Statistical Machine Learning Overview

Lecture 1

Doug Aberdeen, October 8, 2004

National ICT Australia

Definition

Machine learning is the study of how machines can adapt to their environment

- Programmer
- Options in word-processor
- Web advertising
- Voice recognition

Statistical Machine Learning

To deal with uncertainty we use statistics

- Observations from most environments are *noisy*
- Consequences of machine actions are noisy
- Machines need to reason about uncertainty explicitly

Why do this course?

Machine learning can help researchers

- Discover trends (data mining)
- Automate hard problems (inverse kinematics)
- Automatically adapt to changing environments
- Theorise about the way the mind works

(Rough) Course Outline

1. Background complexity and probability
2. Optimisation
3. Bayesian Methods
4. Clustering
5. Neural Networks
6. Kernel Methods
7. More Kernel Methods
8. Principal Component Analysis
9. Graphical Models
10. Markov Decision Processes/RL
11. Inductive Logic Programming/Relational Learning

Admin

- Useful book: Tom Mitchell *Machine Learning* 1997
- Assessment: 3/4 assignments; 2 theory, 2 practical; about 5 hours each
- Monday, Tuesday, Thursday 10am–12. Friday?
- All lectures here except 28/10 (A207 RSISE)
- End of lectures: Nov 4, deadline for assignments Nov 21
- `http://csl.anu.edu.au/~daa/courses.html`

NATIONAL ICT AUSTRALIA

Dichotomies in Machine Learning

Artificial Intelligence: Inductive (SML) vs. Deductive (Logic)

- Inductive
  - Regression, neural networks, support vector machines (SVMs), graphical models
  - Often example based. Past emphasis on treating data as vectors of numbers
- Deductive
  - Forward/Backward chaining, theorem proving, model checking, classical planning
  - Emphasis on structured knowledge representation
- Cross-over: ILP, relational learning, kernels for structure

Representation & Optimisation

- A hypothesis space is defined by the chosen representation and optimisation methods
- A hypothesis is a trained regression or classification tool
- Hypothesis representations: clusters, decision trees, kernel+weights
- Optimisation methods: linear prog., gradient, EM
- Hypothesis space = feasible values of w_1, w_2

$$\begin{aligned}\Pr[rain] =& w_1 * [clouds = 1, fine = 0] + \\ & w_2 * [spring = 1, other = 0]\end{aligned}$$

Use evolution? Gradient descent? EM?

Complexity Review

- P: decided in polynomial time
  E.g. Insertion sort $O(n^2)$, Quicksort $O(n \log n)$
- NP: non-deterministic P: decided in polynomial time by a non-deterministic machine (brain/quantum computer). Verified in poly time
  E.g. Travelling Salesman $O(n!)$
- NP-Complete: difficult problems in NP. All problems in NP can be reduced to an NP-Complete problem
- NP-Hard: at least as hard as NP-Complete, possibly harder
- PSpace: polynomial space
  E.g., Go. $O(x^n)$

NATIONAL ICT AUSTRALIA