Machine Learning 2.01: Introduction

Tom S. F. Haines
T.S.F.Haines@bath.ac.uk

UNIVERSITY OF BATH

Before we get started: Schedule

- Per week:
  - 2 hour lab session.
  - 2 lectures.
- Panopticon should work this time!
- Slides probably after lectures (sorry).
- Everything on Moodle, including a forum.

  (best to ask questions there!)

Before we get started: Coursework

- Marks:
  - 60% from 3 lab write ups (12% each) + 1 group project (24%).
  - 40% from final report (maximum 3000 words).
  - Submission:
    - Lab write ups: Filled in Jupyter workbook.
    - Group project: Individual report + code.
    - Final report: Report + presentation + code.
  - Has different patterns to ML1!

Before we get started: Coursework

- Marks:
  - 60% from 3 lab write ups (12% each) + 1 group project (24%).
  - 40% from final report (maximum 3000 words).
  - Submission:
    - Lab write ups: Filled in Jupyter workbook.
    - Group project: Individual report + code.
    - Final report: Report + presentation + code.
  - Has different patterns to ML1!

- Using Jupyter:
  - Workbook via a web interface.
  - Python 3 + machine learning libraries.
  - For your own computers: `https://www.anaconda.com/download`
  - Careful of which version you're running.
    (Anaconda 3, not 2. Can't tell if you use start bar search!)

Before we get started: Difficulty

- Lots of extremely hard content ...
- ... don't expect to understand everything!

- Some overlap with optional units.

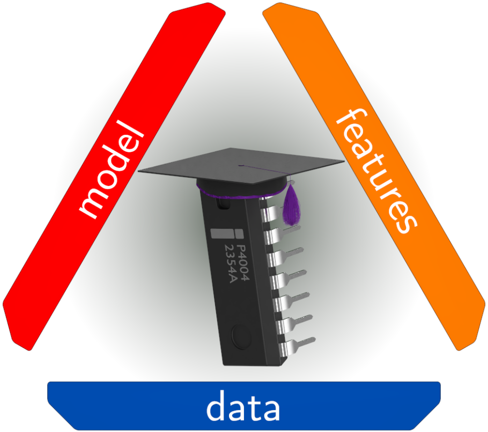
- Aim of unit:
  Can read and implement a research paper with help from Google.

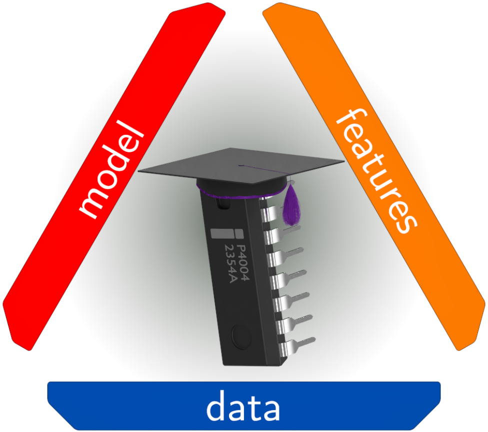
Rest of this lecture

- Overview of machine learning, in terms of ML2.

Machine Learning Triangle

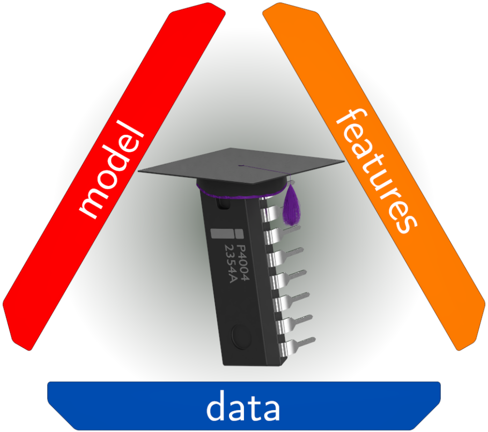
- Good performance $\Longrightarrow$ All done well.





Machine Learning Triangle

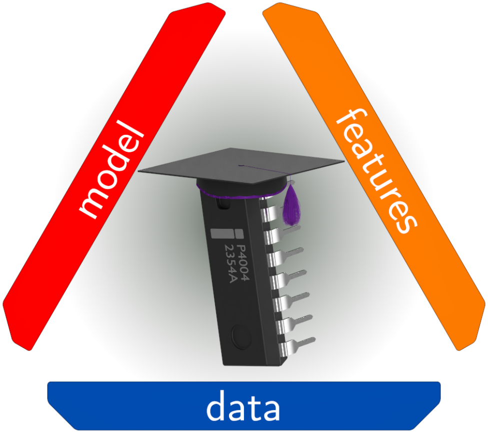
- Good performance $\implies$ All done well.
- Data; new for ML2:
  - New problems $\implies$ new data.
  - But nothing technical.



Machine Learning Triangle

- Good performance $\implies$ All done well.
- Data; new for ML2:
  - New problems $\implies$ new data.
  - But nothing technical.
- Models; new for ML2:
  - More examples.
  - Entirely new "kinds".
  - Often Bayesian.

Machine Learning Triangle



- Good performance $\implies$ All done well.
- Data; new for ML2:
  - New problems $\implies$ new data.
  - But nothing technical.
- Models; new for ML2:
  - More examples.
  - Entirely new “kinds”.
  - Often Bayesian.
- Inference (model fitting); new for ML2:
  - Better techniques!
  - In particular:
    Monte Carlo Markov chain [MCMC] and Variational methods.

Model vs Inference I

- Inference:
  - Definition: Any conclusion reached from deduction and/or data.
  - Sometimes used specifically however,
    e.g. Given a causal model inference is going backwards, predicting going forwards.
  - Specialisations:
    - Prediction
    - Estimation
    - Fitting
    - Learning

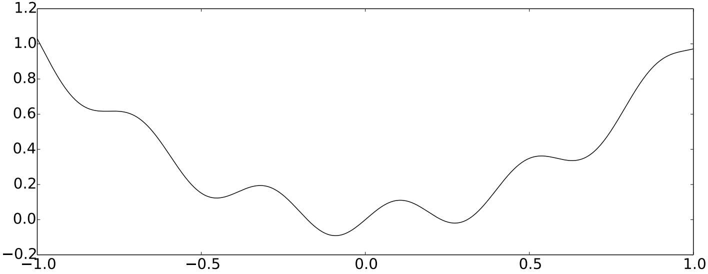
Model vs Inference I

- Inference:
  - Definition: Any conclusion reached from deduction and/or data.
  - Sometimes used specifically however,
    e.g. Given a causal model inference is going backwards, predicting going forwards.
  - Specialisations:
    - Prediction
    - Estimation
    - Fitting
    - Learning

- A model is not it's inference technique(s)!
  (common mistake, as often entwined)
- e.g. Linear model has two methods for inference:
  Analytic solution or gradient descent.

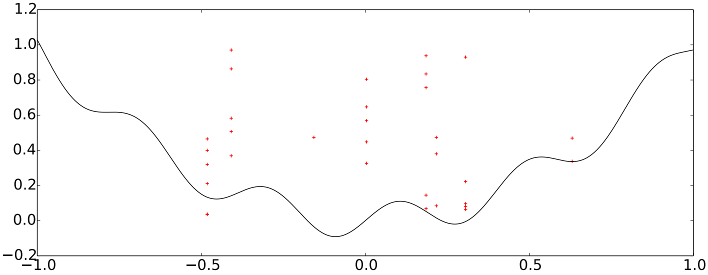
- Linear model: They get the same answer.
- **This is not true in general!**

Model vs Inference II



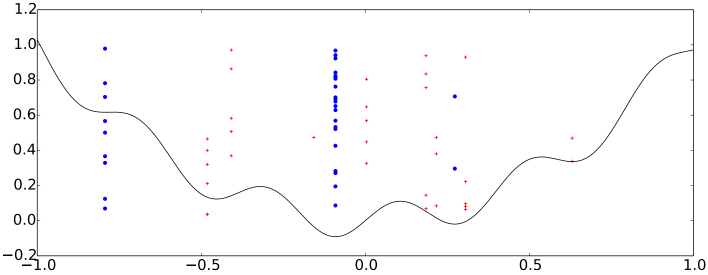
Consider optimisation to find minimum of simple function

Model vs Inference II



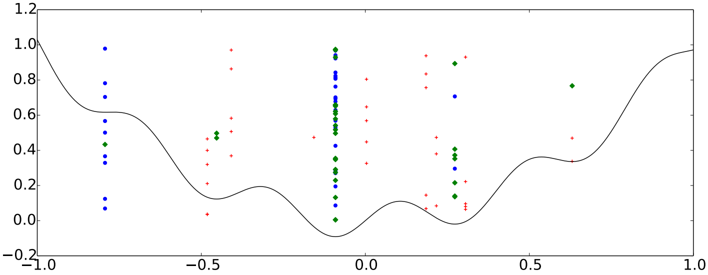
Simple gradient descent (point converged to, y value randomised)

Model vs Inference II



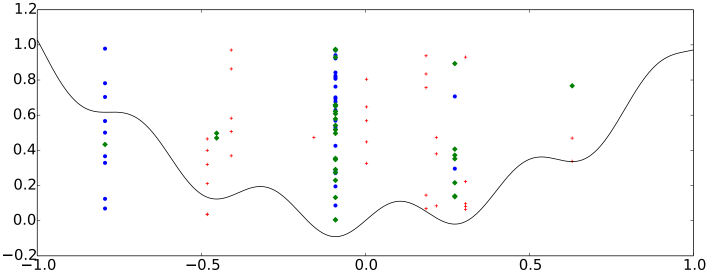
Gradient descent with step size reduction

Model vs Inference II



Gradient descent with step size reduction and momentum

Model vs Inference II



Gradient descent with step size reduction and momentum

- Many research papers: Authors try lots of approaches, publish only the first one that works! (this is publication bias – bad, and not going away)
- Finding real machine learning examples is hence hard!

Model vs Inference III

- Simple model $\implies$ Reliable inference.
- Complex model $\implies$ Inference is hard, unreliable.

Model vs Inference III

- Simple model $\implies$ Reliable inference.
- Complex model $\implies$ Inference is hard, unreliable.

- Have a trade off to make!
- If problem really important and have time:
  Chain models, e.g. learn simple model, use it to initialise inference for more complex model.

Model vs Inference III

- Simple model $\implies$ Reliable inference.
- Complex model $\implies$ Inference is hard, unreliable.

- Have a trade off to make!
- If problem really important and have time:
  Chain models, e.g. learn simple model, use it to initialise inference for more complex model.

- In ML2:
  - MCMC is simple to code, slow to run, gives correct answers.
  - Variational is hard to code, fast to run, gives approximate answers.

Model kinds

Supervised	Deterministic
Semi-supervised	Probabilistic
Weakly-supervised	**Bayesian**
Unsupervised	

- Will be seeing all of above in ML2, but bold ones more.

- Also lots of **Generative** models: Can generate data as well as answers!
- Plus training with missing data.

UNIVERSITY OF BATH

Being Bayesian

$$P(y|x) = \frac{P(x|y)P(y)}{P(x)}$$

You should all know this now, but you may not appreciate **why** it's so important.

Priors are unavoidable I

A

B

- Two bowls, I select one at random.
- I select a ball at random – it is red.
- What is the probability of each bowl?

UNIVERSITY OF BATH

Priors are unavoidable II

We know:

- $P(\texttt{ball} = \texttt{red} | \texttt{bowl} = A) = \frac{3}{6}$
- $P(\texttt{ball} = \texttt{blue} | \texttt{bowl} = A) = \frac{3}{6}$
- $P(\texttt{ball} = \texttt{red} | \texttt{bowl} = B) = \frac{4}{6}$
- $P(\texttt{ball} = \texttt{blue} | \texttt{bowl} = B) = \frac{2}{6}$

But we want $P(\texttt{bowl} | \texttt{ball} = \texttt{red})$. Bayes rule:

$$P(\texttt{bowl} | \texttt{ball} = \texttt{red}) = \frac{P(\texttt{ball} = \texttt{red} | \texttt{bowl}) P(\texttt{bowl})}{P(\texttt{ball} = \texttt{red})}$$

Priors are unavoidable III

To solve we **must** select a prior, $P(\texttt{bowl})$.
You might assume $P(\texttt{bowl} = A) = 0.5$ for instance.

Note:

- There is no limit to the influence of the prior!
- The prior is needed to calculate $P(\texttt{ball})$ in this case.
  But can ignore $P(\texttt{ball})$ and normalise.

Priors=Assumptions

- **All** models include assumptions.
- The Bayesian approach makes them **all** explicit.

- Some people object to priors – they are wrong.

Incremental learning

- Incremental learning = Updating your model as more data arrives.
- Very important for finance, websites etc.
- Lecture on it next week.

- Posterior from the first batch of data can be the prior for the second.
- Incremental learning can only be done both efficiently and correctly if Bayesian.

Non-Bayesian inference

- Ideally, both the model and inference are Bayesian.
- This is hard, but what much of ML2 is about.

Non-Bayesian inference

- Ideally, both the model and inference are Bayesian.
- This is hard, but what much of ML2 is about.

- It can be reasonable to have a Bayesian model but non-Bayesian inference.
- It can also make life a lot easier!
- Typical example:
    - Bayesian inference on model parameters given data.
    - Select maximum a-posteriori [MAP] model parameters.
    - Inference on new data uses MAP model parameters.

Wednesday Machine Learning Seminars

http://www.bath.ac.uk/imi/events/machine-learning.html

All but last in 8 West, Room 2.1, 2.15 - 4.00pm (need to sign up).

- **Good analytics needs good data and that needs good metadata**, Professor Mandy Chessell, IBM, 28 February.
- **Accountable and explainable AI**, Dr Sandra Wachter, Oxford Internet Institute, 7 March.
- **From computational metaphysics towards computational pseudo-ethics**, Dr Christoph Benzmueller, Free University of Berlin, 14 March.
- **A series of mini talks by University of Bath researchers using machine learning**, 21 March.
- **Governance: Arguing for greater transparency, accountability and means for citizen interventions**, Dr Joanna Redden, Cardiff Data Justice Lab Data, 11 April.
- **Of, for, and by the people: The legal lacuna of synthetic persons**, Dr Joanna Bryson, University of Bath, 18 April.
- **Media and artificial intelligence**, Professor Nello Cristianini, University of Bristol, 25 April.
- **Machine learning for functional genomics**, Dr Nick Priest, University of Bath, 2 May.

- **Grand Public Debate on Machine Learning**, 9 May 2018, 5.15 - 7.00pm, The Forum, Bath, BA1 1UG.

UNIVERSITY OF BATH

Racing

- Self-driving F1 car competition at Silverstone in July.
- Electrical engineering is going to enter “something”.
- Whole thing is a car-crash (sorry)...

- Things that may be influenced:
    - Group project
    - Masters projects

Further reading

- “Information Theory, Inference and Learning Algorithms“ by **David J. C. MacKay**.
- “Pattern Recognition and Machine Learning“ by **Christopher M. Bishop**.

Summary

- Overview with a bit of perspective.
- Next lecture: Density estimation.