

MACHINE LEARNING

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

ECS708 Machine Learning

Slide no: 1

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

ECS708 Machine Learning

Slide no: 2

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

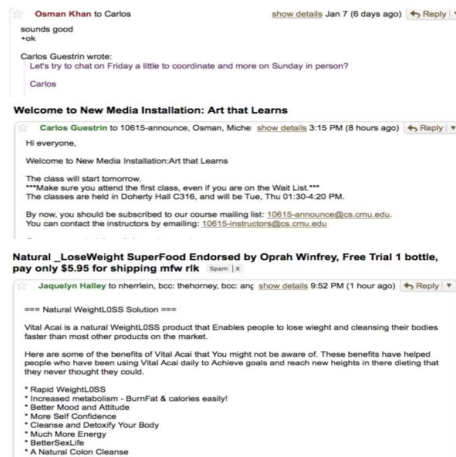
A Few Quotes

- “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

ECS708 Machine Learning

Slide no: 9

Stock market prediction



ECS708 Machine Learning

Slide no: 10

Facial landmark localisation



<https://www.youtube.com/watch?v=ONnabin5GBs>

ECS708 Machine Learning

Slide no: 11

Machine Learning by examples

Clustering

ECS708 Machine Learning

Slide no: 12

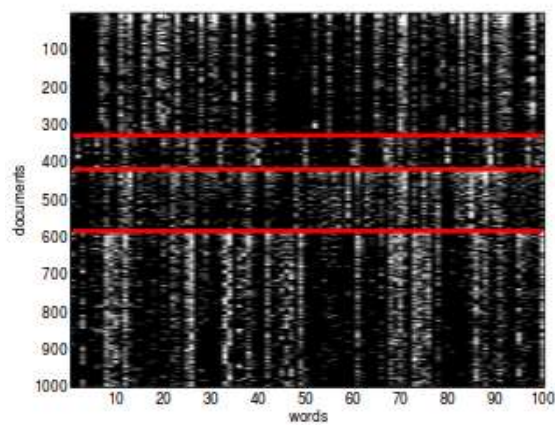
Clustering Images

Clustering images



[Goldberger et al.]

Clustering documents



Machine Learning by examples

Ranking

ECS708 Machine Learning

Slide no: 15

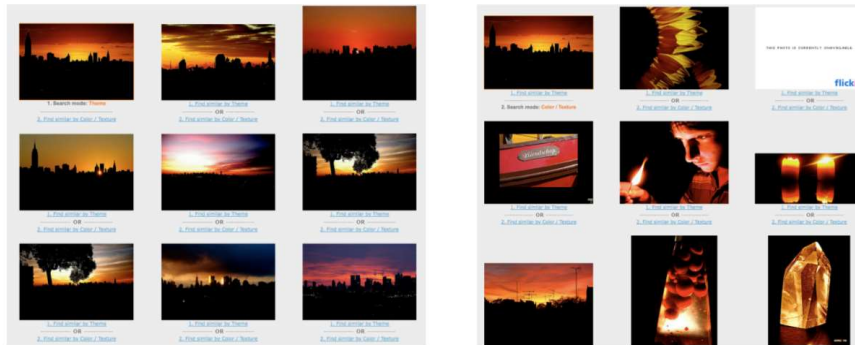
Text-based search

The screenshot shows a Google search interface. The search bar contains the text "learning to rank". Below the search bar, a dropdown menu displays several suggestions: "learning to rank", "learning to rank for information retrieval", "learning to rank using gradient descent", and "learning to rank tutorial". To the right of the search bar is a blue "I'm Feeling Lucky" button. On the left side of the page, there is a sidebar with navigation links: "Search", "Web", "Images", "Maps", "Videos", "News", "Shopping", and "More". Below these links, there is a location setting for "Manhattan, NY 10012" and a "Change location" link. The main content area displays search results. The first result is "Learning to rank - Wikipedia, the free encyclopedia" with a link to "en.wikipedia.org/wiki/Learning_to_rank". The second result is "Learning to rank or machine-learned ranking (MLR) is a type of supervised or semi-supervised machine learning problem in which the goal is to automatically ..." with links to "Applications", "Feature vectors", "Evaluation measures", and "Approaches". The third result is "Yahoo! Learning to Rank Challenge" with a link to "learningtorankchallenge.yahoo.com/". The fourth result is "Learning to Rank Challenge is closed! Close competition, innovative ideas, and fierce determination were some of the highlights of the first ever Yahoo!". The fifth result is "PDF Large Scale Learning to Rank" with a link to "www.eecs.tufts.edu/~dsculley/papers/large-scale-rank.pdf". The sixth result is "File Format: PDF/Adobe Acrobat - Quick View" by D Sculley - Cited by 24 - Related articles. The seventh result is "Pairwise learning to rank methods such as RankSVM give good performance, ... In this paper, we are concerned with learning to rank methods that can learn on ...". The eighth result is "Microsoft Learning to Rank Datasets - Microsoft Research" with a link to "research.microsoft.com/en-us/projects/malr/". The ninth result is "We release two large scale datasets for research on learning to rank: L2R-WEB30k with more than 30000 queries and a random sampling of it L2R-WEB10K ...". The tenth result is "LETOR: A Benchmark Collection for Research on Learning to Rank ..." with a link to "research.microsoft.com/~letor/". The eleventh result is "This website is designed to facilitate research in LEarning TO Rank (LETOR). Much information about learning to rank can be found in the website, including ...".

ECS708 Machine Learning

Slide no: 16

Visual search – find similar images



<http://www.tiltomo.com/>

ECS708 Machine Learning

Slide no: 17

Machine Learning by examples

Recommendation

ECS708 Machine Learning

Slide no: 18

Recommendation



ECS708 Machine Learning

Slide no: 19

Recommendation systems

	← users →				
↑ movies ↓	1	?	3	5	?
	?	1			2
		4	4	5	?

Netflix competition (www.netflixprize.com)

Machine learning competition with \$1m prize

ECS708 Machine Learning

Slide no: 20

Types of Learning Problems

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

ECS708 Machine Learning

Slide no: 21

Supervised Learning: learn a prediction function

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

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

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):

$$P(A) \geq 0$$

$$P(\Omega) = 1$$

$$P(A \cup B) = P(A) + P(B) \quad \text{if } A, B \text{ mutually exclusive}$$

- Random Variables $X: \Omega \rightarrow \mathbb{R}$

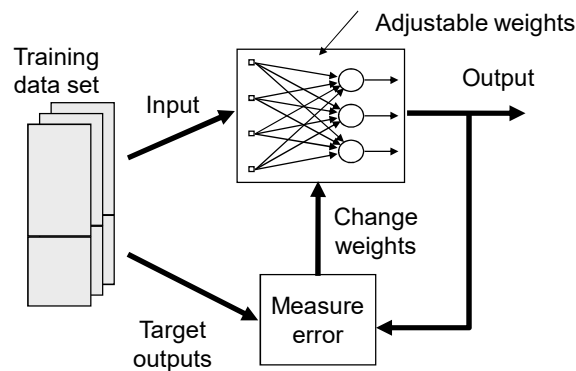
- Distribution function $F_X(x) = P(X \leq x)$

- 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.



ECS708 Machine Learning

Slide no: 33

Statistical Inference

Use of probability theory to estimate the “best” answer

Uses estimate of prob. of observation x given classes ω_i

$$p(x|\omega_i)$$

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

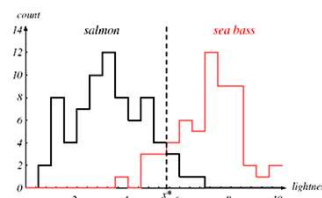
E.g. Classification of Fish:

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]



ECS708 Machine Learning

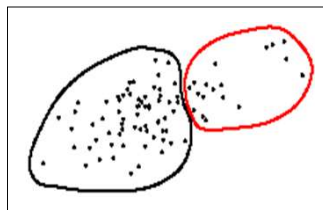
Slide no: 34

Clustering

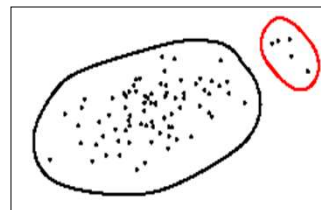
- Collecting together “similar” observations or signals.
- E.g. cluster similar documents, music, bacteria,...

Issues:

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



ECS708 Machine Learning

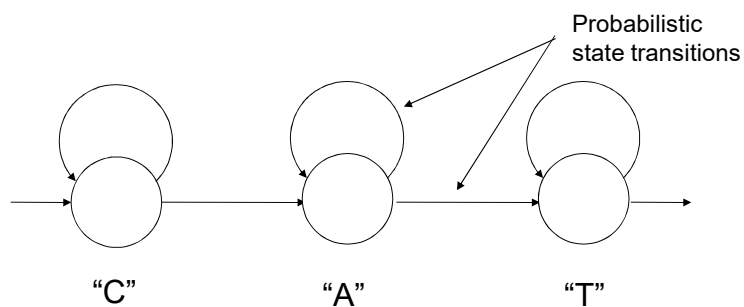


Slide no: 35

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



ECS708 Machine Learning

Slide no: 36

Independent Component Analysis

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

