Learning from Data

Amos Storkey, School of Informatics
University of Edinburgh
http://www.anc.ed.ac.uk/∼amos/lfd/

Semester 1, 2004

Introduction

- Welcome
- Administration
  - Handouts
  - Books: Mitchell: Machine Learning.
  - Assignments
  - Tutorials
  - Course representatives
  - Exams
  - Email

Acknowledgement: I would like to that David Barber and Chris Williams for permission to use course material from previous years.

Course details

- 20 lectures 5.10 to 6.00pm Mon and Thurs
- 8 tutorials (compulsory). Start Thurs week 2.
- 1 week free of lectures
- 2 assessments (20%) (week 5 and week 9)
- 1 exam (80%)

Relationships between courses

LfD Learning from Data. Basic introductory course on supervised and unsupervised learning

RL Reinforcement Learning.

MLSC Machine Learning and Sensorimotor Control.

PMR Probabilistic modelling and reasoning. Focus on probabilistic modelling. Learning and inference for probabilistic models, e.g. Probabilistic expert systems, latent variable models, Hidden Markov models, Kalman filters, Boltzmann machines.

DME Data mining and Exploration. Using methods from PMR to deal with practical issues in learning from large datasets.

Maths and Learning from Data

- Learning from Data will involve a significant number of mathematical ideas and a significant amount of mathematical manipulation.
- For those wanting to pursue research in any of the areas covered, you better understand all (or almost all) the maths.
- Others should understand the *ideas* behind the maths. It is obviously preferable to understand the detail too, but understanding in a procedural way (i.e. how to program an algorithm) will usually be sufficient.
- The detailed derivation of the backpropagation algorithm will not be examinable, but a procedural knowledge of the algorithm is.

Potential difficulties

- I've done next to no mathematics. This course is impossible.
- I understood calculus at 6 months, and got a maths degree shortly after I was potty trained. Why are you wasting my time.
- But what is the philosophy behind...
- Is this really scientific?
- How is this going to help me in course X?
- I am finding this course hard. What shall I do?

Course Outline (not necessarily in order)

- Introduction. Thinking about data.
- Preliminaries: supplementary maths, MATLAB.
- Understanding data, and models of data: generative versus discriminative, supervised or unsupervised.
- Gaussian density estimation, dimensionality reduction.
- Visualisation 1, decision trees.
- Naive Bayes.
- Regression and classification, linear and logistic regression, generalised linear models.
- Mixture models, class conditional classification, visualisation 2.
- Generalisation, perceptron, layered neural networks, radial basis functions, nearest neighbour classifiers, kernel methods.
- Visualisation 3, conclusions.

Why Learn from Data?

- Growing flood of online data.
- People already learn from data. Automated methods increase the scope.
- Recent progress in algorithms and theory.
- Computational power is available.
- Budding industry.
- Because we can.

Examples

- Science (Astronomy, neuroscience, medical imaging, bio-informatics).
- Retail (Intelligent stock control, demographic store placement)
- Manufacturing (Intelligent control, automated monitoring, detection methods)
- Security (Intelligent smoke alarms, fraud detection).
- Marketing
- Management (Scheduling, time tabling, competitor analysis warning systems).
- Finance (risk analysis, micro-elasticity analysis).

Thinking about Data

- This course is not computer science as you know it.
- Computer science and algorithms:
  - Computer science as algorithm generation.
  - If the algorithm works it is good. If it doesn't it is bad.
- Machine Learning: the algorithm and the model.
  - Model encodes understanding about the data.
  - Algorithm comes from the model (and a bit of maths).
  - Algorithms give different approximations.

Thinking about Data

- Learning from data is not magic.
- Prior beliefs/assumptions + data $\rightarrow$ posterior beliefs.
- The model encodes beliefs about the generative process of the data.
- or... The model encodes beliefs about the features/characteristics in the data.
- Can do nothing without some prior input - no connection between data and question.

Illusions

- Logvinenko illusion
- Meteor - internet ray tracing entry.

Example

- 3 Boolean variables. Data set:

$$\begin{array}{ccc} 1 & 0 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & x \end{array}$$

- What is x?
- We cannot say. We have no information at all about how any of these data items is connected.
- “No free lunch”

No Free Lunch

- Try to predict $C \in \{0,1\}$ from $A, B \in \{0,1\}$.
- No noise - given A,B then C is always the same.
- Possible hypotheses. $C = 1$ if and only if values for A,B are in a particular set. One example hypothesis is:
- $\{(1,1),(0,1),(1,0)\}$ (i.e. C = A OR B). Here $(0,1)$ means $A = 0$, $B = 1$.
- If no bias, then there are ${}^4C_0 + {}^4C_1 + {}^4C_2 + {}^4C_3 + {}^4C_4 = 16$ equally possible hypotheses - the hypothesis space.
- Each data point reduces the size of the hypothesis space, but when we attempted to predict C given an unseen set of values of A,B the number of hypotheses predicting $C = 1$ is the same as the number predicting $C = 0$.

No Free Lunch contd.

- Eg suppose we have data $(a = 1, b = 1, c = 0)$, $(a = 0, b = 1, c = 1)$, then the remaining $C = 1$ hypothesis space for (AB) is
- $\{(0, 1)\}$, $\{(0, 1), (1, 0)\}$, $\{(0, 1), (0, 0)\}$, $\{(0, 1), (1, 0), (0, 0)\}$.
- Suppose we now query $(a = 1, b = 0)$. Two of the possible hypotheses predict $C = 1$, and two predict $C = 0$.
- Suppose we now see data $(a = 0, b = 0, c = 0)$. Hypothesis space is $\{(0, 1)\}$, $\{(0, 1), (1, 0)\}$.
- One of the remaining hypotheses predict $C = 1$, and the other predicts $C = 0$.
- No matter what data you receive the number of hypotheses predicting one values for unseen data will equal the number predicting the other value.

Suspect Terms

- Model Free.
- Bias Free. Unbiased.
- No prior information.
- Generally applicable.

Machine Learning and Probability

- Probability theory is key: probabilistic understanding of uncertainty.
- Bayesian methods: machine learning is really just statistics?
- Bayesian methods are non-trivial:
  - Hard to really understand the full implications of a probability distribution.
  - Hard to accurately represent your prior beliefs, and represent them in a way that is amenable to computation.

Summary

- Fairly mathematical course. Try to keep on top of it.
- No free lunch.
- Plethora of practical needs.
- Models not algorithms.
- Probability theory is key.