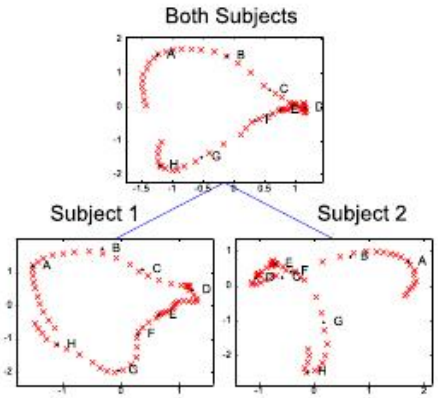
Report on ICML 2008

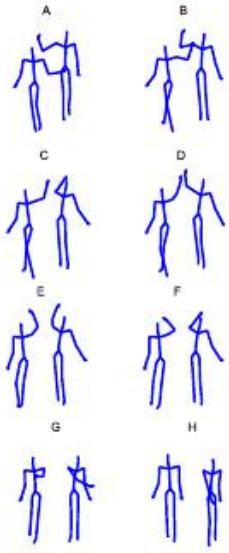
July 5-9 2008, Helsinki, Finland

Tutorials

- Painless embeddings of distributions: the function space view
  Alex Smola, Arthur Gretton, and Kenji Fukumizu
- Stochastic optimal control theory Robotics
  Bert Kappen, Marc Toussaint
- Dimensionality Reduction the Probabilistic Way →
  Neil Lawrence
- Graph. models and variational meth.: Message-passing and relaxations, Martin Wainwright
- Playing Machines: ML Applications in Computer Games
  Ralf Herbrich, Thore Graepel
- Beyond Convexity: Submodularity in Machine Learning ??
  Andreas Krause, Carlos Guestrin
- Theory and Applications of Online Learning
  Shai Shalev-Shwartz, Yoram Singer
- Visual Object Recognition and Retrieval, Rob Fergus
- Sparse Linear Models: Bayesian Inference and Experimental Design
  Matthias Seeger

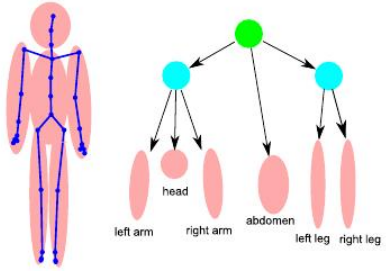
Neil Lawrence, Probabilistic Dimensionality Reduction

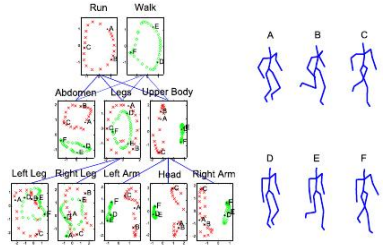




Idea

- Linear Latent Variable Model
  Represent data Y with a lower dimensional set of latent variables X
- $Y = WX + \textit{noise}$
  Roles of W and X are symmetrical; eigenvectors of $Y'Y$ and YY'.





- Make it hierarchical

Can use it generatively too

Wshops

- Bayesian Modelling Applications
- Evaluation Methods for Machine Learning
- Learning and Music
- Learning in Health Care Applications
- Nonparametric Bayes
- **PASCAL Large Scale Learning Challenge**
- Planning to Learn
- Prior Knowledge for Text and Language Processing
- Breakthroughs in Minimum Description Length ML
- Second Annual Reinforcement Learning Competition
- Sparse Optimization and Variable Selection

Invited talks

- Michael Collins, Structured Prediction Problems in Natural Language Processing
- Andrew Ng, STAIR: The STanford Artificial Intelligence Robot project
- Luc De Raedt, Logical and Relational Learning Revisited
- John Winn, Probabilistic models for understanding images

John Winn, Prob. models for understanding images

1. Define a context "world model"
   context: e.g. street $\rightarrow$ walls, buildings, sky
   influence: e.g., shape, which pixel influences which others
2. Generative approaches
   Select some source of variabilities as latent variables (shape $\rightarrow$ texture)
   Estimate those through EM (sensible initialisation !)
3. Discriminative approaches
   A regular grid of parts, + labels
   Classifier: random forests are nice and fast
   Take into account correlations: Layout Consistent Random Field
4. Combine both
   Multi-conditional learning: McCallum et al; Couchard Triggs
   Coupled priors: Lasserre Bishop
   Jigsaw + Potts model (variational EM)

My Top Four

- Empirical Bernstein Stopping, Mnih, Szevespari and Audibert **Use for Racing**
- Apprenticeship Learning with Linear Programming, Syed Bowling and Schapire
- Prospective: From eigenvalue problems to mean squares, Jieping Ye
- Prospective: Knows what it knows, Li, Littman, Walsh Beyond iid, with Active Learning and RL.

Notes on some papers

http://taonew.lri.fr/tiki-index.php?page=ICML+2008

Some other papers

High dimensional space

hyperdisk, Triggs

projection on L_1 ball, Duchi et al.

L_1 and L_2 SVM: Dual coordinate descent, Cho Jui Hsieh

Supervised dimensionality reduction, Rish et al.

Distance learning

Fast Solvers and Mixtures, Weinberger and Saul

From dissimilarities to ranking, from SDP to QP, Gray et al.

Bregman divergences, Cayton

Some others, 2

Reinforcement Learning

Predictive State Representations, the revenge

Learning All Optimal Policies with Multiple Criteria, Barrett et al.

Active RL; Doshi et al.

Multi-task RL, Lazaric et al.

And rare events, Frank, Mannor and Precup

Apprenticeship

Multiple demonstrations, Abbeel et al.

Spaced indexed dynamic programming, Ng

Bandits

Many faces of opportunism, Szita and Lorincz

Banditron, Kakade et al.

Exploration scavenging, Langford

Strategy evalution in Games with Importance Sampling, Bowling et al.

Some others, 3

Active learning (almost everywhere)

Multi-View, Wei Wang et al.

Ranking: active, listwise, ...

with multi-armed bandits. Joachims et al.

+ active, Donmez and Carbonell

Extreme values and quantiles

Saharon Rosset

Multiple Instances

Multiple instance ranking, Bennett et al

Learn from instance labels to bag labels, Hua-Yan Wang et al.

Some others, 4

Deep learning

Semi-Supervised Embedding, Weston, Ratle, Collobert

Quantitative analysis, Salakhutdinov et al.

Extracting and composing robust features, Bengio's band

Restricted Boltman Machines, Tieleman

Runtime, distributed,...

EM, Wolfe et al.

Optimized Cutting Plane, Sonnenburg

SMO on graphics processing units, Catanzaro et al.

Inverse Dependence on Training Set Size, Shalev-Shwartz and Srebro

Non-smooth Optimization

Quasi Newton for non-smooth convex, Yu et al.

Facts and Ambiance

Facts

583 papers submitted; 158 accepted; 539 attendees

Key stuff: optimization; semi-definite; densities (multi-tasks).

Innovations:

- online discussion of papers;
- video recording of talks: PASCAL.

Ambiance

Should ICML become a NIPS-bis?

Have an ACL like organization – tracks

Next

Dual bidding :-)

... Ambiance

I was invited to a “nice bar”
said yes, of course
on the road (hard to find, this bar...)
strange words... they give a fur coat in the bar... can only stay for 2 hours...
Can you guess why ?