Summary of ICML'04
International Conference on Machine Learning

Russ Greiner

ICML'04 Program CoChair

Co-ProgramChair: Dale Schuurmans

GeneralChair: Carla Brodley

ALBERTA INGENUITY CENTRE FOR MACHINE LEARNING

NSERC CRSNG

UNIVERSITY OF ALBERTA

Making IT happen

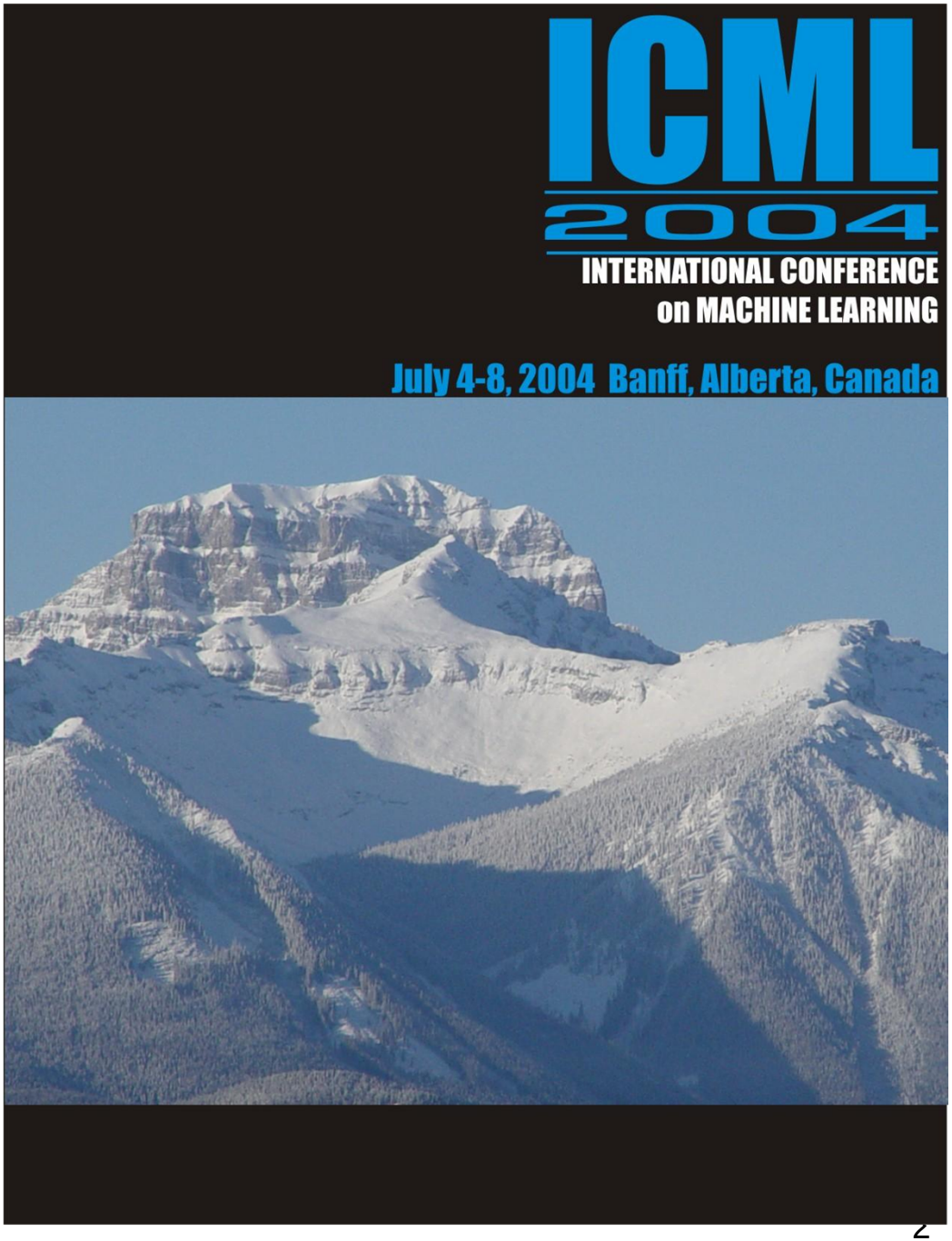
Department of Computing Science

In the beautiful Canadian Rockies...

Twenty-First International Conference on Machine Learning

Banff, Alberta

4 – 8 July 2004




2004
icml

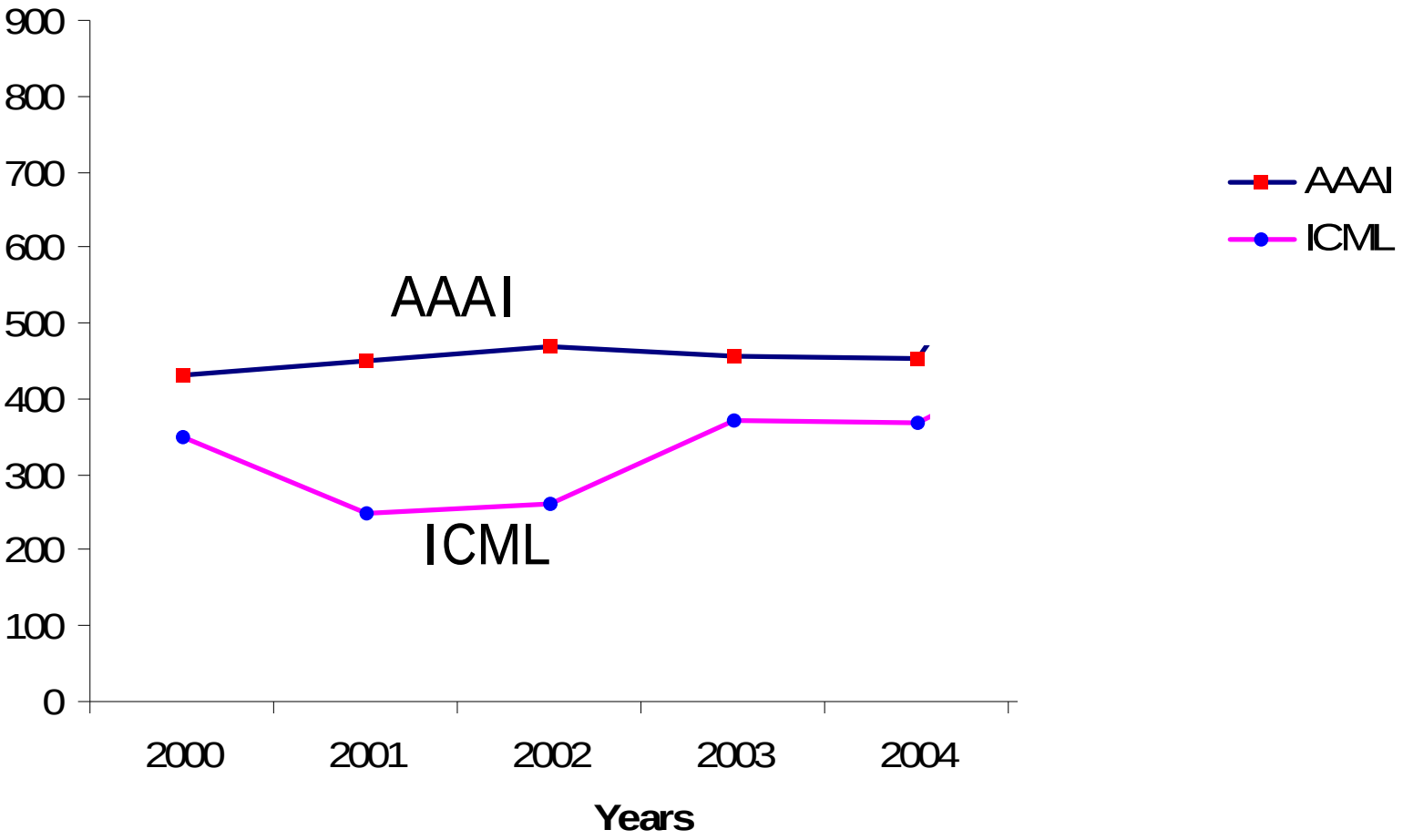
Outline

- Statistics
- High Points
- Comments

Submissions

- 368 submissions
- 118 acceptances
  - 32% acceptance rate

#Submissions to AAAI, ICML



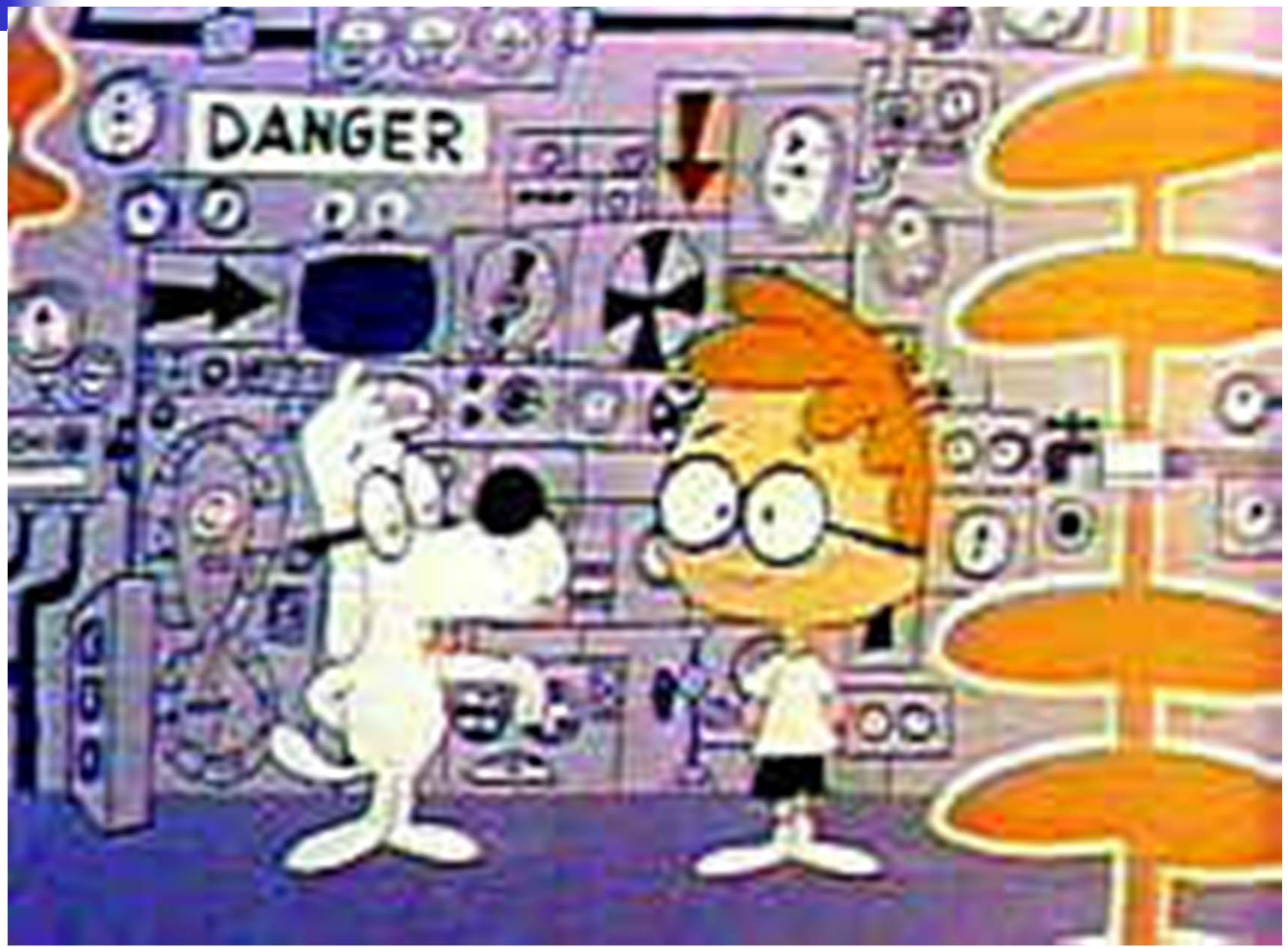
Top 10 Submitted Topics: 2004

Rank	Topic	# Submissions	# Accepts
1	Statistical learning	**49**	**31**
2	Empirical evaluation	**38**	**12**
3	Feature selection	**36**	**14**
4	Reinforcement learning	**36**	**15**
5	Semi-supervised learning	**29**	**6**
6	Kernels	**28**	**20**
7	On-line learning	**26**	**4**
8	Clustering	**25**	**12**
9	Bayesian learning	**24**	**14**
10	Support vector machines	**24**	**15**

Top 10 Submitted Topics: 2003

	Topic	Submitted	Rank (2004)
1.	Statistical learning	65	1
2.	Reinforcement learning	58	4
3.	SVM + kernel	49	≈ 6,10
4.	Applications	48	-
5.	Clustering	47	8
6.	Evaluation	46	2
7.	Learning w/(un)labeled data	40	5
8.	Feature selection	36	3
9.	Text classification	32	-
10	Markov models (HMMs, POMDPs, …)	25	≈9

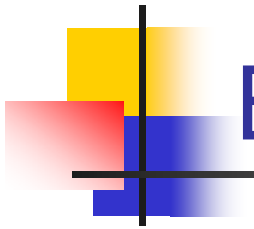
Way back... 1994 ...



- Naoki Abe, Hiroshi Mamitsuka: A New Method for Predicting Protein Secondary Structures Based on Stochastic Tree Grammars. 3-11
- David W. Aha, Stephane Lapointe, Charles X. Ling, Stan Matwin: Learning Recursive Relations with Randomly Selected Small Training Sets. 12-18
- Lars Asker: Improving Accuracy of Incorrect Domain Theories. 19-27
- Rich Caruana, Dayne Freitag: Greedy Attribute Selection. 28-36
- Mark Craven, Jude W. Shavlik: Using Sampling and Queries to Extract Rules from Trained Neural Networks. 37-45
- Michael de la Maza: The Generate, Test, and Explain Discovery System Architecture. 46-52
- Harris Drucker, Corinna Cortes, Lawrence D. Jackel, Yann LeCun, Vladimir Vapnik: Boosting and Other Machine Learning Algorithms. 53-61
- Tapio Elomaa: In Defense of C4.5: Notes Learning One-Level Decision Trees. 62-69
- Johannes Fürnkranz, Gerhard Widmer: Incremental Reduced Error Pruning. 70-77
- Melinda T. Gervasio, Gerald DeJong: An Incremental Learning Approach for Completable Planning. 78-86
- Yolanda Gil: Learning by Experimentation: Incremental Refinement of Incomplete Planning Domains. 87-95
- Attilio Giordana, Lorenza Saitta, Floriano Zini: Learning Disjunctive Concepts by Means of Genetic Algorithms. 96-1
- Matthias Heger: Consideration of Risk in Reinformance Learning. 105-111
- Chun-Nan Hsu, Craig A. Knoblock: Rule Introduction for Semantic Query Optimization. 11
- George H. John, Ron Kohavi, Karl Pfleger: Irrelevant Features and the Subs
- Jörg-Uwe Kietz, Marcus Lübbe: An Efficient Subsumption Algori
- Moshe Koppel, Alberto Maria Segre, Ronen Feldm
- David D. Lewis, Jason Catlett: Heter
- Michael L. Littman: Mark
- Sridhar Mah
- g R Learning and Q Learning. 164-172
- Rule-Bases. 173-180
- g Cross Validation Error. 190-198
- oduction System Rule-Bases. 199-207
- enetic Search to Refine Knowledge-based Neural Networks. 208-216
- pher J. Merz, Patrick M. Murphy, Kamal Ali, Timothy Hume, Clifford Brunk: Reducing Misclassification Costs. 217-225
- ald J. Williams: Incremental Multi-Step Q-Learning. 226-232
- J. Ross Quinlan: The Minimum Description Length Principle and Categorical Theories. 233-241
- John Rachlin, Simon Kasif, Steven Salzberg, David W. Aha: Towards a Better Understanding of Memory-based Reasoning Systems. 242-250
- Justinian P. Rosca, Dana H. Ballard: Hierarchical Self-Organization in Genetic programming. 251-258
- Cullen Schaffer: A Conservation Law for Generalization Performance. 259-265
- Robert E. Schapire, Manfred K. Warmuth: On the Worst-Case Analysis of Temporal-Difference Learning Algorithms. 266-274
- Michèle Sebag: A Constraint-based Induction Algorithm in FOL. 275-283
- Satinder P. Singh, Tommi Jaakkola, Michael I. Jordan: Learning Without State-Estimation in Partially Observable Markovian Decision Processes.
- David B. Skalak: Prototype and Feature Selection by Sampling and Random Mutation Hill Climbing Algorithms. 293-301
- Irina Tchoumatchenko, Jean-Gabriel Ganascia: A Baysian Framework to Integrate Symbolic and Neural Learning. 302-308
- Chen K. Tham, Richard W. Prager: A Modular Q-Learning Architecture for Manipulator Task Decomposition. 309-317
- Paul E. Utgoff: An Improved Algorithm for Incremental Induction of Decision Trees. 318-325
- Raúl E. Valdés-Pérez, Aurora Pérez: A Powerful Heuristic for the Discovery of Complex Patterned Behaviour. 326-334
- Sholom M. Weiss, Nitin Indurkhya: Small Sample Decision tree Pruning. 335-342
- John M. Zelle, Raymond J. Mooney, Joshua B. Konvisser: Combining Top-down and Bottom-up Techniques in Inductive Logic Programming. 343-351
- Jean-Daniel Zucker, Jean-Gabriel Ganascia: Selective Reformulation of Examples in Concept Learning. 352-360

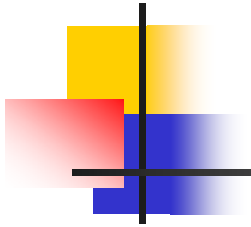
Invited Talks

- Michael I. Jordan: A Statistical Approach to Decision Tree Modeling. 363-370
- Stephen Muggleton: Bayesian Inductive Logic Programming. 371-379



Both Then (1994) and Now (2004)

- Classification
  - Supervised learning
  - Boosting
- Reinforcement Learning
  - MDP, POMDP, ...
- Applications
  - Bioinformatics
  - Planning
- ...



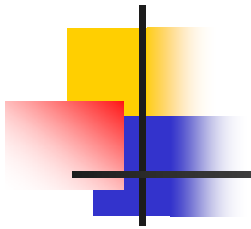
Then vs Now ...

1994:

- Dimensionality Reduction
  - Feature Selection

2004:

- Dimensionality Reduction
  - PCA, ...



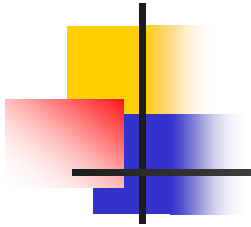
Then vs Now ...

1994:

- Probabilistic Models
  - Rare...
  - Bayesian nets

2004:

- Probabilistic Models
  - Everywhere!
  - Bayesian nets, Markov Random Fields, Conditional Random Fields,



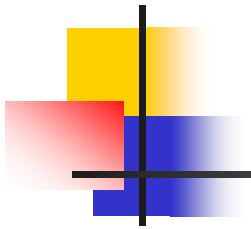
Then vs Now ...

1994:

- Symbolic learning
  - Decision Tree
  - Inductive Logic Programming

2004:

- Statistical Learning
  - 49/118 !
  - Lasso, EM, MCMC,
  - Robust,
  - ...



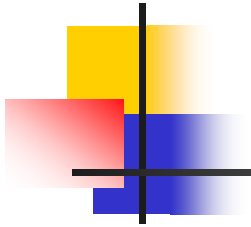
Then vs Now ...

1994:

- Concept Learning
- Neural Nets

2004:

- SVMs
- Kernel
- Numeric Methods
  - Linear Algebra
  - ...
- Clustering
  - 12 papers, vs 0



Then vs Now ...

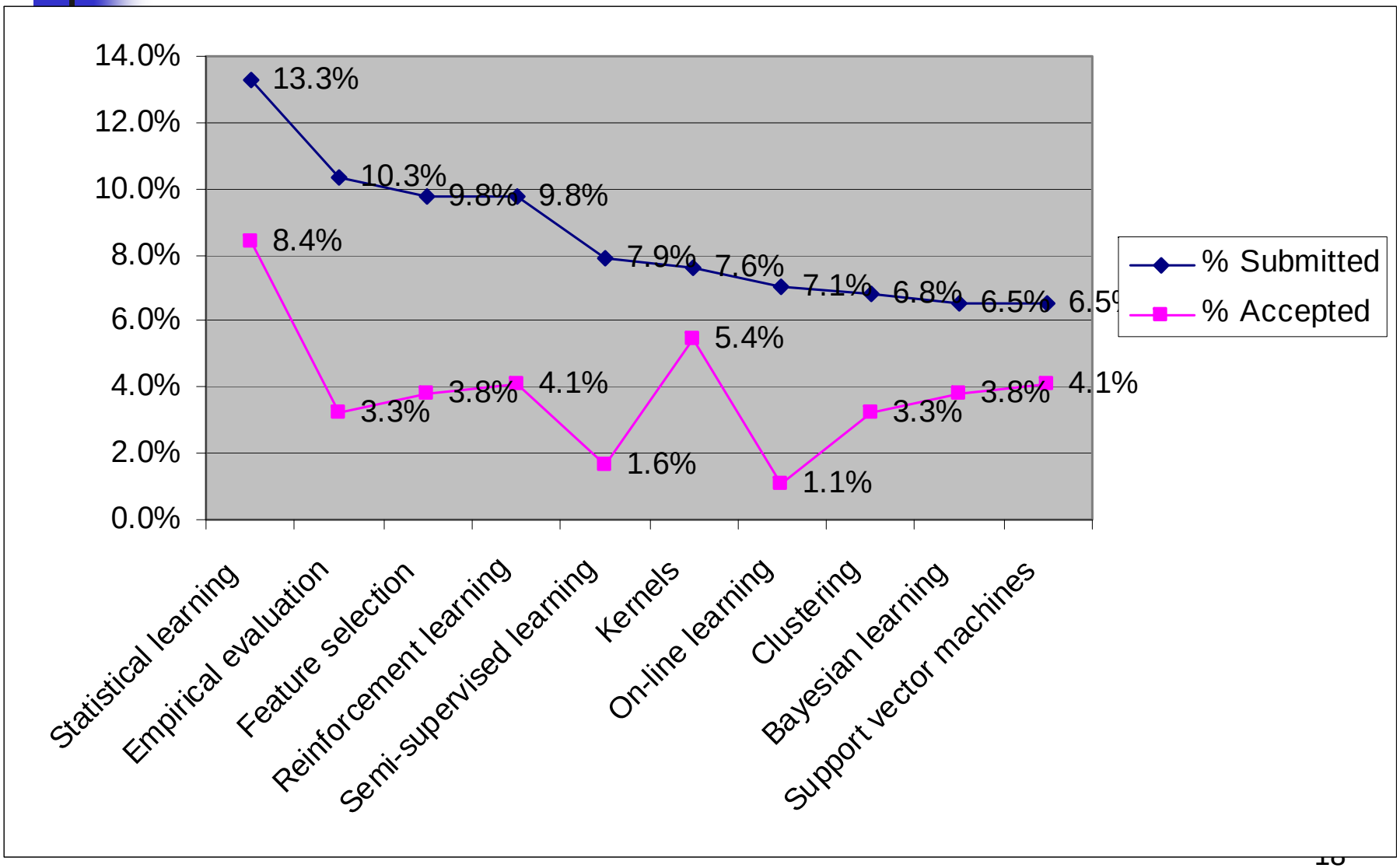
1994:

- Validation:
  - Better than C4.5 on some UCI
- Some Real Applications

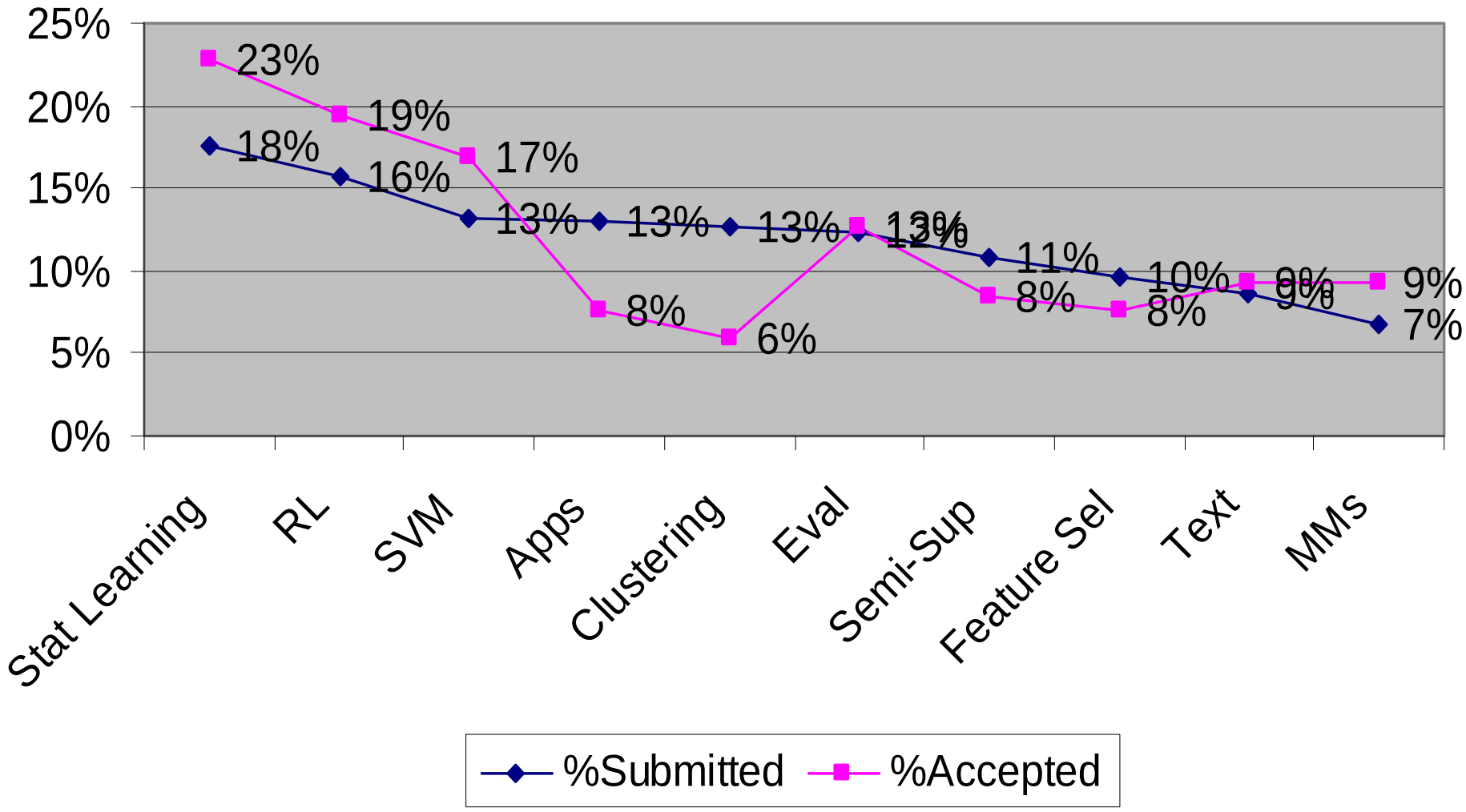
2004:

- Validation:
  - Better than SVM on MANY UCI
- Most involve Real applications

Top 10 Submitted Topics 2004



Top 10 Submitted Topics 2003



How to get a paper accepted in ICML 2004???

w	P(accept|w)

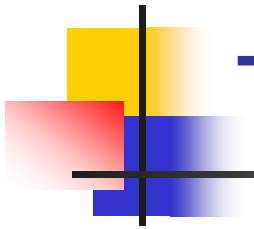
Attendance

- ICML'04 Conference: 397

  - + 30 for Workshop/Tutorial
- Collocated with
  - UAI 2004
  - 

Workshops

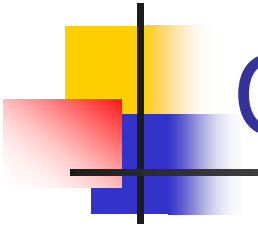
- Statistical Relational Learning
  - 84
- Relational Reinforcement Learning
  - 28
- Predictive Representations of World Knowledge
  - 23
- Physiological Data Modeling
  - 9



Tutorials

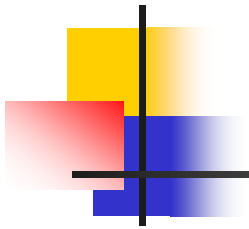
- *Bayesian Methods for Machine Learning*: 144
- *Kernels for Structured Data* : 127

- *Data Structures for Fast Statistics*: 90
- *Game-theoretic Learning* : 81
- *Spectral Clustering* : 61
- *Probabilistic Logic Learning*: 60
- *Junk E-mail Filtering*: 50
- *The Many Faces of ROC Analysis* : 42



Outline

- Statistics
- High Points
  - Invited Speakers
  - Best Papers
  - Kernel Day
- Comments



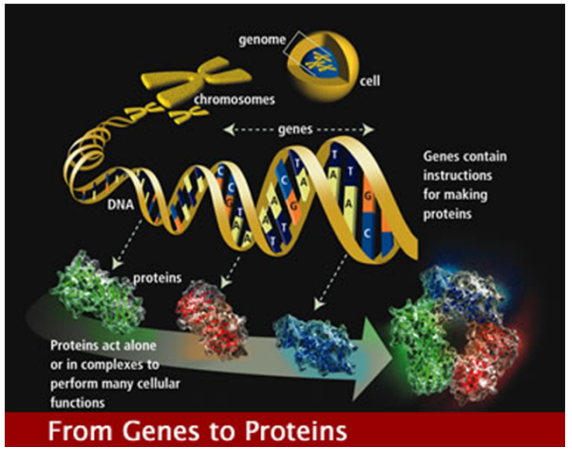
Invited Presentations

- Identify areas ADJACENT to ML
  - *BioInformatics*
  - *Vision*
  - *Computational Finance*
- Invite leading exponent of that area

Invited Talk: Bioinformatics

- **Gene Myers**
  - Celera Genomics
  - *BLAST*
    - ACM Kanellakis Prize in 2001
  - Member: National Academy of Engineering

- *Whole Genome Sequencing, Comparative Genomics, and Systems Biology*

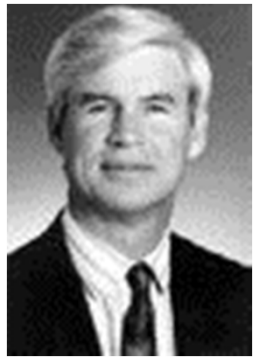


Invited Talk: Vision

- **Michael Black**
  - leading researcher in computer vision & AI
  - *Learning to See People*

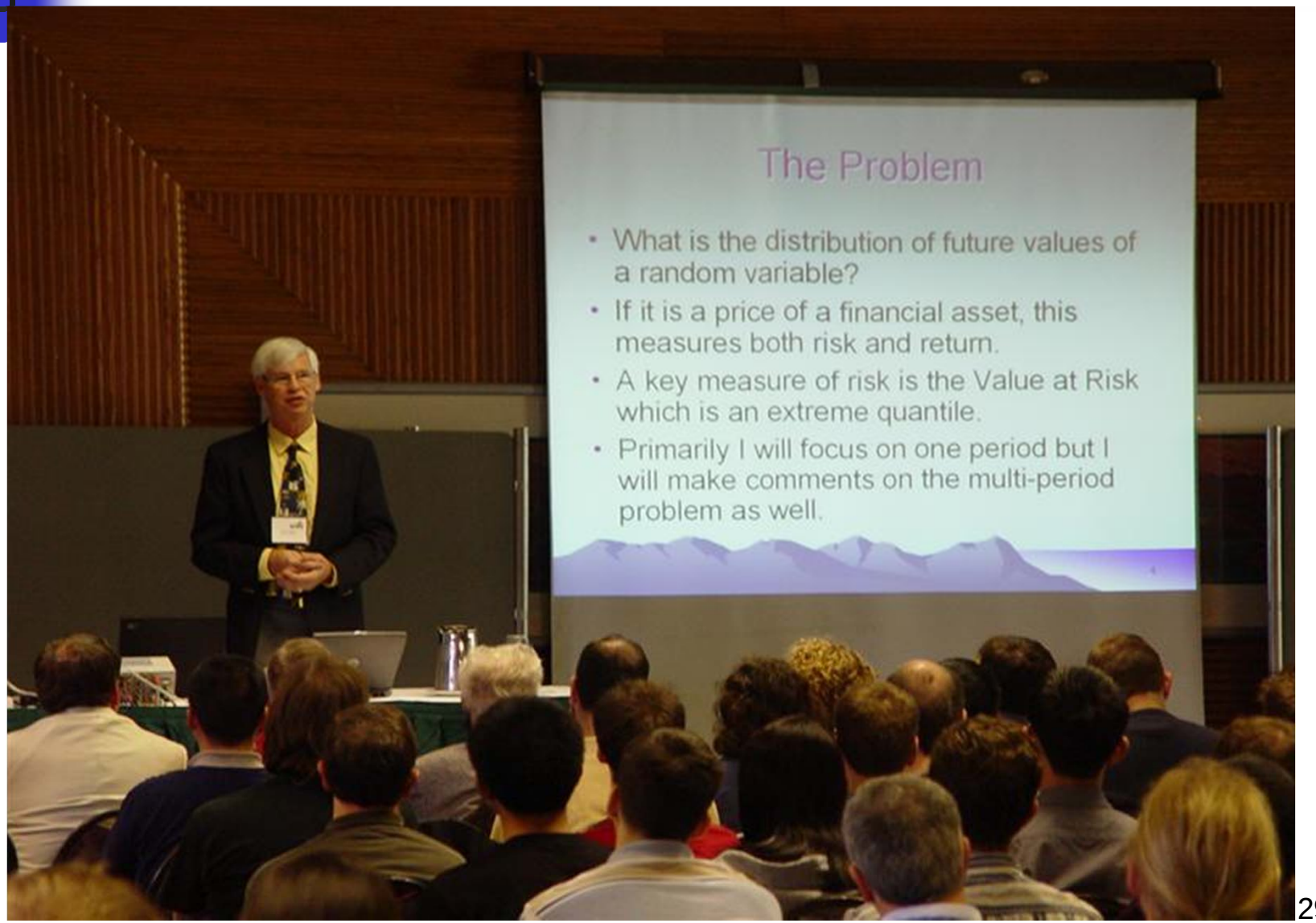
Invited Presentations: Computational Finance

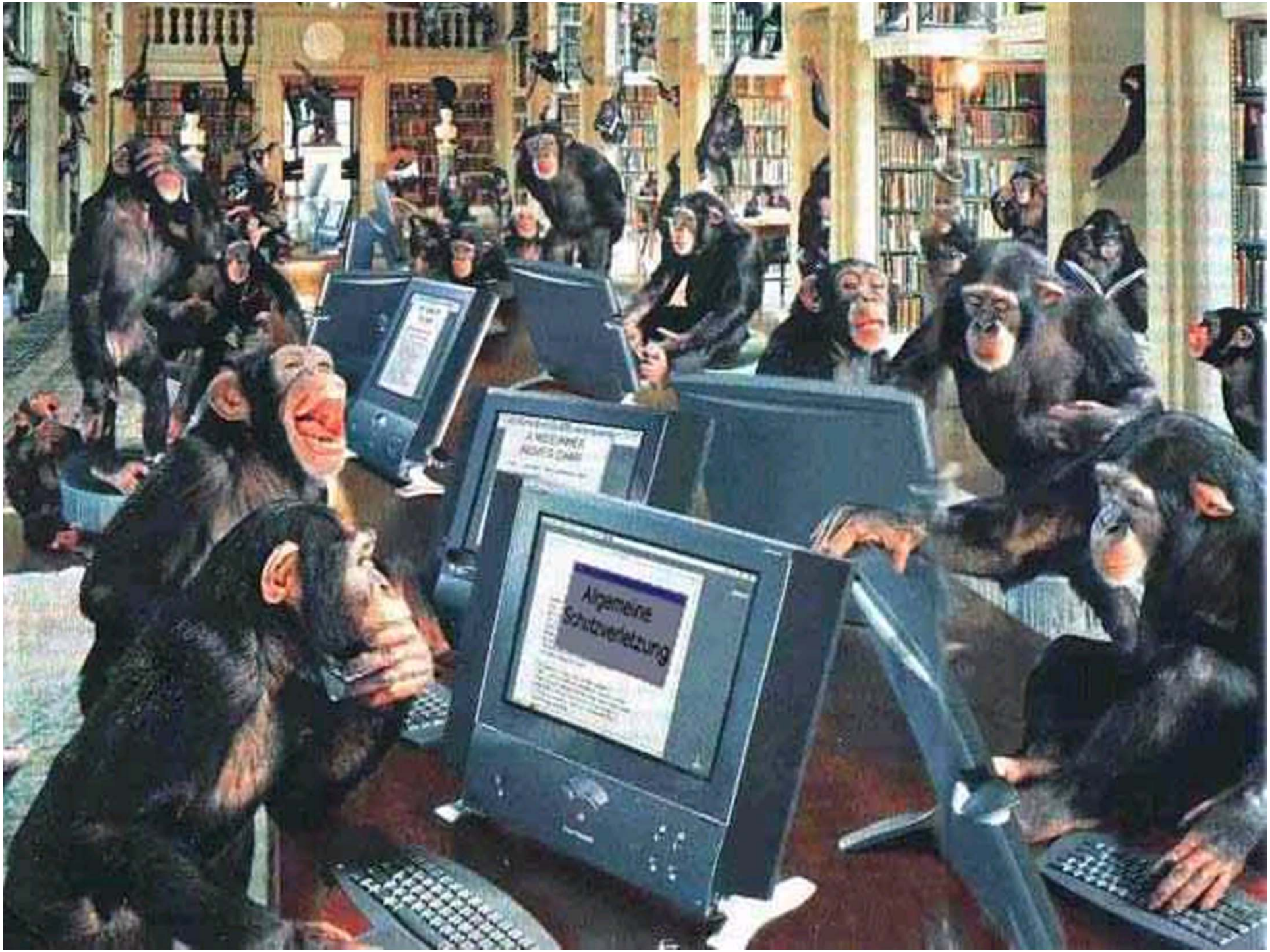
- **#1:** 2003 Nobel Prize Winner, Economics
  - for methods of analyzing economic time series with time-varying volatility (ARCH)
  - **Robert Engle**



- **#2: …**
- **#3: …**
- **#4: …**
- **…**

Models of Dynamic Uncertainty in Univariate and Multivariate Systems




Allgemeine
Schutzverletzung

ARCH Model...

Outstanding Student Papers

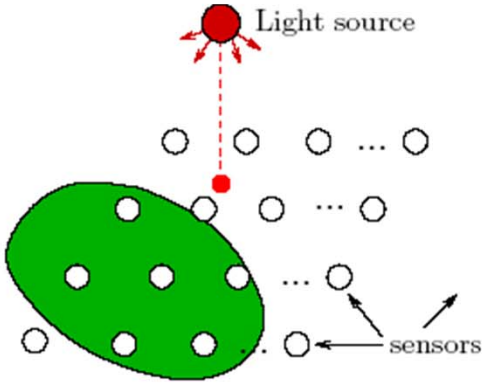
- **Generalized Low Rank Approximations of Matrices**
  - *Jieping Ye*
- **Decentralized Detection and Classification using Kernel Method**
  - *XuanLong Nguyen,* Martin Wainwright, Michael Jordan
- **Learning a Kernel Matrix for Nonlinear Dimensionality Reduction**
  - *Kilian Weinberger, Fei Sha,* Lawrence Saul

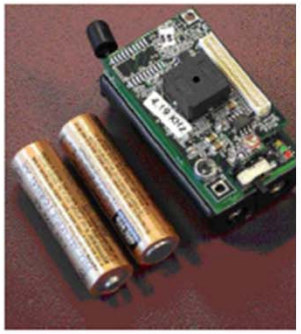
Generalized Low Rank Approximations of Matrices

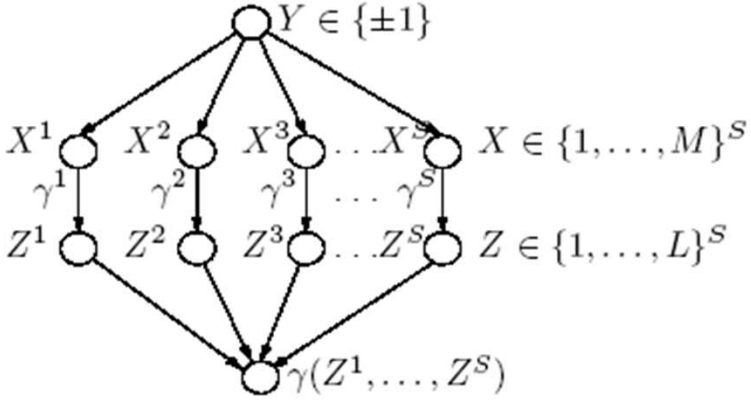
- *Jieping Ye*

Decentralized Detection and Classification using Kernel Method

- *XuanLong Nguyen*
- Martin Wainwright
- Michael Jordan





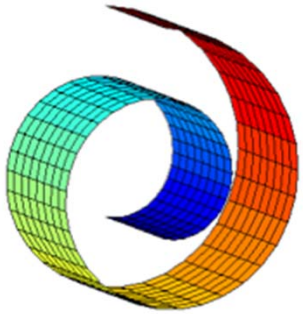


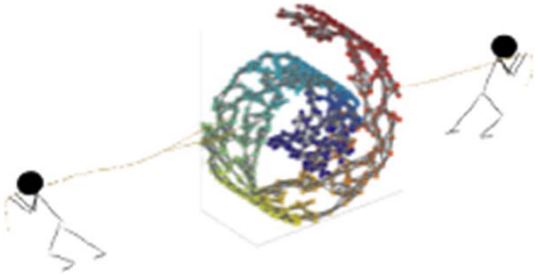
- Problem: Given training data $(x_i, y_i)_{i=1..n}$
- Find: decision rules $(\gamma^1, \ldots, \gamma^s; \gamma)$ that minimize misclassification probability: $P(Y \neq \gamma(Z^1, \ldots, Z^s))$

Learning a Kernel Matrix for Nonlinear Dimensionality Reduction

- *Kilian Weinberger*

- *Fei Sha*
- Lawrence Saul





Honorable Mention for Outstanding Paper Award

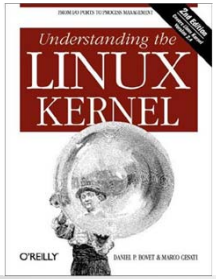
- **Multiple Kernel Learning, Conic Duality, and the SMO Algorithm**
  - *Francis Bach, Gert Lanckriet, Michael Jordan*
- **Efficient Hierarchical MCMC for Policy Search**
  - *Malcolm Strens*
- **Authorship Verification as a One-Class Classification Problem**
  - *Moshe Koppel, Jonathan Schler*

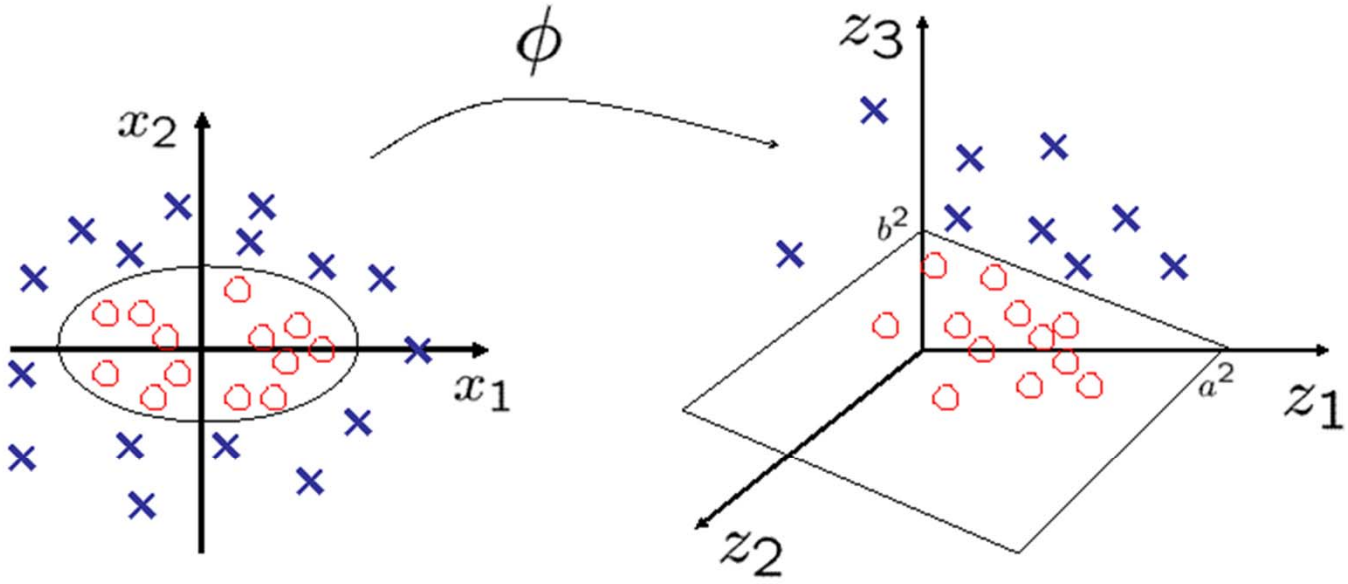
Linear Separators

- Linear Separator of $\{ \langle x_i, \pm \rangle \}$
  - $f(x) = w \bullet x = \Sigma_i \alpha_i x_i \bullet x$
  - ... so w is sum of training data x_i
- If re-represent data using
  - $\Phi: \Re^r \rightarrow \Re^k$

  then
  - $f(x) = \Sigma_i \alpha_i \Phi(x_i) \bullet \Phi(x)$
- Sometimes...
  - use $K(x,z) = \Phi(x) \bullet \Phi(z)$
  - If *n* datapoints, only $n \times n$, even if $k=\infty$

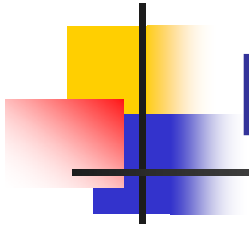
Kernel Trick





$$\phi : (x_1, x_2) \longrightarrow (x_1^2, \sqrt{2}x_1x_2, x_2^2)$$

$$\left(\frac{x_1}{a}\right)^2 + \left(\frac{x_2}{b}\right)^2 = 1 \longrightarrow \frac{z_1}{a^2} + \frac{z_3}{b^2} = 1$$



Kernel Day

COLT Session: Kernels

- Bayesian Networks and Inner Product Spaces
- Inequality for Nearly Log-concave Distributions with Applications to Learning
- Bayes and Tukey Meet at the Center Point
- Sparseness vs Estimating Conditional Probabilities: Some Asymptotic Results
- A Statistical Mechanics Analysis of Gram Matrix Eigenvalue Spectra
- Statistical Properties of Kernel Principal Component Analysis
- Kernelizing Sorting, Permutation and Alignment for Minimum Volume PCA
- Regularization and Semisupervised Learning on Large Graphs

ICML Session: Kernels

- Support Vector Machine Learning for Interdependent and Structured Output Spaces
- Probabilistic Tangent Subspace: A Unified View
- Bayesian Inference for Transductive Learning of Kernel Matrix Using the Tanner-Wong Data Augmentation Algorithm
- Kernel-Based Discriminative Learning Algorithms for Labeling Sequences, Trees and Graphs
- A Kernel View of the Dimensionality Reduction of Manifolds
- Multiple Kernel Learning, Conic Duality, and the SMO Algorithm
- Learning with Non Positive Kernels
- Extensions of Marginalized Graph Kernels

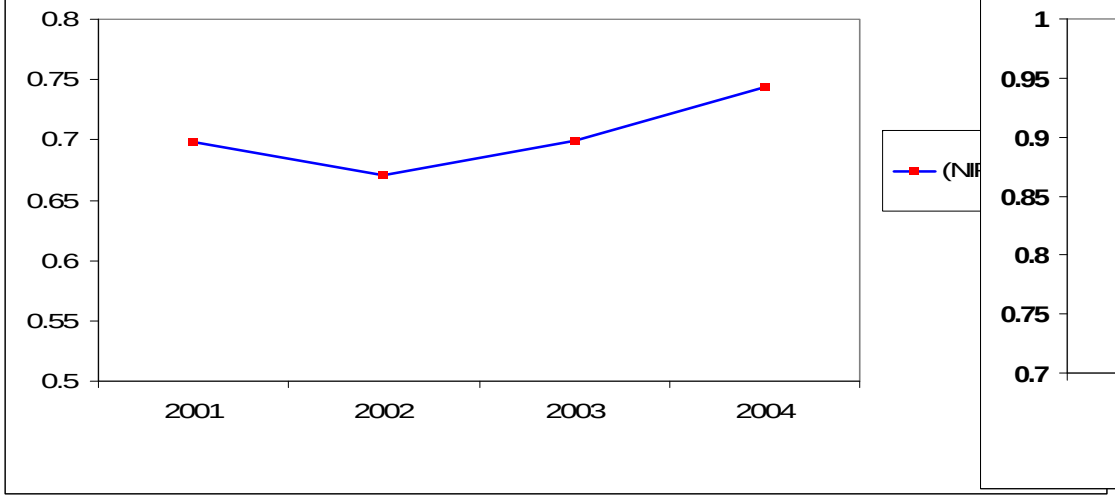
Distribution of Papers

- Lots of numerical methods
  - Linear algebra
  - Dimensionality reduction techniques
  - ...
- Better know the “kernel trick” !
- Is ICML becoming more like NIPS?
- Similarity (cosine) of
  - counts of all words in NIPS-xx
  - counts of all words in ICML-xx

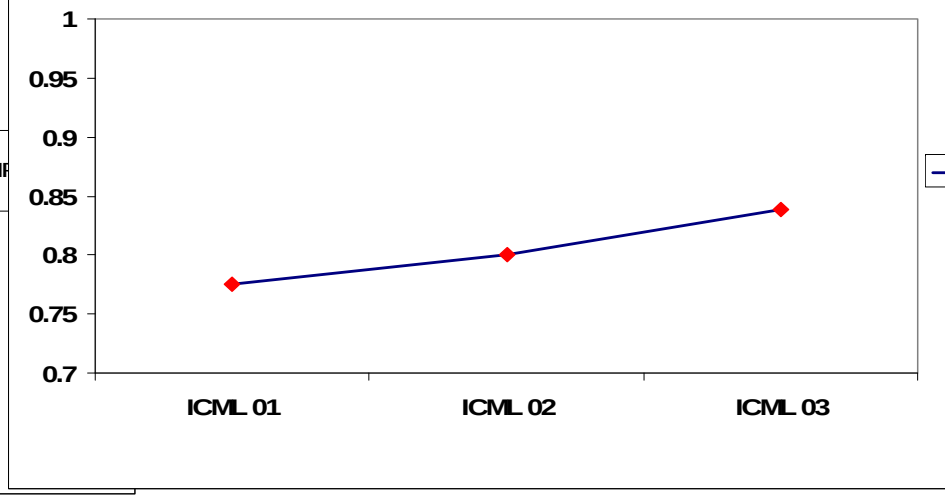
Neural Information Processing Systems Conference

ICML'04 ~ NIPS ??

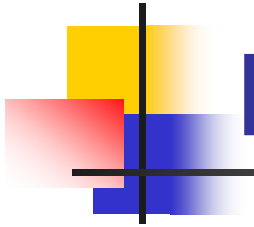
NIPS-xx vs ICML-xx



ICML'04 vs ICML-xx



- Yes,
  NIPS-xx titles are getting more similar to ICML-xx each year...
- But ...
  ICML'04 is still very similar to earlier ICML-xx's



Distribution of Papers, II

- Hoped for MANY applications...
  - Great ACs in Vision, Games, BioInfo, NaturalLanguage
  - Some great application papers
  - ... still not that many submissions in these areas

28 Area Chairs

+ 200 Program Committee Members

- Yoshua Bengio

NLU/Text Eric Brill

BioInfo Mark Craven

- Charles Elkan
- Peter Flach

Vision David Fleet

- Johannes Fürnkranz
- Lise Getoor
- Zoubin Ghahramani
- Sally Goldman
- Geoff Gordon
- Thomas Hofmann
- David Jensen

NLU/Text Mark Johnson

BioInfo Simon Kasif

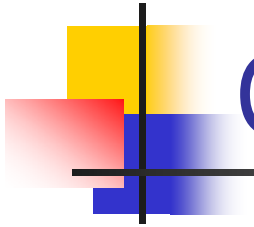
BioInfo Eamonn Keogh

- John Langford
- Michael Littman
- Risto Miikkulainen
- Andrew Moore
- Andrew Ng
- Sam Roweis

NLU/Text Mehran Sahami

Games Jonathan Schaeffer

- Bernhard Schoelkopf
- Alex Smola
- Rich Sutton
- Chris Williams



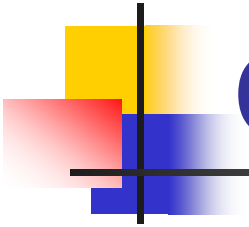
Comments on ICML'04

- Double-Blind reviewing
  - Seemed to work... stay tuned
- Many Conditional Accepts
  - Appropriate, as Archival! (like journal)
- *General* chair!
  - in addition to Program Chairs
  - Thanks, Carla!

Field of Machine Learning

- Conference went VERY VERY well.
- Because the field is ...Alive and Well!
- Lots of new Ideas and Insights
- Many new challenges!
- Interacting with + Contributing to...
  - NLU, BioInfo, Vision, Robotics, Web, ...
- *If it deals with real data*, *it's machine learning.*

- ... most of AI now deals with real data!



Outline

- Statistics
- High Points
- Comments

Sponsors

- Local (Alberta)

Making IT happen Computing Science

Alberta INGENUITY Centre for Machine Learning

PIMS

INFORMATICS iCORE CIRCLE OF RESEARCH EXCELLENCE

- Gov't (US)

DARPA

NSF

- Industry

Google

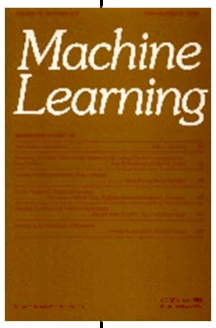
YAHOO! Research Labs

SIEMENS

Microsoft

Sun microsystems

We make the net work.



Thanks to ...

- All Award Winners, Invited Speakers
  - for giving great talks then
  - sending me their slides!
- ICML'04 Committee

  Carla Brodley, Johannes Furnkranz, Rob Holte,
  Michael DeMarco, Kiri Wagstaff, Jennifer Dy, Stephen Scott,
  David Jensen, David Woloschuk + web team, C-H Lee
- Area chairs, reviewers, authors, attendees, ...
- My funders
  - Alberta Ingenuity Centre for Machine Learning
  - NSERC
  - UofAlberta Dept of Computing Science

ALBERTA INGENUITY CENTRE FOR MACHINE LEARNING

NSERC CRSNG

Making IT happen

Department of Computing Science

Questions ?

- ICML 2004: 4-8/July Banff, Alberta
  - http://www.aicml.cs.ualberta.ca/_banff04/icml/
- ICML 2005: 7-11/Aug Bonn, Germany
  - http://icml2005.kdnet.org/icml.php
- ICML 2006: 24-30/June Pittsburgh, PA
  - http://icml2006.org/

  - Pictures...

Gene Myers


Whole Genome Sequencing, Comparative Genomics, & Systems Biology
Gene Myers
University of California
Berkeley

Computation of marginalized kernels

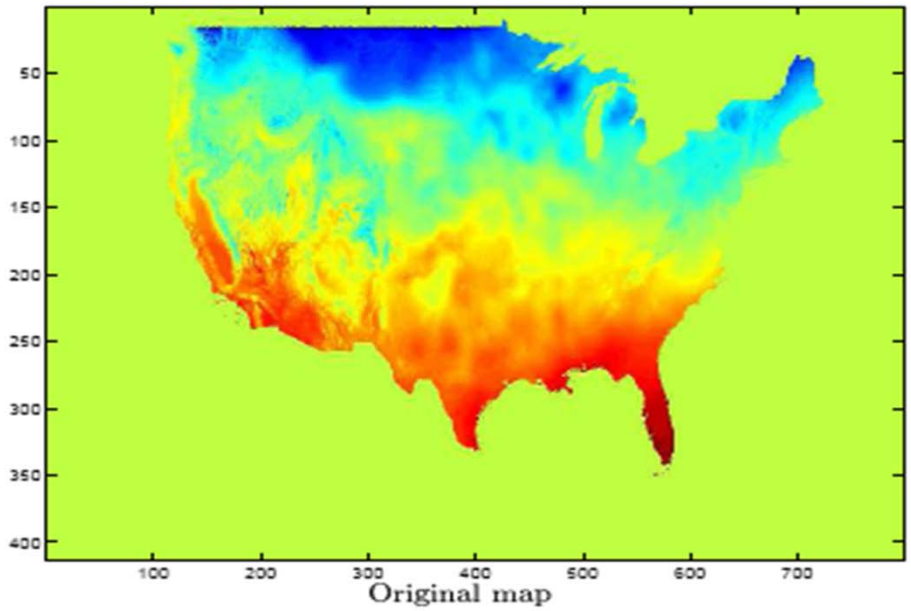
(Tsuda et al, 2002; Jaakkola & Haussler, 1999)

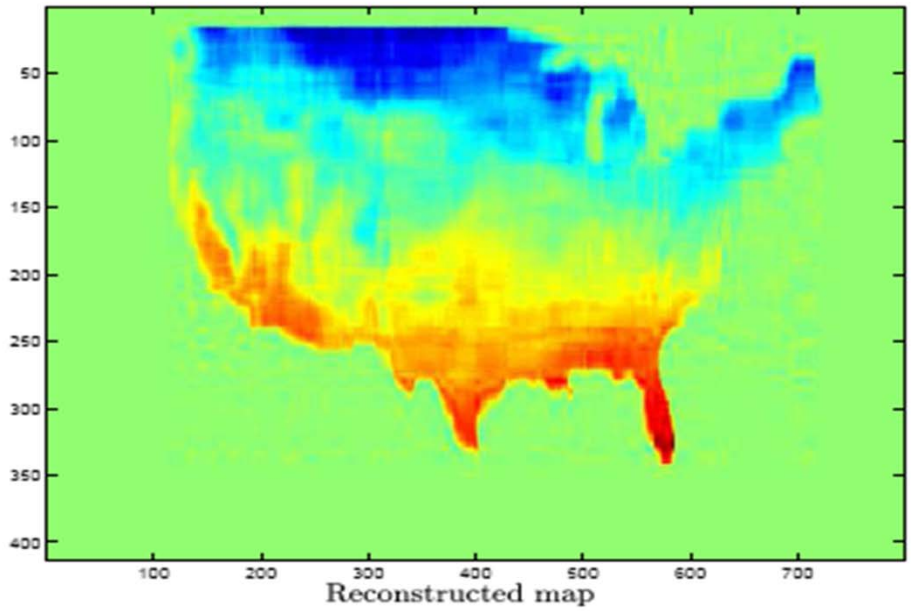
- Marginalized kernel $K_Q(x, x')$ is defined as:

$$K_Q(x, x') := \sum_{z,z'} \underbrace{Q(z|x)Q(z'|x')}_{\text{Factorized distributions}} \underbrace{K_z(z, z')}_{\text{Base kernel}} ,$$

- If $K_z(z, z')$ is decomposed into smaller components of z and z', then $K_Q(x, x')$ can be computed efficiently (in polynomial-time).

Nguyen/Wainwright/Jordan





- Original size: $6570 \times 413 \times 798 \times 8 = 17.3$ GB.
- Reduced size: $(6570 \times 25 \times 50 + 413 \times 25 + 798 \times 50) \times 8 = 66.2$ MB.
- Compression ratio = 17.3 GB/66.2 MB $= 262$.
- SVD is not applicable for this dataset.
- GLRAM scales to large datasets.

GLRAM: The main results

- **Theorem 2** *Let L, R, and $\{M_i\}_{i=1}^n$ minimize $\sum_{i=1}^n ||A_i - LM_iR^T||_F^2$ subject to $L^TL = I_{\ell_1}$ and $R^TR = I_{\ell_2}$. Then for every i, $M_i = L^TA_iR$.*
  - Expand $\sum_{i=1}^n ||A_i - LM_iR^T||_F^2$ as

$$\begin{aligned}\sum_{i=1}^n ||A_i - LM_iR^T||_F^2 &= \sum_{i=1}^n \text{trace}\left((A_i - LM_iR^T)(A_i - LM_iR^T)^T\right) \\ &= \sum_{i=1}^n \text{trace}(A_iA_i^T) + \sum_{i=1}^n \text{trace}(M_iM_i^T) - 2\sum_{i=1}^n \text{trace}(LM_iR^TA_i^T).\end{aligned}$$

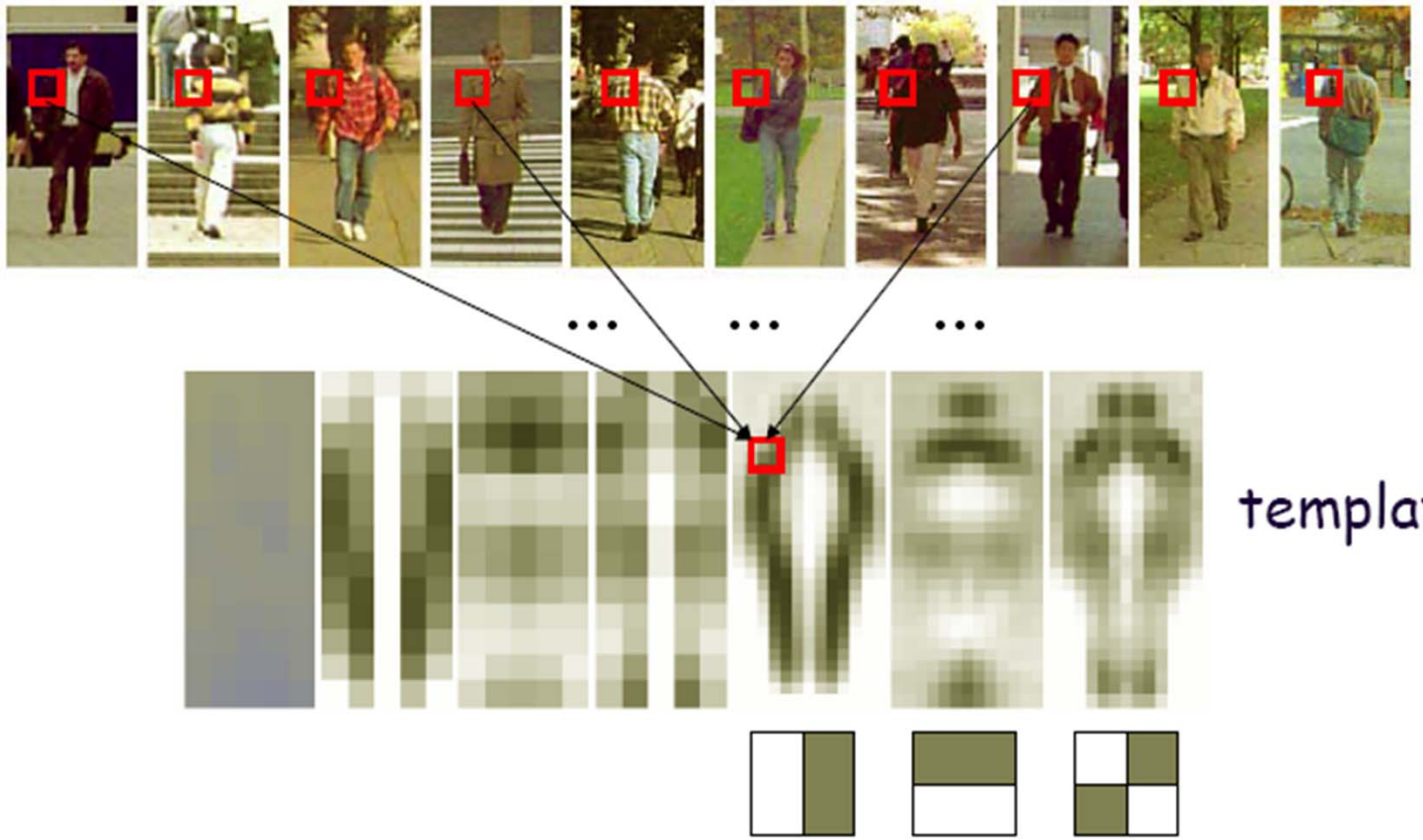
  - Take the derivative with respect to M_i and get $2M_i - 2L^TA_iR = 0 \Longrightarrow M_i = L^TA_iR$.

- **Theorem 3** *Let L and R be defined as in Theorem 2. Then L and R maximize $\sum_{i=1}^n ||L^TA_iR||_F^2$.*
  - Plugging $M_i = L^TA_iR$ into $\sum_{i=1}^n ||A_i - LM_iR^T||_F^2$, we obtain

$$\sum_{i=1}^n ||A_i - LM_iR^T||_F^2 = \sum_{i=1}^n ||A_i||_F^2 - \sum_{i=1}^n ||L^TA_iR||_F^2.$$

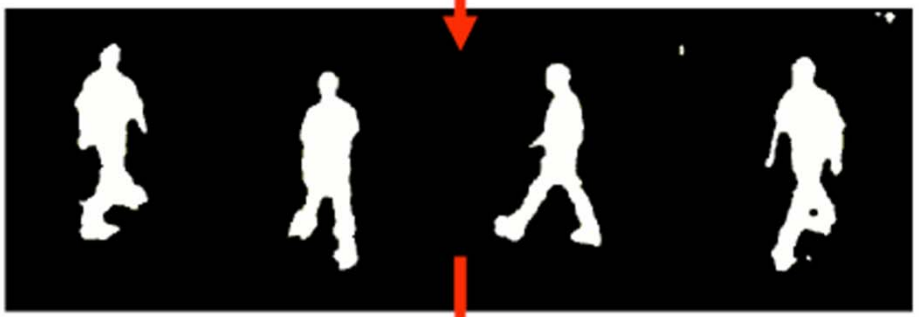
  - The minimization of $\sum_{i=1}^n ||A_i - LM_iR^T||_F^2$ is equivalent to the maximization of $\sum_{i=1}^n ||L^TA_iR||_F^2$.

Support Vector Machines



"Pedestrian detection using wavelet templates," Oren *et al* CVPR'97.

What about 3D Pose?



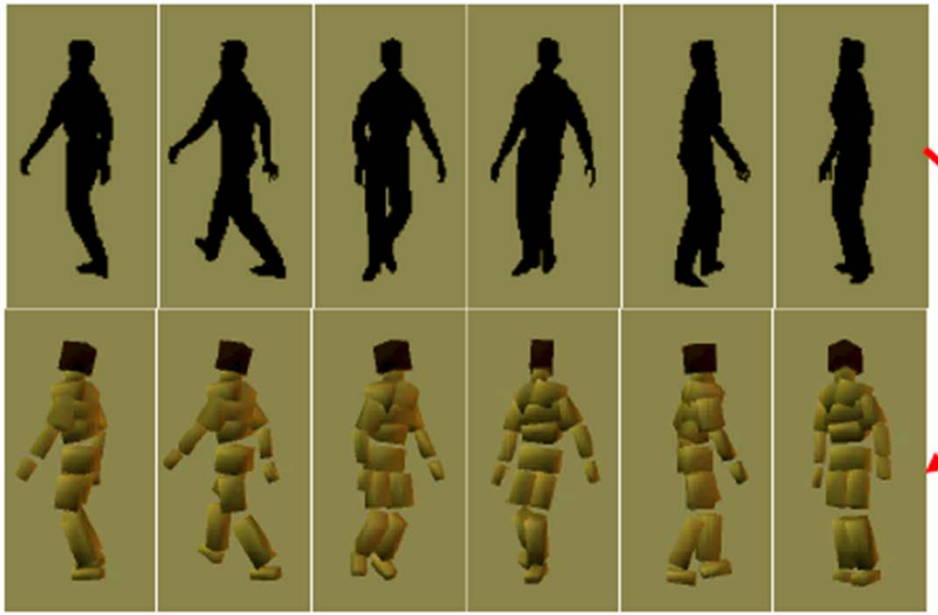
Contour Points / Shape Model

Learned mapping

30+ dimensions

K. Grauman, G. Shakhnarovich, T. Darrell, ICCV'03

Single View to 3D Pose



Given synthetic training data, learn the mapping from silhouette contours to 3D pose.

"Gaussian kernel RVM", Agarwal and Triggs CVPR04

"Fast Pose Estimation with Parameter Sensitive Hashing", Shakhnarovich, G., Viola, P., & Darrell, T. CVPR'03.

Problems

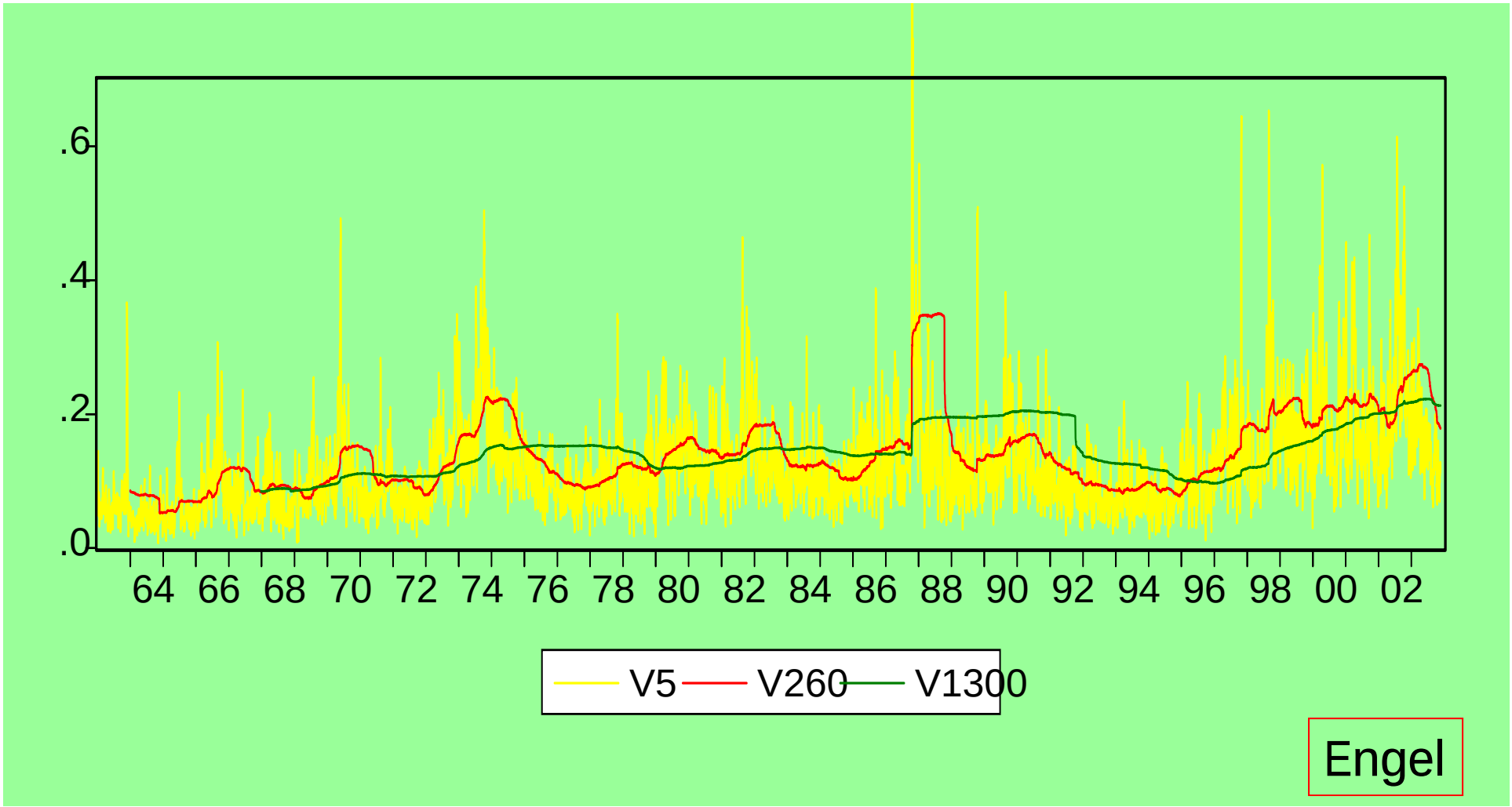
Accidental alignment

Motion blur.

(nothing to match)

ROLLING WINDOW VOLATILITIES

NUMBER OF DAYS=5, 260, 1300



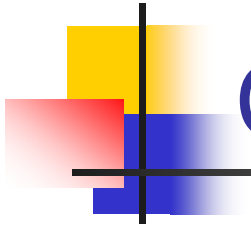

2004
icml

Invited Presentations

- **Michael Black**
  - leading researcher in computer vision & AI
  - *Learning to See People*

- **Gene Myers**
  - Celera; National Academy of Engineering
  - *Whole Genome Sequencing, Comparative Genomics, and Systems Biology*

- **Robert Engle**
  - 2003 Nobel Prize Winner, Economics
  - *Models of Dynamic Uncertainty in Univariate and Multivariate Systems*



Organizing Committee

- Conference Chair: *Carla Brodley*
- Workshops+ Tutorials Chair: *Johannes Fürnkranz*
- Local Arrangements Chairs: Rob Holte (Michael DeMarco)
- Registration Chair: Kiri Wagstaff
- Proceedings Chair: *Jennifer Dy*
- Treasurer: *Stephen Scott*
- Government Funding Chair: *David Jensen*
- Volunteer Coordinator: *Kiri Wagstaff*
- Webmasters: *David Woloschuk, Bryce Larson, ...*

Advisory committee

- Andrea Danyluk
- Tom Dietterich
- Tom Fawcett
- Achim Hoffmann
- Leslie Pack Kaelbling
- Nina Mishra
- Foster Provost
- Claude Sammut

Governing Bodies

- The International Machine Learning Society
- Kernel Machines

28 Area Chairs – International


Yoshua Bengio

- Eric Brill
- Mark Craven
- Charles Elkan


Peter Flach


David Fleet


Johannes Fürnkranz

- Lise Getoor


Zoubin Ghahramani

- Sally Goldman
- Geoff Gordon
- Thomas Hofmann
- David Jensen
- Mark Johnson
- Simon Kasif
- Eamonn Keogh
- John Langford
- Michael Littman
- Risto Miikkulainen
- Andrew Moore
- Andrew Ng


Sam Roweis

- Mehran Sahami


Jonathan Schaeffer


Bernhard Schoelkopf

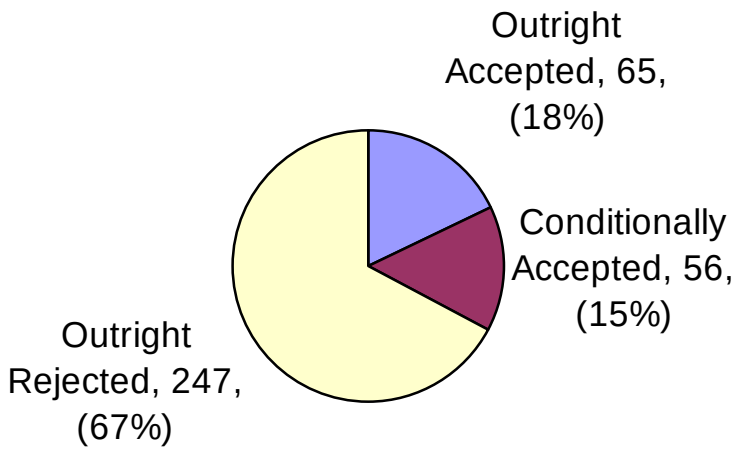

Alex Smola


Rich Sutton

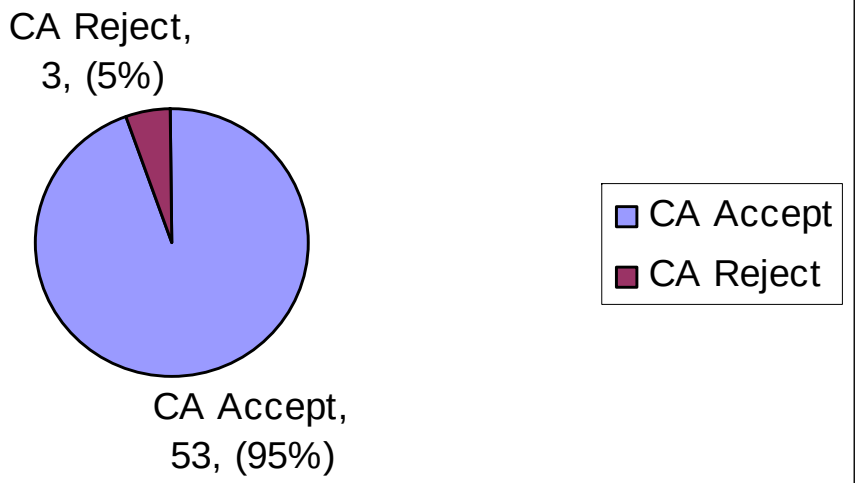

Chris Williams

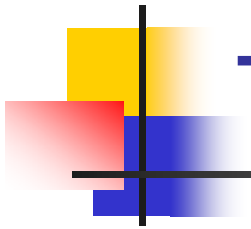
Acceptance Rates

368 submissions



56 CA's





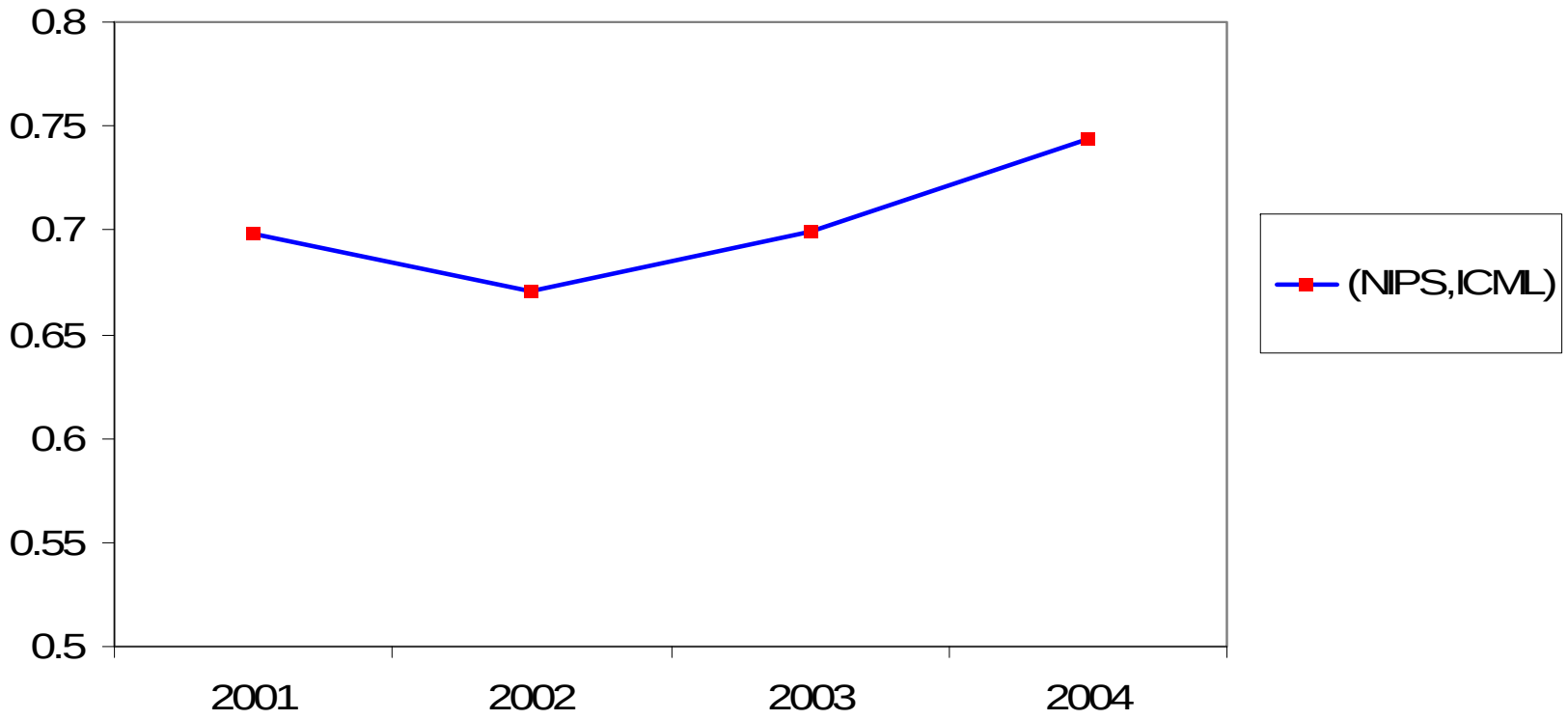
Thanks to...

- Organizing Committee :

  Carla Brodley, Johannes Furnkranz, Rob Holte, Michael DeMarco, Kiri Wagstaff, Jennifer Dy, Stephen Scott, David Jensen, David Woloschuk + web team

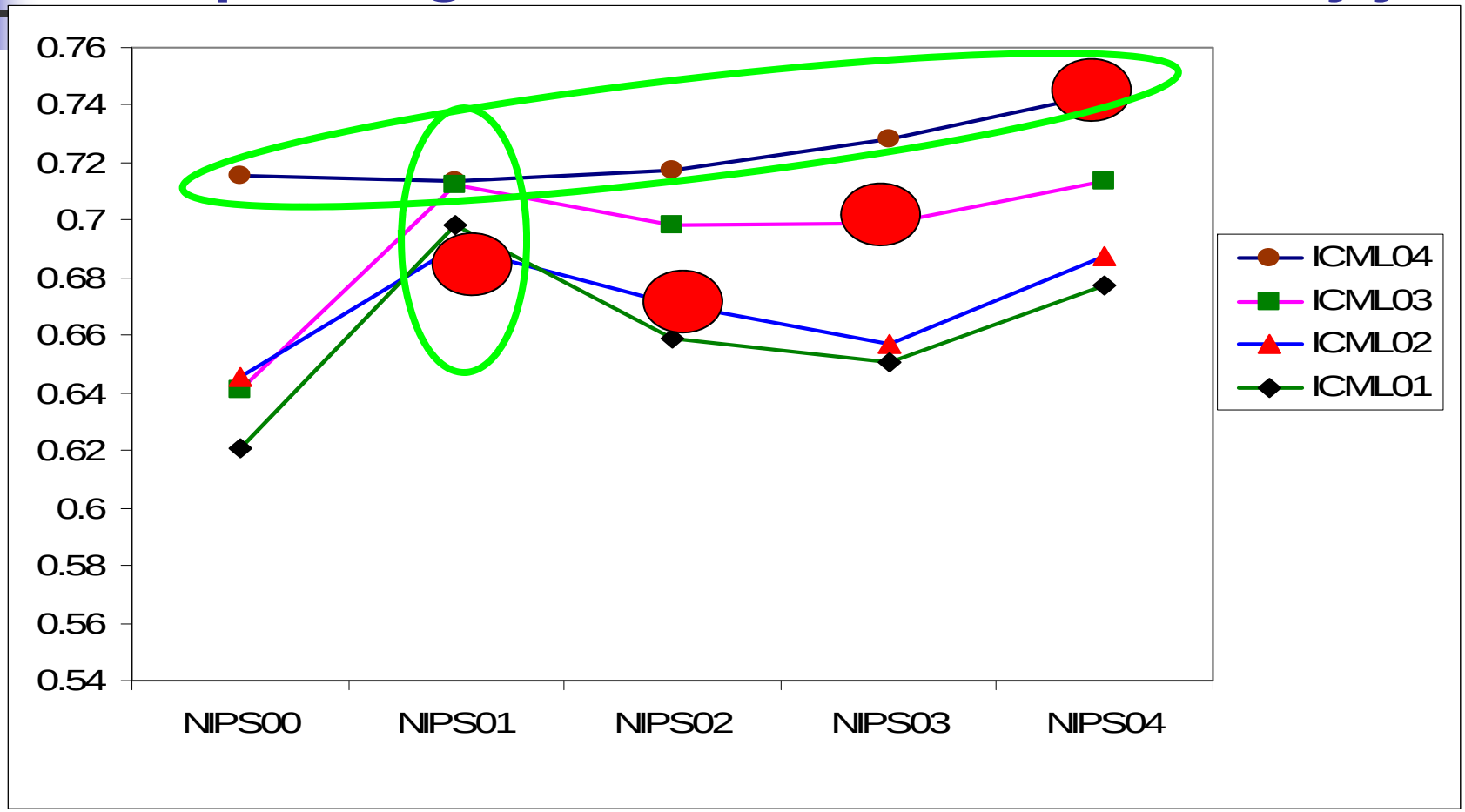
- Workshop/Tutorial Organizers
- Invited Speakers
- Area Chairs + Reviewers
- Authors
- Registrants

ICML-200x vs NIPS-200x



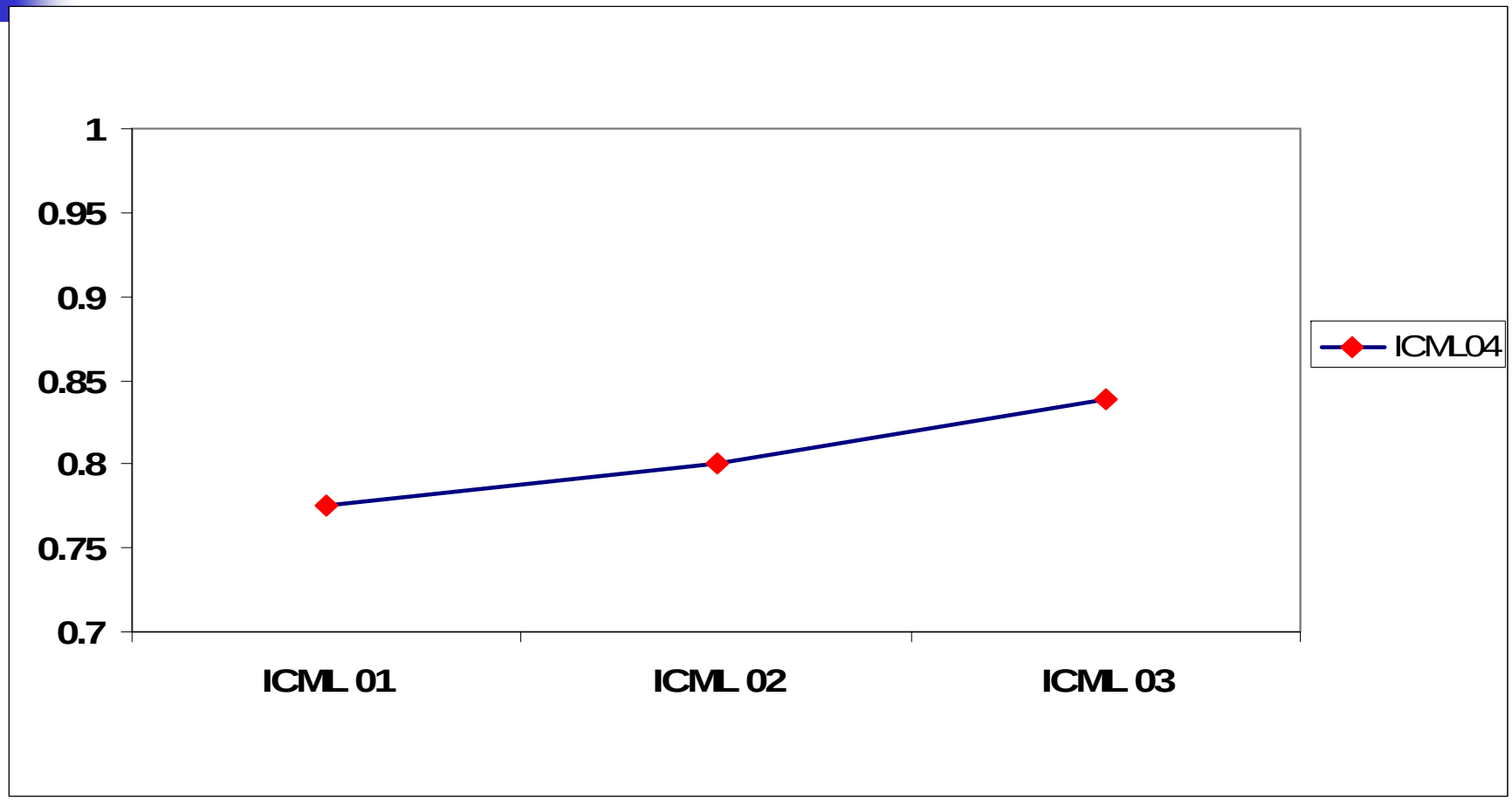
- Yes,
  NIPS-xx titles are getting more similar to ICML-xx each year...

Comparing ICML-xx with NIPS-yy



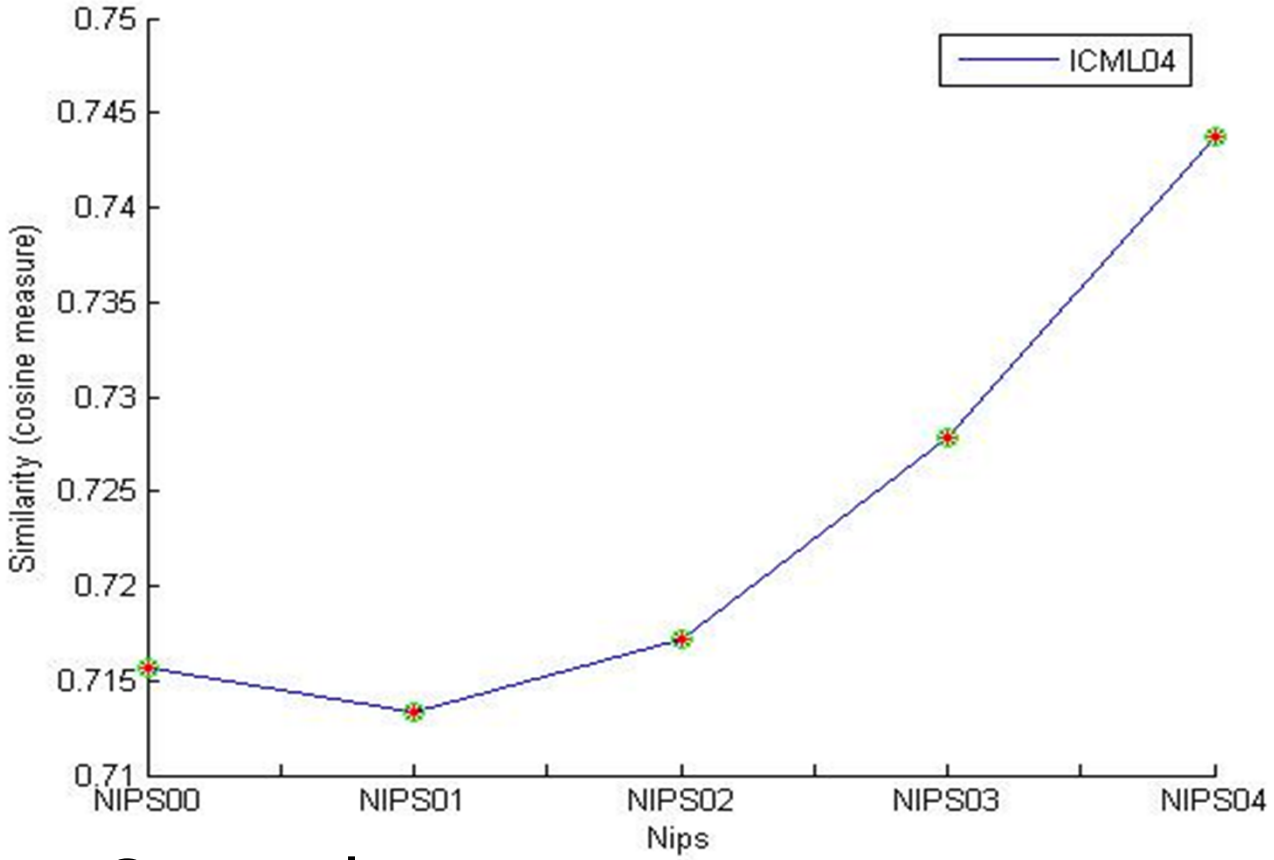
- Ok... maybe ICML'04 is most NIPS-y...
- So???

ICML'04 vs ICML-xx



- Of course, ICML'04 isn't that different from earlier ICML-xx's

ICML'04 vs NIPS-x



- Or maybe...
  NIPS-xx is getting more and more like ICML04??