

Syllabus

- Project Stuff: you can do one on your own, but you'll be graded the same as those who work in teams.
Hopefully there will be visitors to watch the final presentations, and we will WOW them with our learning and this class will become permanent
- Books: *Pattern recognition and machine learning* – claims to be for first-year PhD students with zero math background, but this is a lie
 - A good reference, but will be painful to learn machine learning from. You are forewarned.

A first course in Machine Learning – available free from UoA library. More friendly.

Not the Syllabus

Features = characteristics of the data

Machine learning: how to transform the features of the data to make them more informative

Paradigms of Machine Learning:

- 1) Supervised training
 - a. Set of labelled data
 - b. Algorithm
 - c. Classification tool
- 2) Validation methods
 - a. Cross-validation
 - b. Bias-variance analysis (for example: in the case of cancer screening, it's much better to throw a false positive than a false negative)
 - Bias-variance analysis is “probably the cornerstone of this class”
 - It isn't performed enough

Knowledge discovery – when we don't know what we're looking for

- Unsupervised learning – is a lot of what's done in planetary science
- Q: did gmail use user-labelled sets to train originally? A: no?

Machine learning books often have a **and then a miracle occurs** step that makes them bloody difficult to follow

Intro to python

numpy: python's matlabby equivalent (note matlab counts from 1, python sensibly starts from 0)

Q: what indicates that you have a list vs an array? An excellent question! **Add** an explanation of the *shape* command, which would show you *1x4* vs *size 4*

So basically *rand* gives different output based on how many args it gets

Linalg – linear algebra library of numpy

Ask me about the zero: it's referring to a topo, which technically has 2 elements, and using [0] means the 1st element