Machine Learning and Statistics: What's the Connection?

Chris Williams

Institute for Adaptive and Neural Computation
School of Informatics, University of Edinburgh, UK

August 2006

Outline

- The roots of machine learning (ML)
- Problems addressed by ML
- Some examples
- Machine learning and statistics

What is Machine Learning?

- The goal of machine learning is to build computer systems that can adapt and learn from their experience. (Dietterich, 1999)
- Machine learning usually refers to changes in systems that perform tasks associated with artificial intelligence (AI). Such tasks involve recognition, diagnosis, planning, robot control, prediction, etc. (Nilsson, 1996)
- Some reasons for adaptation:
  - Some tasks can be hard to define except via examples
  - Adaptation can improve a human-built system, or track changes over time
- Goals can be autonomous machine performance, or enabling humans to learn from data (data mining)

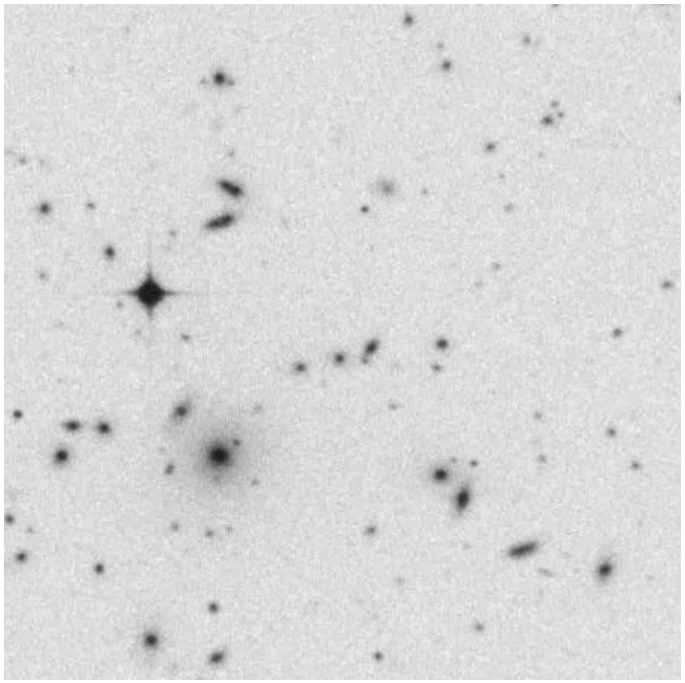
Roots of Machine Learning

- Statistical pattern recognition, adaptive control theory (EE)
- Artificial Intelligence: e.g. discovering rules using decision trees, inductive logic programming
- Brain models, e.g. neural networks
- Psychological models
- Statistics

Problems Addressed by Machine Learning

- **Supervised Learning**
  model $p(y|\mathbf{x})$: regression, classification, *etc*
- **Unsupervised Learning**
  model $p(\mathbf{x})$: not just clustering!
- **Reinforcement Learning**
  Markov decision processes, POMDPs, planning.

Supervised Learning: Examples

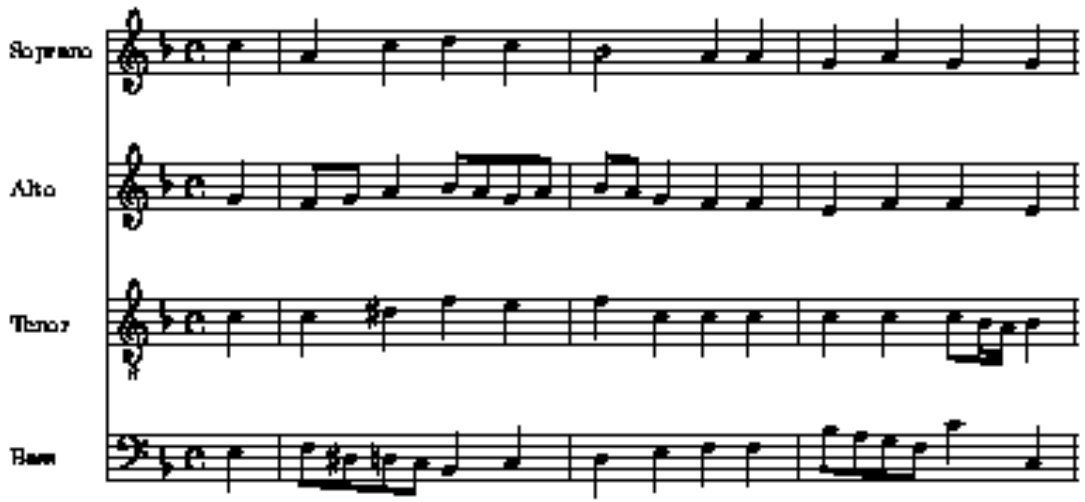


Star/galaxy classification
Fayyad, Weir, Djorgovski (1993)

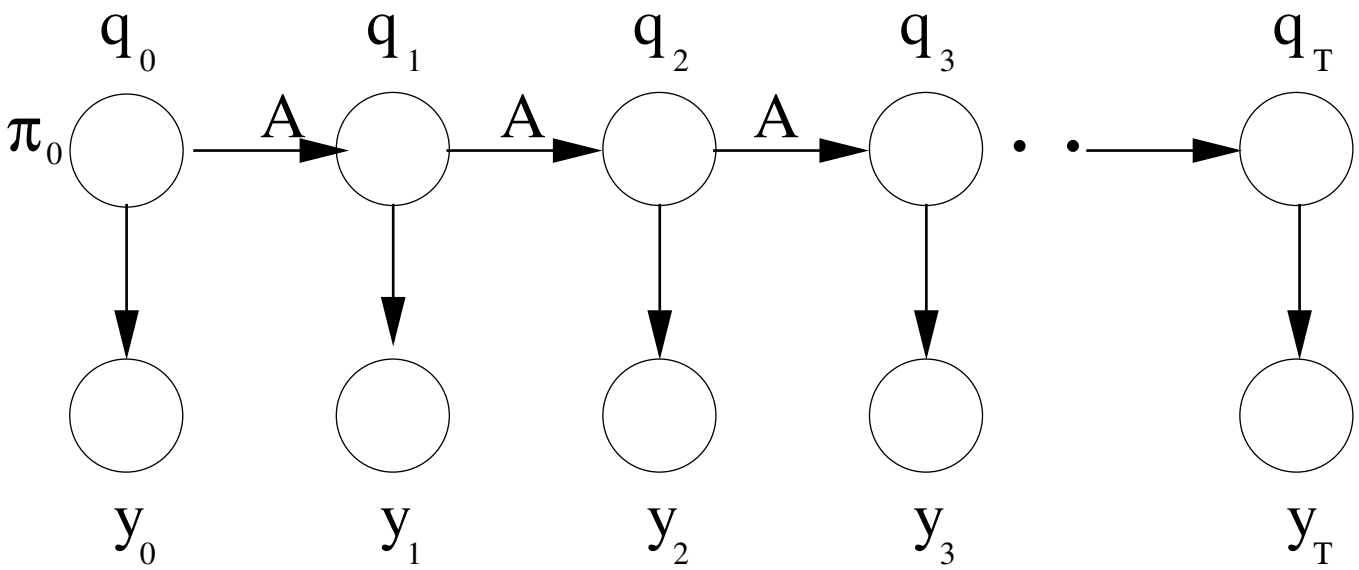
Robot arm inverse dynamics
Vijayakumar and Schaal (2000)

Example: Harmonizing Chorales in the Style of J S Bach

Bach chorale K54



Hidden Markov Model



- Visible states are the melody (quarter notes)
- Hidden states are the harmony (which chord)

- A supervised learning task
- Trained using labelled melody/harmony data
- Task: find Viterbi alignment for harmony given melody (or sample from $P(\mathrm{harmony}|\mathrm{melody})$)
- Actually uses HMMs for three subtasks: harmonic skeleton, chord skeleton, ornamentation
- Online demo at `http://www.anc.inf.ed.ac.uk/demos/hmmbach/`
- Moray Allan and Chris Williams (NIPS 2004)
- Bach chorale K54, HMM harmonization

Sprites: Learning about Objects from Images

An unsupervised task, latent variable modelling

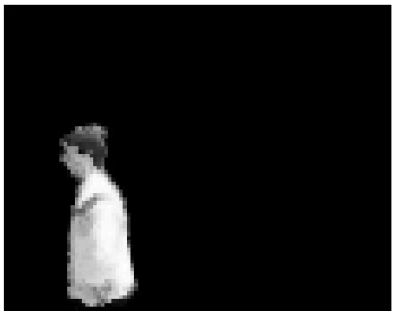
1 2 3



4 5 6



- *Sprite*: a graphics term for a "cardboard cutout" model having a shape and an appearance
- Also need to cope with image transformations (e.g. object motion)



foreground (original)

mask (original)

foreground (transformed)

mask (transformed)

- For a single object

$$p(\mathbf{x}) = \sum_{j=1}^{J} p_j p(\mathbf{x}|T_j)$$

  J can be very large. Model fitted using EM
- Note that multiple objects combine by *occlusion*
- With L objects, would naively need to enumerate $O(J^L)$ settings of the latent variables
- Jojic and Frey (2001): variational inference; Williams and Titsias (2003), greedy algorithm

Movie

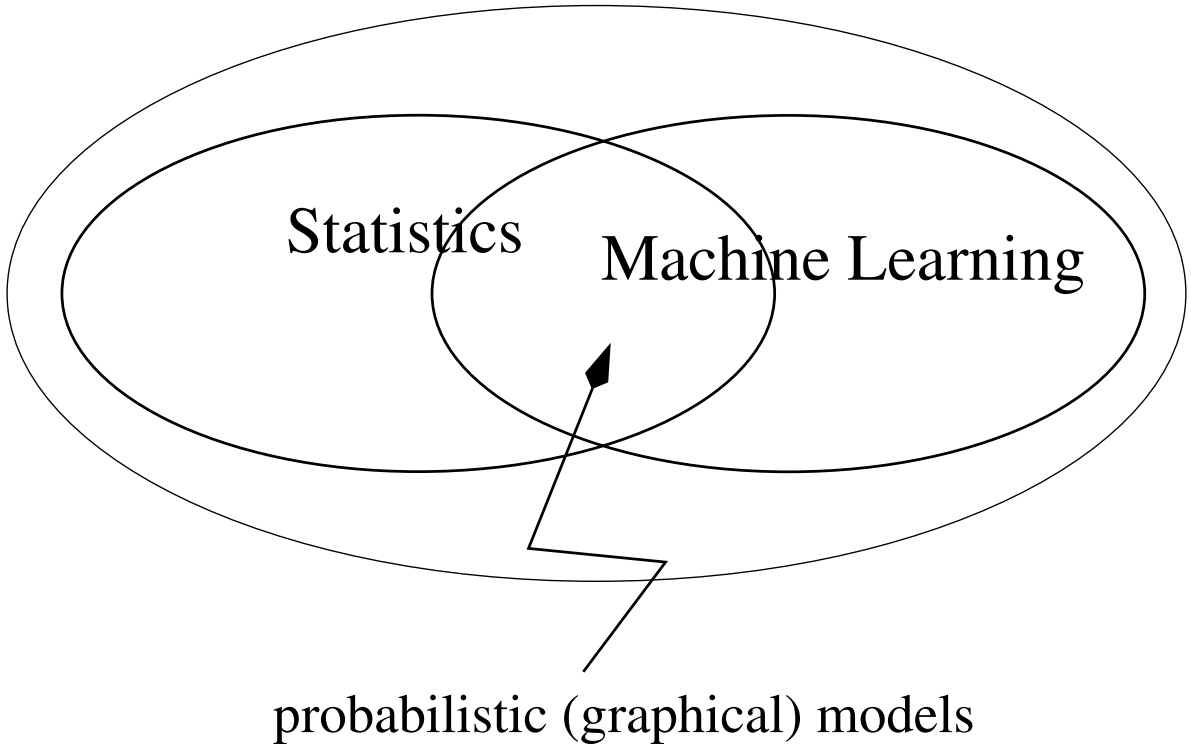
Other Examples

- Automatic recommender systems (collaborative filtering)
- Bioinformatics (sequence data, microarrays, etc)
- Information retrieval (Google)
- (Neonatal) condition monitoring
- Fraud detection
- Topic models for documents
- Statistical natural language processing (e.g. parsing)
- : RL: TD-gammon, elevator dispatching

Some Theoretical Issues in Machine Learning

- Computational Learning Theory (COLT) community
- Uniform convergence bounds, VC dimension, learning curves PAC-Bayesian analysis etc
- Prediction using expert advice (Cesa-Bianchi and Lugosi); e.g. combining/selecting predictions from a pool of experts
- Algorithms for and analysis of approximate inference methods in probabilistic graphical models

Machine Learning and Statistics



- Same models, but different problems?
- Not all machine learning methods are based on probabilisic models, e.g. SVMs, non-negative matrix factorization

Some Differences

- Statistics: focus on understanding data in terms of models
- Statistics: interpretability, hypothesis testing
- Machine Learning: greater focus on prediction
- Machine Learning: focus on the analysis of learning algorithms (not just large dataset issues)

Summary

- Rich variety of problems addressed by machine learning in supervised, unsupervised and reinforcement learning
- Some differences, but a nexus wrt probabilistic models, graphical models
- Bidirectional exchanges of ideas and methods
- As statisticians, what is your view?

Slide from Rob Tibshirani (early 1990s)

NEURAL NETS	STATISTICS
network	model
weights	parameters
learning	fitting
generalization	test set performance
supervised learning	regression/classification
unsupervised learning	density estimation
optimal brain damage	model selection
large grant = \$100,000	large grant= \$10,000
nice place to have a meeting: Snowbird, Utah, French Alps	nice place to have a meeting: Las Vegas in August

What is data mining?

Data mining is the analysis of (often large) observational data sets to find unsuspected relationships and to summarize the data in novel ways that are both understandable and useful to the data owner. Hand, Mannila, Smyth

We are drowning in information, but starving for knowledge! Naisbett

[Data mining is the] extraction of interesting (non-trivial, implicit, previously unknown and potentially useful) information or patterns from data in large databases. Han