Learning from Data

Amos Storkey, School of Informatics

Semester 1

http://www.anc.ed.ac.uk/∼amos/lfd/

Introduction

- Welcome
- Administration
  - Online notes
  - Books: See website
  - Assignments
  - Tutorials
  - Exams

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

Course details

- 18 lectures 5.10 to 6.00pm Mon and Thurs
- 7 tutorials (compulsory). Start Thurs week 3.
- 2 assignments (20%) (week 4 and week 8)
- 1 exam (80%)
- Course notes.

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.

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.
- 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.

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).
- Over to you...

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.

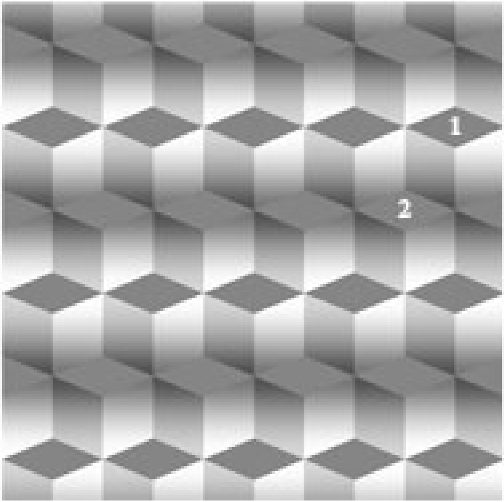
Thinking about Data

- 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.

Logvinenko Illusion



Example

- 3 Boolean variables. Data set:

1	0	1
1	1	1
1	1	0
0	1	0
0	0	x

- 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.

Prior assumption

- Prior assumption: say something about the ways C is allowed to relate to A, B.
- e.g. ($C = A$ OR B) OR ($C = A$ AND B) OR ($C = A$ XOR B).
- $(a = 1, b = 0, c = 1)$, $(a = 0, b = 1, c = 1)$:
- So either OR or XOR. Can predict $(a = 0, b = 0, c = ?)$.

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.