and Noah Smith. We were delighted to be able to organize our workshops once again jointly with COLT and UAI as part of a special “overlap day,” consisting of nine workshops selected and arranged collaboratively by the respective workshop chairs of the three conferences. This day provided a rich opportunity for interaction among the attendees of the conferences.

This year, ICML continued its award offerings to help build our community, celebrate our advances, and encourage applications and long-term thinking. In addition to our previously traditional “Best Paper” and “Best Student Paper” awards, we also gave awards for “Best Application Paper” and “10-year Best Paper” (for the best paper of ICML 1999). We thank Springer, publisher of *Machine Learning*, for sponsoring our student paper awards.

The organization of ICML 2009 involved efforts from many people, to whom we are extremely grateful. As programme co-chairs, we worked closely with the general chair, Andrea Danyluk, who brought her positive attitude and her experience from for serving as program co-chair of the highly successful ICML 2001 conference. Doina Precup handled local arrangements, including venue, budgeting, website, refreshments, registration, banquet, and a million other things. The tutorials chair, Jennifer Neville, and the workshop chair, Chris Williams, pulled together excellent offerings. The publications chair, Kiri Wagstaff, handled the critical and relatively thankless job of working with authors to edit the accepted papers into a unified volume. We thank Lise Getoor for her remarkable efforts as the funding chair in this financially difficult year; Drew Bagnell and Nicholas Roy, the student funding co-chairs, who dispersed student travel awards; and Joelle Pineau, the volunteer chair, who arranged for student volunteers, and made sure the conference ran smoothly. We also thank Steven Scott, the treasurer of IMLS, for his support and advice on financial issues, and Monica Dinculescu for her contributions as webmaster for the [icml2009.org](http://icml2009.org) website. The staff of Softconf.com, especially Rich Gerber, were extremely helpful in tailoring the START V2 conference-management software to our rather unorthodox reviewing procedure.

For more general support, we are grateful to the members of IMLS for their advice. We are also very grateful to the many financial sponsors of ICML (who are listed elsewhere in these proceedings) for their generous support of this conference.

No technical conference is possible without the efforts of reviewers, so we wish to thank everyone who participated in the process, from the dedicated reviewers, to the Area Chairs, to the authors themselves for their roles in the process. No reviewing procedure is perfect, but to the extent that this year’s program was successful, it was because of those people who applied their prodigious intellectual talents to help maintain and improve the quality of work in our community. We were awed and impressed to see how dependent the research process is on your efforts and how seriously you take your role in keeping this exciting and important research area healthy.

We hope the machine-learning community find this volume a useful resource and we are happy to have had the opportunity to contribute.

Sincerely,

*Léon Bottou and Michael L. Littman*  
*ICML 2009 Programme Co-chairs*





---

# ICML 2009 Organization

---

## **General Chair**

Andrea Danyluk, *Williams College, U.S.A.*

## **Programme Co-chairs**

Léon Bottou, *NEC Laboratories America, U.S.A.*

Michael Littman, *Rutgers University, U.S.A.*

## **Workshops Chair**

Chris Williams, *University of Edinburgh, United Kingdom*

## **Tutorials Chair**

Jennifer Neville, *Purdue University, U.S.A.*

## **Publications Chair**

Kiri L. Wagstaff, *Jet Propulsion Laboratory, California Inst. of Technology, U.S.A.*

## **Volunteers Chair**

Joelle Pineau, *McGill University, Canada*

## **Student Funding Co-chairs**

Drew Bagnell, *Carnegie Mellon University, U.S.A.*

Nicholas Roy, *Massachusetts Institute of Technology, U.S.A.*

## **Fundraising Chair**

Lise Getoor, *University of Maryland, College Park, U.S.A.*

## **Local Arrangements Chair**

Doina Precup, *McGill University, Canada*

## Area Chairs

Naoki Abe	Yasemin Altun	Francis Bach	Samy Bengio
David Blei	Carla Brodley	Florence d'Alche-Buc	Kurt Driessens
Alan Fern	David Forsyth	Johannes Fuernkranz	Kenji Fukumizu
Marie desJardins	John Langford	Mirella Lapata	Neil Lawrence
Yann LeCun	Sofus Attila Macskassy	Yishay Mansour	Steven Minton
Dunja Mladenic	Tim Oates	Mike Pazzani	Massimiliano Pontil
Pascal Poupart	Luc De Raedt	Carl Rasmussen	Martin Riedmiller
Dan Roth	Volker Roth	Michele Sebag	Fei Sha
Yoram Singer	Nathan Srebro	Yee Whye Teh	Luis Torgo
David Wingate	Nevin Zhang	Martin Zinkevich	

## Reviewers

Pieter Abbeel	Jake Abernethy	Ryan Adams	Alekh Agarwal
Deepak Agarwal	Shivani Agarwal	Amrudin Agovic	Babak Ahmadi
Fabio Aiolli	Edoardo Airoldi	Shotaro Akaho	David Albrecht
Kamal Ali	Mauricio Alvarez	Christophe Ambroise	Massih-Reza Amini
Adam Anthony	Annalisa Appice	Cedric Archambeau	Andreas Argyriou
Tom Armstrong	Alexandre d'Aspremont	Jean-Yves Audibert	Peter Auer
Michael Aupetit	Ali Bahramisharif	Gokhan Bakir	Pierre Baldi
Arindam Banerjee	Kobus Barnard	Kevin Bartz	Sugato Basu
Kedar Bellare	Shai Ben-David	Yael Ben-Haim	Yoshua Bengio
Bettina Berendt	Alex Berg	Tamara Berg	Alex Berg
Alina Beygelzimer	Indrajit Bhattacharya	Jinbo Bi	Steffen Bickel
Nicolas Le Bihan	Mikhail Bilenko	Adriana Birlutiu	Ronny Bjarnason
John Blitzer	Hendrik Blockeel	Phil Blunsom	Karsten Borgwardt
Daniel Borrajo	Zoran Bosnić	Henrik Boström	Alexandre Bouchard-Côté
Jordan Boyd-Graber	Janez Brank	Mikio Braun	Ulf Brefeld
Gavin Brown	Nicolas Brunel	Emma Brunskill	Cassio Campos
Joaquin Quiñero Candela	Stephane Canu	Olivier Cappe	Xavier Carreras
Vitor Carvalho	Luca Cazzanti	Tonatiuh Pena Centeno	Nicòlo Cesa-Bianchi
Chia-Hui Chang	Jonathan Chang	Yu-Han Chang	Olivier Chapelle
Kamalika Chaudhuri	Nitesh Chawla	Gal Chechik	Silvia Chiappa
Sumit Chopra	Wei Chu	William Cohen	Trevor Cohn
Ronan Collobert	Corinna Cortes	Koby Crammer	Daniel Cremers
Tom Croonenborghs	Lehel Csato	Botond Cseke	James Cussens
Marco Cuturi	Aaron D'Souza	David L. Dowe	Walter Daelemans
Tom Dean	Marc Deisenroth	Gerald Dejong	Ofer Dekel
Janez Demsar	Meghana Deodhar	Chris Diehl	Nan Ding
Arnaud Doucet	Adam Drake	Greg Druck	Christopher Drummond
John Duchi	Miroslav Dudik	Pierre Dupont	Pinar Duygulu
Jennifer Dy	Saso Dzeroski	Eric Eaton	Jacob Eisenstein
Ran El-Yaniv	Tina Eliassi-Rad	Charles Elkan	Dan Ellis
Barbara Engelhardt	Gwenn Engleblenne	Dumitru Erhan	Damien Ernst
Eyal Even-Dar	Ian Fasel	Andrew Fast	Ad Feelders
Rob Fergus	Xiaoli Fern	Shai Fine	Michael Fink
Jenny Finkel	Bernd Fischer	Francois Fleuret	Blaz Fortuna
Charless Fowlkes	Eibe Frank	Nando de Freitas	Yoav Freund
Jurgen Van Gael	Thomas Gaertner	Brian Gallagher	Fred Garcia
Natalia Gardiol	Matthew Gardner	Cenk Gazen	Melinda Gervasio

Pierre Geurts	Rayid Ghani	Mohammad Ghavamzadeh	Ali Ghodsi
Remi Gilleron	Attilio Giordana	Christophe Giraud-Carrier	Mark Girolami
Robert Givan	Amir Globerson	Xavier Glorot	Jacob Goldberger
Sally Goldman	Nicholas Gorski	Dilan Gorur	Thore Graepel
Yves Grandvalet	David Grangier	Russ Greiner	Arthur Gretton
Tom Griffiths	Marko Grobelnik	Amit Gruber	Peter Grunwald
Yuhong Guo	Amaury Habrard	Raia Hadsell	Patrick Haffner
Peter Haider	Barbara Hammer	Steve Hanneke	Katja Hansen
Nikolaus Hansen	Zaid Harchaoui	David Hardoon	Stefan Harmeling
Tamir Hazan	Matthias Hein	Katherine Heller	Tom Heskes
Shohei Hido	Colin de la Higuera	Julia Hockenmaier	Thomas Hofmann
Derek Hoiem	Tamas Horvath	Andreas Hotho	Daniel Hsu
Pei-Yun Hsueh	Jens Humrich	Tuyen Huynh	Aapo Hyvarinen
Eyke Hüllermeier	Hal Daume III	Tsuyoshi Ide	Eugene Ie
Christian Igel	Alexander Ihler	Shiro Ikeda	Akihiro Inokuchi
Nathan Intrator	Laurent Jacob	Manfred Jaeger	Viren Jain
Stefanie Jegelka	David Jensen	Qiang Ji	Thorsten Joachims
Jeffrey Johns	Nick Jong	Murray Jorgensen	Adam Tauman Kalai
Satyen Kale	Alexandros Kalousis	Shivaram Kalyanakrishnan	Takafumi Kanamori
Bert Kappen	Nikos Karampatziakis	Hisashi Kashima	Samuel Kaski
Motoaki Kawanabe	Balazs Kegl	Eamonn Keogh	Kristian Kersting
Roni Khardon	Alexandre Klementiev	Marius Kloft	Levente Kocsis
Pranam Kolari	Alek Kolcz	Risi Kondor	George Konidaris
Igor Kononenko	Kevin Korb	Robert Krauthgamer	Brian Kulis
Ravi Kumar	Ondrej Kuzelka	Dima Kuzmin	James Kwok
Simon Lacoste-Julien	Michail Lagoudakis	Wai Lam	Nicolas Lambert
Christoph Lampert	Niels Landwehr	Tobias Lang	Helge Langseth
Hugo Larochelle	Pavel Laskov	Martin Lauer	Nada Lavrac
Victor Lavrenko	Alessandro Lazaric	Svetlana Lazebnik	Quoc Le
Guy Lebanon	Honglak Lee	Jun won Lee	Gayle Leen
Oliver Lemon	Kristina Lerman	Jure Leskovec	Geoffrey Levine
Lihong Li	Wei Li	Percy Liang	Wenhui Liao
Katrina Ligett	Chih-Jen Lin	Yuanqing Lin	Charles Ling
Michael Littman	Han Liu	Yan Liu	Dan Lizotte
Nicolas Loeff	Phil Long	Gaelle Loosli	Aurelie Lozano
Rich Maclin	Omid Madani	Sridhar Mahadevan	Markus Maier
Julien Mairal	Donato Malerba	Marcus Maloof	Shie Mannor
Mario Marchand	Oded Margalit	Dragos Margineantu	Shaul Markovitch
Lluís Marquez	Bhaskara Marthi	David McAllester	Ryan McDonald
Brendan McMahan	Andrew McCallum	Neville Mehta	Marina Meila
Prem Melville	Srujana Merugu	Matt Michelson	Lilyana Mihalkova
Brian Milch	Kurt Tadayuki Miller	Marta Milo	David Mimno
Andriy Mnih	Daichi Mochihashi	Shakir Mohamed	Mehryar Mohri
Michi Momma	Joris Mooij	Raymond Mooney	João Mendes Moreira
Katharina Morik	Clay Morrison	Alessandro Moschitti	Sourav Mukherjee
Remi Munos	Hector Munoz-Avila	Noboru Murata	Kevin Murphy
Iain Murray	Ion Muslea	Ramesh Nallapati	Sriaram Natarajan
Jennifer Neville	David Newman	Ann Nicholson	Hannes Nickisch
Alexandru Niculescu-Mizil	Siegfried Nijssen	Mahesan Niranjan	Richard Nock
Ann Nowe	Sebastian Nowozin	Shigeyuki Oba	Guillaume Obozinski
Emanuele Olivetti	Cheng Soon Ong	Peter Orbanz	Sarah Osentoski
Simon Osindero	Martijn Van Otterlo	Patrick de Oude	David Page
Jean-Francois Paiement	Thomas Palmer	Charles Parker	Andrea Passerini
Hélène Paugam-Moisy	Michael Pazzani	Dmitry Pechyony	Elzbieta Pekalska

Jose Pena	Fernando Perez-Cruz	Claudia Perlich	Jan Peters
Slav Petrov	Bernhard Pfahringer	Joelle Pineau	Christian Plagemann
Vid Podpecan	Jean Ponce	Doina Precup	Vasin Punyakanok
Emmanuel Rachelson	Yaron Rachlin	Ali Rahimi	Alain Rakotomamonjy
Liva Ralaivola	Ganesh Ramakrishnan	Sudhir Raman	Deva Ramanan
Jan Ramon	MarcAurelio Ranzato	Carl Edward Rasmussen	Soumya Ray
Ben Recht	Xiaofeng Ren	Sebastian Riedel	Irina Rish
Marko Robnik-Sikonja	Lorenzo Rosasco	Saharon Rosset	Fabrice Rossi
Afshin Rostami	Juho Rousu	Benjamin Rubinstein	Cynthia Rudin
Ulrich Rückert	Stefan Rüping	Maytal Saar-Tsechansky	Yunus Saatci
Yvan Saey	Mehran Sahami	Paul Sajda	Ruslan Salakhutdinov
Scott Sanner	Taisuke Sato	Christoph Sawade	Robert Schapire
Tobias Scheffer	Bruno Scherrer	Cordelia Schmid	Mikkel Schmidt
Matt Schmill	Marc Sebban	Matthias Seeger	Sathya Selvaraj
Kevin Seppi	Gregory Shakhnarovich	Shai Shalev-Shwartz	Ohad Shamir
Hanhuai Shan	Jude Shavlik	John Shawe-Taylor	Alexander Sherstov
Yuichi Shiraishi	Ricardo Silva	David Silver	Ozgur Simsek
Vikas Sindhwani	Tomas Singliar	Kevin Small	David Smith
Matthew Smith	Alex Smola	Barry Smyth	Carlos Soares
Le Song	Soeren Sonnenburg	David Sontag	Finnegan Southey
Diana Spears	Dan Spielman	Suvrit Sra	Nati Srebro
Karthik Sridharan	Ashwin Srinivasan	NIsheeth Srivastava	Oliver Stegle
Florian Steinke	David Stern	David Stracuzzi	Matt Streeter
Alex Strehl	Nathan Sturtevant	Erik Sudderth	Masashi Sugiyama
Charles Sutton	Johan Suykens	Taiji Suzuki	Csaba Szepesvari
Prasad Tadepalli	Junichi Takeuchi	Erik Talvitie	Shu Tanaka
Jie Tang	Brian Tanner	Pham Dinh Tao	Matthew Taylor
Olivier Teytaud	Bo Thiesson	Jin Tian	Kai Ming Ting
Mike Tipping	Peter Tischer	Ivan Titov	Michalis Titsias
Hannu Toivonen	Ryota Tomioka	Lisa Torrey	Andrea Torsello
Marc Toussaint	Kristina Toutanova	Bill Triggs	Erik Trumbelj
Ivor Tsang	Koji Tsuda	Karl Tuyls	Raquel Urtasun
Jarno Vanhatalo	Manik Varma	Sergei Vassilivitski	Michel Verleysen
Jean-Philippe Vert	S V N Vishwanathan	Michalis Vlachos	Nikos Vlassis
Kiri Wagstaff	Hanna Wallach	Li-Lun Wang	Tao Wang
Xuerui Wang	Geoff Webb	Louis Wehenkel	Kilian Weinberger
Gary Weiss	Jason Weston	Adam White	Shimon Whiteson
Yair Wiener	Marco Wiering	Eric Wiewiora	John Winn
Britton Wolfe	Frank Wood	Jenn Wortman	Stefan Wrobel
Hui Xiong	Zhao Xu	Fusun Yaman	Yoshihiro Yamanishi
Liu Yang	Jieping Ye	Dit-Yan Yeung	Scott Wen-tau Yih
Cheung Yiuming	Sungwook Yoon	Kai Yu	Ming Yuan
Yisong Yue	Francois Yvon	Bianca Zadrozny	Mikhail Zaslavskiy
Filip Zelezny	ChengXiang Zhai	Duo Zhang	Hau Helen Zhang
Nevin L. Zhang	Tong Zhang	Xinhua Zhang	Alice Zheng
Dengyong Zhou	Xiaojin Zhu		

## Student Volunteers

Amin Atrash	James Bergstra	Abdeslam Boularias	Thomas Buhler
Keith Bush	Louis-Charles Caron	Pablo Castro	Maria Fernanda Caropreso
Laurent Charlin	Hamid Chinaei	Olivier Delalleau	Meltem Demirkus
Guillaume Desjardins	Dumitru Erhan	Mahdi Milani Fard	Pascal Germain
Sebastien Giguere	Marcos Ginestra	Yogesh Girdhar	Xavier Glorot
Arthur Guez	Masoumeh Izadi	Ou Jin	Fabian Kaelin
Robert Kaplow	George Konidaris	Alexandre Lacasse	Alexandre Lacoste
Hong Li	Jun Liu	Dan Lizotte	Zhengdong Lu
Pierre-Antoine Manzagol	Malika Meghjani	Cosmin Paduraru	Mark Palatucci
Olivier Remillard	Francois Rivest	Javad Sadri	Guillaume Saulnier
Lei Tang	Jonathan Taylor	Claudius Todor	Robert Vincent
Robert West	Chun-Nam Yu	Jia Yuan Yu	Zhi Zeng
De-Chuan Zhan	Kai Zhang	Jun Zhu	

## Website

Monica Dinculescu      Jordan Frank

## Logo and Cover Design

Marilyne Lavoie

## Registration System

Marie Perreault

## Board of the International Machine Learning Society

<b>President</b> , Raymond J. Mooney	<b>Secretary</b> , Satinder Singh	<b>Treasurer</b> , Stephen Scott	
<b>Funding co-chairs</b> , Rich Caruana and Lise Getoor			
William Cohen	Luc De Raedt	Tom Dietterich	Zoubin Ghahramani
Rob Holte	Michael Jordan	Pat Langley	Nada Lavrač
Andrew McCallum	Andrew Moore	Foster Provost	Stuart Russell
Rob Schapire	Bernhard Schölkopf	Ian Witten	Stefan Wrobel

## Governing Bodies

The International Machine Learning Society (IMLS), <http://www.machinelearning.org>.



---

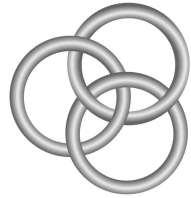
# ICML 2009 Funding

---

The organizers of ICML 2009 would like to thank the following sponsors:

## Platinum

Microsoft®  
**Research**



**MITACS**



## Silver

**GERAD** Groupe d'études et de recherche  
en analyse des décisions

**Google**

**YAHOO!**  
LABS

## Supporting

 ALBERTA INGENUITY CENTRE FOR  
MACHINE LEARNING

**C** CENTRE  
**R** DE RECHERCHES  
**M** MATHÉMATIQUES

**IBM**®



**SIEMENS**





---

**ICML 2009 Invited Talks**

---



# Can learning kernels help performance?

*Corinna Cortes*

*Google Research, U.S.A.*

## **Abstract:**

Kernel methods combined with large-margin learning algorithms such as SVMs have been used successfully to tackle a variety of learning tasks since their introduction in the early 90s. However, in the standard framework of these methods, the choice of an appropriate kernel is left to the user and a poor selection may lead to sub-optimal performance. Instead, sample points can be used to select a kernel function suitable for the task out of a family of kernels fixed by the user. While this is an appealing idea supported by some recent theoretical guarantees, in experiments, it has proven surprisingly difficult to consistently and significantly outperform simple fixed combination schemes of kernels. This talk will survey different methods and algorithms for learning kernels and will present novel results that tend to suggest that significant performance improvements can be obtained with a large number of kernels.

(Includes joint work with Mehryar Mohri and Afshin Rostamizadeh.)

## **Biography:**

Corinna Cortes is the Head of Google Research, NY, where she is working on a broad range of theoretical and applied large-scale machine learning problems. Prior to Google, Corinna spent more than ten years at AT&T Labs Research, formerly AT&T Bell Labs, where she held a distinguished research position. Corinna's research work is well-known in particular for her contributions to the theoretical foundations of support vector machines (SVMs) and her work on data-mining in very large data sets for which she was awarded the AT&T Science and Technology Medal in the year 2000. Corinna received her MS degree in Physics from the Niels Bohr Institute in Copenhagen and joined AT&T Bell Labs as a researcher in 1989. She received her Ph.D. in computer science from the University of Rochester in 1993. Corinna is also a competitive runner, placing third in the More Marathon in New York City in 2005, and a mother of two.

# How do infants bootstrap into spoken language?: Models and challenges

*Emmanuel Dupoux*

*Ecole Normale Supérieure, Ecole des Hautes Etudes en Sciences Sociales,  
Centre National de la Recherche Scientifique, France*

## **Abstract:**

Human infants learn spontaneously and effortlessly the language(s) spoken in their environments, despite the extraordinary complexity of the task. Here, I will present an overview of the early phases of language acquisition and focus on one area where a modeling approach is currently being conducted using tools of signal processing and automatic speech recognition: the unsupervised acquisition of phonetic categories. During their first year of life, infants construct a detailed representation of the phonemes of their native language and lose the ability to distinguish nonnative phonemic contrasts. Unsupervised statistical clustering is not sufficient; it does not converge on the inventory of phonemes, but rather on contextual allophonic units or subunits. I present an information-theoretic algorithm that groups together allophonic variants based on three sources of information that can be acquired independently: the statistical distribution of their contexts, the phonetic plausibility of the grouping, and the existence of lexical minimal pairs. This algorithm is tested on several natural speech corpora. We find that these three sources of information are probably not language specific. What is presumably unique to language is the way in which they are combined to optimize the emergence of linguistic categories.

## **Biography:**

Emmanuel Dupoux is the director of the Laboratoire de Sciences Cognitives et Psycholinguistique in Paris. He conducts research on the early phases of language and social acquisition in human infants, using a mix of behavioral and brain-imaging techniques as well as computational modeling. He teaches at the Ecole des Hautes Etudes en Sciences Sociales where he has set up an interdisciplinary graduate program in Cognitive Science.

# Drifting games, boosting and online learning

*Yoav Freund*

*University of California, San Diego, U.S.A.*

## **Abstract:**

Drifting games is a mathematical framework for modeling learning problems. In this talk I will present the framework and show how it is used to derive a new boosting algorithm called Robustboost and a new online prediction algorithm called NormalHedge. I will present two sets of experiments using these algorithms on synthetic and real world data. The first experiments demonstrate that Robustboost outperforms Adaboost and Logitboost when there are many outliers in the training data. The second set of experiments demonstrate that a tracking algorithm based on NormalHedge is more robust against noise than particle filters.

## **Biography:**

Yoav Freund is a professor of Computer Science and Engineering in the University of California, San Diego. His work is in the areas of machine learning, computational statistics, information theory and their applications. He is best known for his joint work with Dr. Robert Schapire on the Adaboost algorithm. For this work Freund and Schapire were awarded the 2003 Godel Prize and the 2004 Kanellakis Prize. Freund was elected fellow of AAAI in 2008. Freund is included in the Thompson list of most highly cited scientists: ISIHighlyCited.com

.



---

**ICML 2009**  
**Tutorial and Workshop Summaries**

---





---

## Overview of Tutorials and Workshops

---

As in previous years we were pleased to have a strong program of tutorials for ICML 2009. These were held on June 14, immediately preceding the main conference. The program featured nine tutorials covering a wide range of methods for, and applications of, machine learning. There were tutorials on: active learning (Dasgupta, Langford); convergence of natural dynamics in multi-agent games (Even-Dar, Mirrokni); machine learning for large social and information networks (Leskovec); learning with dependencies between several response variables (Tresp, Yu); machine learning in information retrieval (Bennett, Bilenko, Collins-Thompson); the neuroscience of reinforcement learning (Niv); reductions in machine learning (Beygelzimer, Langford, Zadrozny); structured prediction for natural language processing (Smith); and a survey of boosting from an optimization perspective (Warmuth, Vishwanathan). We would like to thank the community for the high-quality tutorial proposals that were received, the presenters for their extensive efforts in preparing and delivering the selected tutorials, and the local arrangements, program, and general chairs of ICML for their hard work in organizing such a stimulating conference.

*Jennifer Neville*  
*ICML 2009 Tutorial Chair*

Once again, ICML solicited and hosted world-class workshops on topics related to machine learning. This year, we were delighted to collaborate with the program co-chairs of UAI (Jeff Bilmes and Andrew Ng) and the COLT workshops chair (Sham Kakade) to put together an exciting joint program. We constructed a slate of nine workshops that represent a wide range of perspectives and fields, as seen in the summaries below. All workshops were held on June 18th, immediately after the main conference days. We would like to thank all of the workshop organizers for their service to the community in putting together these high-quality meetings. We also thank the outstanding local arrangement chairs and the general and program chairs for ICML and the other conferences for creating another exciting and successful conference.

*Chris Williams*  
*ICML 2009 Workshop Chair*

---

# Tutorials

---

## **T1: Reductions in Machine Learning**

*Alina Beygelzimer, IBM T. J. Watson Research Center, U.S.A.  
John Langford, Yahoo! Research, U.S.A.  
Bianca Zadrozny, Fluminense Federal University, Brazil*

Machine learning reductions are about reusing solutions to simple, core problems in order to solve more complex problems. A basic difficulty in applying machine learning in practice is that we often need to solve problems that don't quite match the problems solved by standard machine learning algorithms. Reductions are techniques that transform such practical problems into core machine learning problems. These can then be solved using any existing learning algorithm whose solution can, in turn, be used to solve the original problem. The material that we plan to cover is both algorithmic and analytic: We will discuss existing and new algorithms along with the methodology for analyzing and creating new reductions. In our experience, this approach is an effective tool for designing empirically successful, automated solutions to learning problems.

## **T2: Convergence of Natural Dynamics to Equilibria**

*Eyal Even-Dar, Google, U.S.A.  
Vahab Mirrokni, Google, U.S.A.*

Recently, a lot of effort has been devoted to analyzing response dynamics in various games. Questions about the dynamics themselves and their convergence properties attracted a great deal of attention. This includes, for example, questions like "How long do uncoordinated agents need to reach an equilibrium?" and "Do uncoordinated agents quickly reach a state with low social cost?". An important aspect in studying such dynamics is the learning model employed by self-interested agents in these models. Studying the effect of learning algorithms on the convergence rate of players is crucial for developing a solid understanding of the corresponding games. In this tutorial, we first describe an overview of the required terminology from game theory. Then, we survey results about the convergence of myopic and learning-based best responses of players to equilibria and approximately optimal solutions, and study the effect of various learning algorithms in convergence (rate). Throughout the tutorial, we describe fundamental connections between local search algorithms and learning algorithms with the convergence of best-response dynamics in multi-agent games.

### **T3: Learning with Dependencies between Several Response Variables**

*Volker Tresp, Siemens Corporate Technologies, U.S.A.  
Kai Yu, NEC Laboratories, U.S.A.*

We analyze situations where modeling several response variables for a given input improves the prediction accuracy for each individual response variable. Interestingly, this setting has appeared in different context and a number of different but related approaches have been proposed. In all these approaches some assumptions about the dependency structure between the response variables is made. Here is a small selection of labels describing relevant work: multitask learning, multi-class classification, multi-label prediction, hierarchical Bayes, inductive transfer learning, hierarchical linear models, mixed effect models, partial least squares, canonical correlation analysis, maximal covariance regression, multivariate regression, structured prediction, relational learning, . . . The large number of approaches is confusing for the novice, and often even for the expert. In this tutorial we systematically introduce some of the major approaches and describe them from a common viewpoint.

### **T4: Survey of Boosting from an Optimization Perspective**

*Manfred K. Warmuth, University of California, Santa Cruz, U.S.A.  
S.V.N. Vishwanathan, Purdue University, U.S.A.*

Boosting has become a well known ensemble method. The algorithm maintains a distribution on the binary labeled examples and a new base learner is added in a greedy fashion. The goal is to obtain a small linear combination of base learners that clearly separates the examples. We focus on a recent view of Boosting where the update algorithm for distribution on the examples is characterized by a minimization problem that uses a relative entropy as a regularization. The most well known boosting algorithms is AdaBoost. This algorithm approximately maximizes the hard margin, when the data is separable. We focus on recent algorithms that provably maximize the soft margin when the data is noisy. We will teach the new algorithms, give a unified and versatile view of Boosting in terms of relative entropy regularization, and show how to solve large scale problems based on state of the art optimization techniques.

## **T5: The Neuroscience of Reinforcement Learning**

*Yael Niv, Princeton University, U.S.A.*

One of the most influential contributions of machine learning to understanding the human brain is the (fairly recent) formulation of learning in real world tasks in terms of the computational framework of reinforcement learning. This confluence of ideas is not limited to abstract ideas about how trial and error learning should proceed, but rather, current views regarding the computational roles of extremely important brain substances (such as dopamine) and brain areas (such as the basal ganglia) draw heavily from reinforcement learning. The results of this growing line of research stand to contribute not only to neuroscience and psychology, but also to machine learning: human and animal brains are remarkably adept at learning new tasks in an uncertain, dynamic and extremely complex world. Understanding how the brain implements reinforcement learning efficiently may suggest similar solutions to engineering and artificial intelligent problems. This tutorial will present the current state of the study of neural reinforcement learning, with an emphasis on both what it teaches us about the brain, and what it teaches us about reinforcement learning.

## **T6: Machine Learning in IR: Recent Successes and New Opportunities**

*Paul Bennett, Microsoft Research, U.S.A*

*Misha Bilenko, Microsoft Research, U.S.A*

*Kevyn Collins-Thompson, Microsoft Research, U.S.A*

This tutorial will focus on the interplay between information retrieval (IR) and machine learning. The intersection of these research areas has seen tremendous growth and progress in recent years, much of it fueled by incorporating machine learning techniques into the core of information retrieval technologies, including Web search engines, e-mail and news filtering systems, music and movie recommendations, online advertising systems, and many others. As the complexity, scale, and user expectations for retrieval technology increase, it is becoming increasingly important for each field to keep pace with and inform the other. With that goal in mind, this tutorial covers: the nature of the challenging learning problems faced at many levels by search technology systems today; successful applications of machine learning methods to key IR tasks; and opportunities for joint future progress and emerging research problems which will benefit both machine learning and information retrieval.

## **T7: Active Learning**

*Sanjoy Dasgupta, University of California, San Diego, U.S.A.  
John Langford, Yahoo! Research, U.S.A.*

Active learning is defined by contrast to the passive model of supervised learning where all the labels for learning are obtained without reference to the learning algorithm, while in active learning the learner interactively chooses which data points to label. The hope of active learning is that interaction can substantially reduce the number of labels required, making solving problems via machine learning more practical. This hope is known to be valid in certain special cases, both empirically and theoretically. Variants of active learning has been investigated over several decades and fields. The focus of this tutorial is on general techniques which are applicable to many problems. At a mathematical level, this corresponds to approaches with provable guarantees under weakest-possible assumptions since real problems are more likely to fit algorithms which work under weak assumptions. We believe this tutorial should be of broad interest. People working on or using supervised learning are often confronted with the need for more labels, where active learning can help. Similarly, in reinforcement learning, generalizing while interacting in more complex ways is an active research topic.

## **T8: Large Social and Information Networks: Opportunities for ML**

*Jure Leskovec, Carnegie Mellon University, U.S.A.*

Emergence of the web, social media and online social networking websites gave rise to detailed traces of human social activity. This offers many opportunities to analyze and model behaviors of millions of people. For example, we can now study “planetary scale” dynamics of a full Microsoft Instant Messenger network of 240 million people, with more than 255 billion exchanged messages per month. Many types of data, especially web and “social” data, come in a form of a network or a graph. This tutorial will cover several aspects of such network data: macroscopic properties of network datasets; statistical models for modeling large scale network structure of static and dynamic networks; properties and models of network structure and evolution at the level of groups of nodes and algorithms for extracting such structures. I will also present several applications and case studies of blogs, instant messaging, Wikipedia and web search. Machine learning as a topic will be present throughout the tutorial. The idea of the tutorial is to introduce the machine learning community to recent developments in the area of social and information networks that underpin the Web and other on-line media.

## **T9: Structured Prediction for Natural Language Processing**

*Noah Smith, Carnegie Mellon University, U.S.A.*

This tutorial will discuss the use of structured prediction methods from machine learning in natural language processing. The field of NLP has, in the past two decades, come to simultaneously rely on and challenge the field of machine learning. Statistical methods now dominate NLP, and have moved the field forward substantially, opening up new possibilities for the exploitation of data in developing NLP components and applications. However, formulations of NLP problems are often simplified for computational or practical convenience, at the expense of system performance. This tutorial aims to introduce several structured prediction problems from NLP, current solutions, and challenges that lie ahead. Applications in NLP are a mainstay at ICML conferences; many ML researchers view NLP as a primary or secondary application area of interest. This tutorial will help the broader ML community understand this important application area, how progress is measured, and the trade-offs that make it a challenge.

---

# Workshops

---

## **W1: Seventh Annual Workshop on Bayes Applications**

*John Mark Agosta, Intel Corp., U.S.A.*  
*Russell Almond, Educational Testing Service, U.S.A.*  
*Dennis Buede, Innovative Decisions, U.S.A.*  
*Marek J. Druzdzal, University of Pittsburgh, U.S.A.*  
*Judy Goldsmith, University of Kentucky, U.S.A.*  
*Silja Renooij, Universiteit Utrecht, The Netherlands*

The Bayes Applications Workshop presents projects in several areas where researchers have demonstrated the use of Bayes networks / graphical models in business, military, education and product development. To work in the real world, researchers often have to integrate such analytic techniques with the needs of larger systems and to reconcile apparently conflicting demands of theory and practice. Fortunately the availability of mature academic and commercial software tools has spawned numerous opportunities for such attempts; the stories of how this has been accomplished reveal valuable lessons for all.

## **W2: Automated Interpretation and Modelling of Cell Images**

*Robert F. Murphy, Carnegie Mellon University, U.S.A.*  
*Chun-Nan Hsu, Academia Sinica, Taiwan*  
*Loris Nanni, University of Bologna, Italy*

Dramatic advances in fluorescent probe development, new fluorescence microscope designs to achieve greatly improved temporal and spatial resolution, and significant advances in digital camera and computer technology have enabled increasing use of fluorescence microscopy for quantitative, large scale studies of cell behavior. The high volume and high quality of images resulting from these studies has created and will continue to create many opportunities for computational analysis, especially in the realm of computer vision, machine learning and UAI. This workshop is to bring together interdisciplinary researchers to present and discuss emerging challenges and research issues that arise when realizing fully-automated intelligent analysis of cell images due to recent advances in cell imaging capabilities to discover new biological knowledge about cell structure and function. Discussions of new issues overlooked in the major conferences will be especially encouraged.



### **W3: Workshop on Learning Feature Hierarchies**

*Kay Yu, NEC Laboratories America, U.S.A.*  
*Ruslan Salakhutdinov, University of Toronto, Canada*  
*Yann LeCun, New York University, U.S.A.*  
*Geoff Hinton, University of Toronto, Canada*  
*Yoshua Bengio, University of Montreal, Canada*

Building intelligent systems that are capable of extracting high-level representations from high-dimensional sensory data lies at the core of solving many AI related tasks, including object recognition, speech perception, and language understanding. Theoretical and biological arguments strongly suggest that building such systems requires deep architectures that involve many layers of nonlinear processing. Recent research in machine learning has seen a notable advance in learning feature hierarchies via deep architectures from labeled and unlabeled data. The learned high-level representations have been shown to give promising results in many challenging supervised learning problems, where data patterns often exhibit a high degree of variations. Through having a series of invited talks, a poster session, and a panel discussion, this workshop is expected to assess the current state of the field, discuss key challenges, and identify future promising directions of investigation.

### **W4: Results of the 2009 Reinforcement Learning Competition**

*David Wingate, Massachusetts Institute of Technology, U.S.A.*  
*Carlos Diuk, Rutgers University, U.S.A.*  
*Lihong Li, Rutgers University, U.S.A.*  
*Matthew Taylor, University of Southern California, U.S.A.*  
*Jordan Frank, McGill University, Canada*

The annual Reinforcement Learning competition invites researchers from around the world to apply their latest methods to a suite of exciting and diverse challenge problems. The aim of the competition is to facilitate direct comparisons between learning methods on important and realistic domains. This competition can stimulate development and verification of increasingly practical algorithms on events like tetris, helicopter control, and real-time strategy environments. The 2009 Reinforcement Learning ICML workshop will feature the results of the competition, presentations by competitors regarding their methods, insights, and challenges they overcame, as well as invited speakers and a poster session.

## **W5: The Fourth Workshop on Evaluation Methods for Machine Learning**

*Chris Drummond, NRC Institute for Information Technology, Canada*

*Nathalie Japkowicz, University of Ottawa, Canada*

*William Klement, University of Ottawa, Canada*

*Sofus Macskassy, Fetch Technologies, U.S.A.*

The fourth in a series, this workshop intends to continue the debate within the machine learning community into how we evaluate new algorithms. We aim to discuss what properties of an algorithm need to be evaluated (e.g., accuracy, comprehensibility, conciseness); to solicit views and suggestions for other approaches than those currently used; to investigate alternate methods that could be useful. The three previous workshops focused on issues that have captured the interest of the community, such as; the role of experiments in evaluation, the use of one, community wide, evaluation measure (e.g., Accuracy, AUC, F-measure), the relevance of statistical tests to evaluation, the effectiveness of the UCI data sets for evaluation, the need for sharing and characterizing benchmark data sets in general, and how to promote the views of this workshop to the rest of the community. The 2008 ICML workshop concluded with an agreement that we, as a scientific community, should substantially change how evaluation is performed in machine learning. We, however, disagreed on the direction that this change should take. As a continuation of the same theme, this workshop aims to solicit views, intuitions and visions of alternatives to change existing evaluation methods. We hope to make progress but still carry forward the good methods and experiences we already have acquired.

## **W6: On-line Learning with Limited Feedback**

*Jean-Yves Audibert, Université Paris-Est, France*

*Peter Auer, University of Leoben, Austria*

*Alessandro Lazaric, INRIA, France*

*Remi Munos, INRIA, France*

*Daniil Ryabko, INRIA, France*

*Csaba Szepesvári, University of Alberta, Canada*

The main focus of the workshop is the problem of on-line learning when only limited feedback is available to the learner. In on-line learning, at each time step the learner has to predict the outcome corresponding to the next input based on the feedbacks obtained so far. Unlike the usual supervised problem, in which after each prediction the learner is revealed sufficient information to evaluate the goodness of all predictions he could have made, in many cases only limited feedback may be available to the learner. Depending on the nature of the limitation on the feedback, different classes of problems can be identified, such as reinforcement learning, on-line control problems, multi-armed bandits, indirect feedback. Although some aspects of on-line learning with limited feedback have been already thoroughly analyzed (e.g., multi-armed bandit problems), many problems are still open. For instance, bandits with large action spaces and side information, learning with delayed reward, on-line optimization, etc., are of primary concern in many recent works on on-line learning. Furthermore, on-line learning with limited feedback has strong connections with a number of other fields of Machine Learning such as active learning, semi-supervised learning, and multi-class classification. The goal of the workshop is to provide researchers with the possibility to present their current research on these topics and to encourage the discussion about the main open issues and the possible connections between the different sub-fields.

## **W7: Numerical Mathematics in Machine Learning**

*Matthias Seeger, Saarland University and Max Planck Institute for Informatics, Germany*

*Suvrit Sra, Max-Planck Institute for Biological Cybernetics, Germany*

*John P. Cunningham, Stanford University, U.S.A.*

Many machine learning methods naturally reduce to numerical mathematics algorithms, such as conjugate gradients, Lanczos, matrix factorizations, ODE solvers, or quadrature. These techniques are often used without proper understanding of numerical stability issues or knowledge of techniques to improve convergence (such as preconditioning). By bringing together experts from both fields, we aim to identify major gaps along this interface of growing importance, to find tractable remedies, and to feed back specific machine learning demands to numerical mathematicians.

## **W8: Abstraction in Reinforcement Learning**

*Özgür Şimşek, Max Planck Institute for Human Development, Germany  
George Konidaris, University of Massachusetts, Amherst, U.S.A.*

Although reinforcement learning methods have been effectively applied to a number of problems of practical importance, successful large-scale applications remain the exception rather than the norm. Problems with large state spaces still pose considerable challenges to existing algorithms.

Abstraction is the process of factoring out irrelevant details, in other words, of focusing only on the information that is relevant for a particular purpose. For a number of years, the research community has been exploring various forms of abstraction as potential mechanisms for scaling up reinforcement learning algorithms to large, complex problems. State abstraction approaches and temporal abstraction methods have become well established, while recent representation-discovery methods have shown a great deal of promise.

The goal of this workshop is to promote interaction between researchers that work on various forms of abstraction in reinforcement learning, to explore possible areas of synergy between existing approaches, and to open up discussion on novel techniques that can harness the existing strengths of different types of abstractions.

## **W9: Sparse Methods for Music Audio**

*Douglas Eck, University of Montreal, Canada  
Dan Ellis, Columbia University, U.S.A.  
Philippe Hamel, University of Montreal, Canada*

Sparse coding is gaining attention as alternative to traditional, orthogonal-basis approaches, able to find more interesting or more useful solutions to underconstrained, high-dimensional problems. Music audio provides an excellent candidate for sparse coding, being very high dimensional (e.g., over 80,000 values in one second of music from a CD), yet usefully described as the combination of a small number of separate signals – such as individual instruments – each subject to a large number of mutual constraints. This description can be applied at multiple levels, from the raw audio through to compositional structure.

There have been only a few publications on applying sparse techniques in music. The goal of the workshop is to bring together researchers with an interest in this topic, to focus, develop, and refine the various perspectives and approaches possible. We hope to raise the profile of these ideas, both to those already working with music audio, and to machine learning researchers who may be curious about working with music audio data.



---

**ICML 2009 Papers**

---

