MACHINE LEARNING

Dr Ioannis Patras
i.patras@qmul.ac.uk

Course Format

- 12 weeks, 2 (+1 occasionally) hours of lectures per week
- 2 coursework assignments organised in 6 lab sessions of 2 hrs each as follows:
  - 2hrs lab sessions for part 1 of assignment X
  - 2hrs lab sessions for part 2 of assignment X
  - 2hrs lab sessions for assessment of part1 and part2
- Assesment: 80% final exam, 20% coursework

Recommended texts

- Kevin Murphy. Machine Learning. A probabilistic perspective. MIT Press. http://www.cs.ubc.ca/~murphyk/MLbook/
  [Main book for the module]
- D. Barber: Bayesian Reasoning and Machine Learning
  [Comprehensive, a bit advanced, free online pdf]

- Duda, Hart and Stork: Pattern Classification (2001)
  [Good, comprehensive]
- Bishop: Pattern Recognition and Machine Learning (2006)
  [Good and reasonably affordable.]
- Barber (2002). [Thorough coverage of Probability, RVs, etc.]

Machine Learning

- Use of “intelligent” techniques for analysis and processing of signals and data.

Applications include:

- Spam Filtering
- Stock market prediction
- Autonomous vehicles
- Medical diagnosis
- Search engines (google search, google image search)
- Speech recognition
- News grouping, Ad placement
- Netflix recommendations

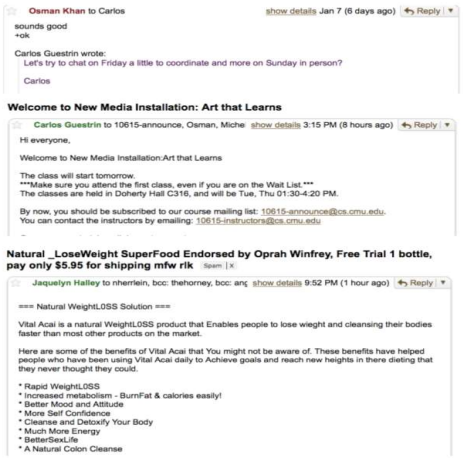
A Few Quotes

- “A breakthrough in machine learning would be worth ten Microsofts” (Bill Gates, Chairman, Microsoft)
- “Machine learning is the next Internet” (Tony Tether, former director, DARPA)
- “Machine learning is the hot new thing” (John Hennessy, President, Stanford)
- “Web rankings today are mostly a matter of machine learning” (Prabhakar Raghavan, former Dir. Research, Yahoo)
- “Machine learning is going to result in a real revolution” (Greg Papadopoulos, former CTO, Sun)

Machine Learning by examples

Classification

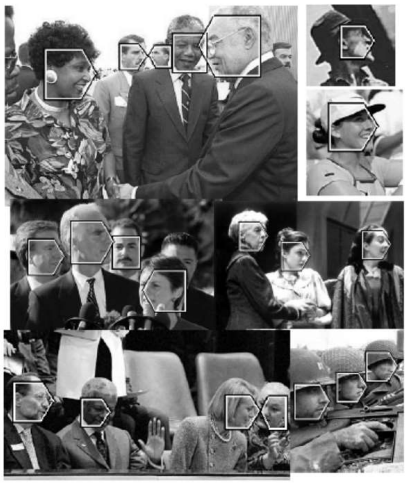
Spam Filtering



Spam/no spam

Face detection

Example training images for each orientation

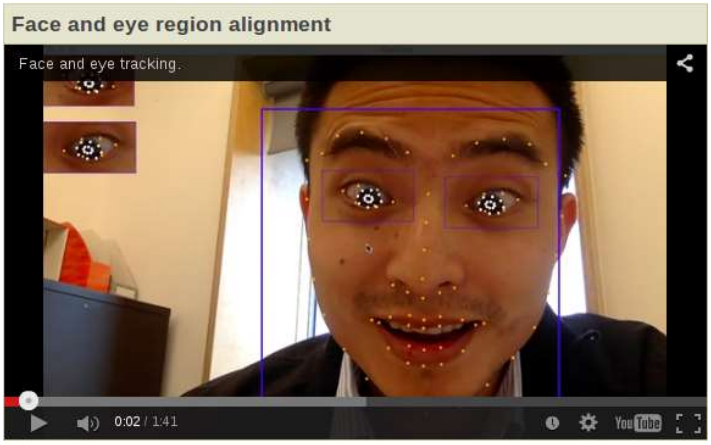


Machine Learning by examples

Regression

Stock market prediction

Facial landmark localisation



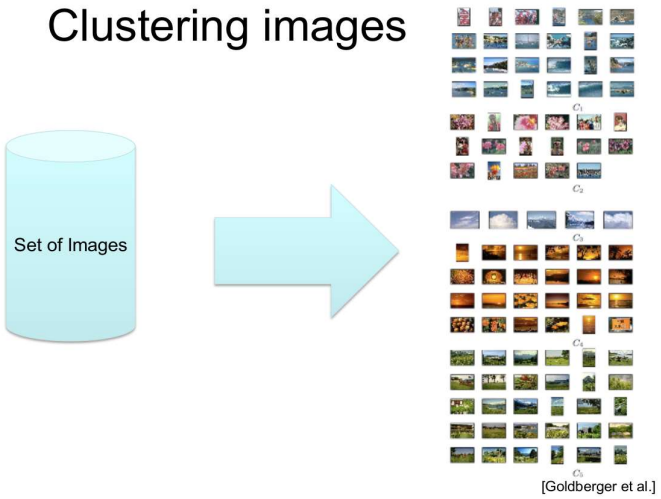
https://www.youtube.com/watch?v=ONnobin5GBs

Machine Learning by examples

Clustering

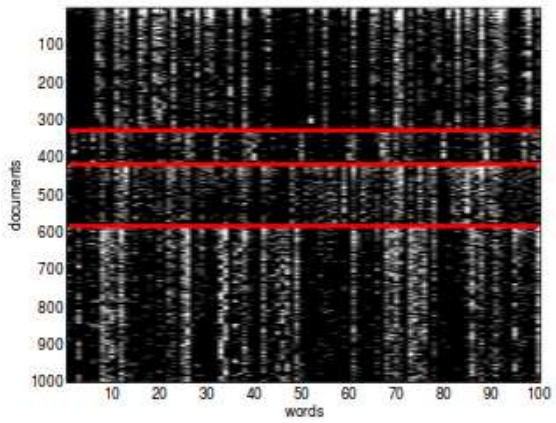
Clustering Images

Clustering images



[Goldberger et al.]

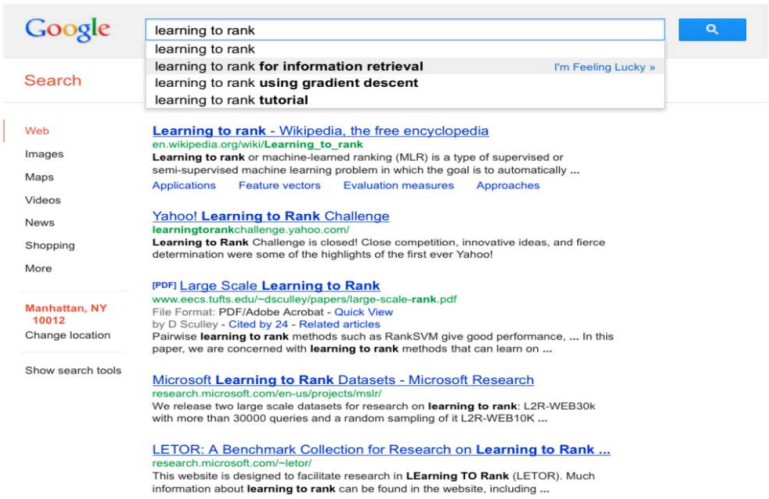
Clustering documents



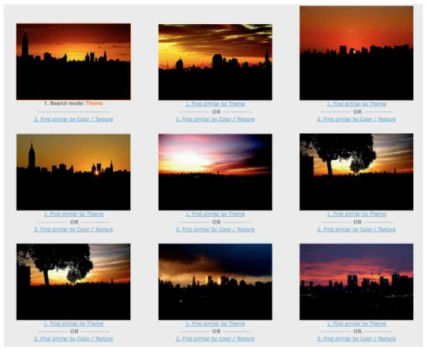
Machine Learning by examples

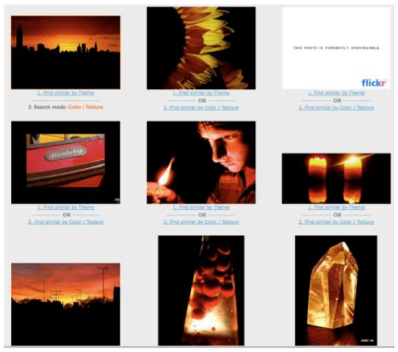
Ranking

Text-based search



Visual search – find similar images





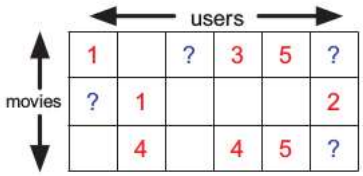
http://www.tiltomo.com/

Machine Learning by examples

Recommendation

Recommendation

Recommendation systems



Netflix competition (www.netflixprize.com)

Machine learning competition with $1m prize

Types of Learning Problems

- Supervised Learning
  Regression, Classification
- Unsupervised Learning
  Clustering
- Reinforcement Learning
  Policy learning

Supervised Learning: learn a prediction function

Learning a function f when the target is known for the training data.

Given $\{(x_i, y_i)\}^N_{\{i=1\}}$, $x_i \in \mathrm{X}, y_i \in Y$

Learn $f: X \rightarrow Y$

House price prediction [R]

Stock market prediction [R]

Categorizing (Classification).

Supervised Learning (regression)

Learn to predict the price of a house (target) given the size of the house (features)

Supervised Learning (classification)

Learn to predict the whether a tumour of a given size (feature) is malignant or not (target)

Unsupervised Learning

Categorise these LEGO bricks into groups

Can you write an algorithm to explain how you grouped them?

No explicit target was given.

Reduce dimensions

Compress data

Visual Hierarchies

Unsupervised Learning (clustering)

Given the size and hue/colour (features) of several fruits, group them into clusters

Reinforcement Learning

- LEGO example. Choose 2 blocks one after the other, and I'll tell you how well you've done.
- A kind of denuded supervised learning where you just have a hotter/colder signal, not the complete right answer. Its as if I give you an exam and a mark, but never the model answers. At least there is more guidance than with unsupervised learning.

Reinforcement Learning

Given the outcome (reward) of previous checkers games, learn the move (action) you should make given a checkerboard configuration (state).

Other ways for Classifying Machine Learning Methods

Parametric Methods: Learn a low dimensional set of parameters, e.g. weights in a neural network, throwing away the training data points.

Non-parametric: Keeps the training data points throughout, e.g. k-nearest neighbor methods.

Think of the whole of ML like this

Feature Selection: Pre-process the data based on domain specific/expert knowledge.

Model: Choose a machine to make the prediction using these features.

Cost Function: Write an equation that describes how good or bad your model is doing.

Minimize cost function: Write an algorithm to minimize/maximize your cost function, preferably provably so.

Generalization: Check that you have not overfitted or underfitted the data, e.g. the football predicting octopus.

Techniques

Techniques covered in ECS708 include:

- Probability and Random Variables
- Neural Networks
- Bayesian Inference

- Clustering
- Hidden Markov Models (HMMs)
- Principal Components Analysis, Independent Component Analysis

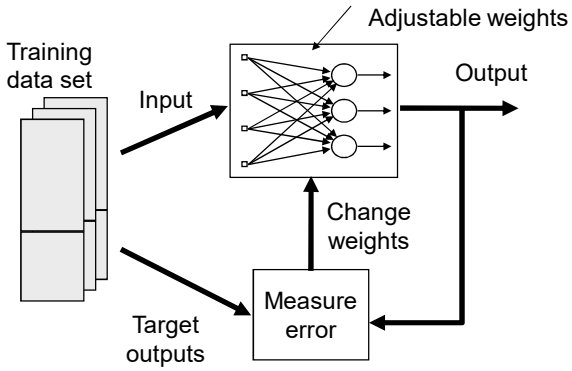
Probability and Random Variables

- Probability of an event (set of possible outcomes):
  $P(A) \geq 0$
  $P(\Omega) = 1$
  $P(A \cup B) = P(A) + P(B)$ if A, B mutually exclusive
- Random Variables $X : \Omega \rightarrow IR$
  - Distribution function $F_X(x) = P(X \leq x)$
  - Density function (pdf) $f_X(x) = \frac{d}{dx} F_X(x)$

    (More familiar as $p(x)$)

Neural Networks

Adjust “weight” parameters in a network of simple calculating units to produce desired behaviour.



Statistical Inference

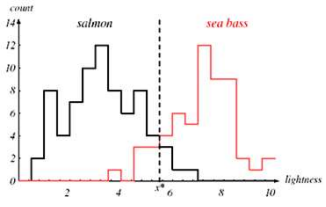
Use of probability theory to estimate the “best” answer

Uses estimate of prob. of observation $\mathbf{x}$ given classes ω_i

$$p(x|\omega_i)$$

1) Maximum Likelihood – Choose $\max_i p(x|\omega_i)$

E.g. Classification of Fish:



2) Use Bayes Theorem:

$$p(\omega_i|x) = \frac{p(x|\omega_i)p(\omega_i)}{p(x)}$$

Choose $\max_i p(\omega_i|x)$

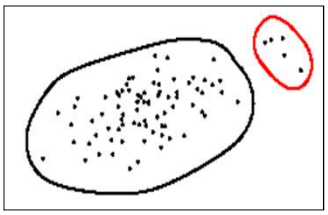
[After all, we have x and want to know ω_i]

Clustering

- Collecting together “similar” observations or signals.
- E.g. cluster similar documents, music, bacteria,…

Issues:

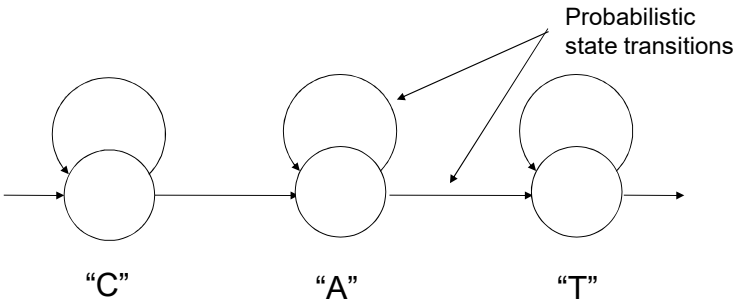
- Similarity measures to use?
- Same / different sized clusters?
- How many clusters?



Hidden Markov Models (HMMs)

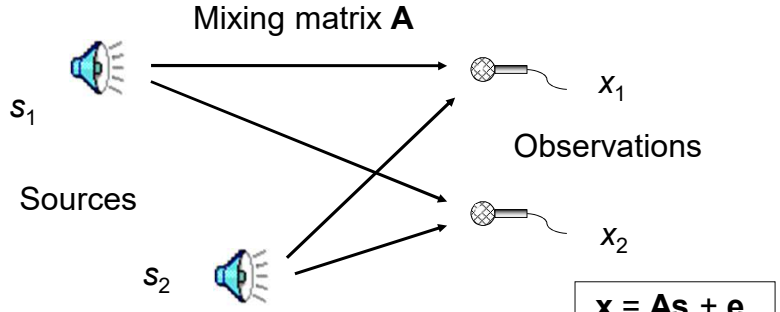
Hidden Markov Models (HMMs)

- Statistical model of speech utterances
- We don’t “see” the states (“hidden”), only their output
- Choose the most probable word given the utterance



Independent Component Analysis

Blind Sources Separation - want to “unmix” observed signals that contain more than one source signal – the “Cocktail Party Problem”



$\mathbf{x} = \mathbf{As} + \mathbf{e}$

noise

n sources, $m \geq n$ observations