Introduction

Machine Learning and Pattern Recognition

Chris Williams

School of Informatics, University of Edinburgh

August 2014

(All of the slides in this course have been adapted from
previous versions by Charles Sutton, Amos Storkey, David Barber.)

What is Machine Learning?

- It's about finding patterns in data, and using the patterns to make predictions
- There are lots of problems where
  - We'd like to be able to solve them with a computer
  - We don't know how to write a computer program to solve them
  - We can collect examples
- Let's look at some example problems ...

Example 1: Categorizing Documents

Input: Text of Document

Tory fury as Lib Dem peers join Labour to delay boundary review

Review delayed for five years, seriously damaging David Cameron's chances of winning overall majority in 2015

Patrick Wintour, political editor
The Guardian, Monday 14 January 2013 19.45 GMT
Jump to comments (120)

David Cameron's chances of winning an overall majority in 2015 have been seriously damaged by the delay to the boundary review. Photograph: Kerim Okten/EPA

Coalition relations plummeted on Monday when the Liberal Democrats were accused by Conservatives of double crossing, cynicism, cheating and opportunism as Nick Clegg's peers joined Labour to delay a constituency boundary review that had been likely to gift the Tories 20 extra seats.

The review will now be delayed for five years, leaving the next election to be fought on the existing constituency boundaries, so seriously

NFL playoffs: Colin Kaepernick and Ray Lewis keep Super Bowl dreams alive

An exhilarating weekend of NFL playoffs saw Peyton Manning, Russell Wilson, Aaron Rodgers and Rob Gronkowski depart; while Ray Lewis, Colin Kaepernick, and Matt Ryan go through

Golden Tate celebrates a Seattle touchdown as the Seahawks contributed to a helter skelter weekend of NFL action. Photograph: Mike Ehrmann/Getty Images

John Fox the real villain in Denver's defeat

In the heady moments that followed Baltimore's double overtime victory over the Broncos on Saturday, Ray Lewis cut to the chase: "When all the emotions calm down," he said, "this will probably go down as one of the greatest victories in Ravens history."

If that is not such a big claim to make for a team which is still just 16 years old, then the truth is that this game will be remembered far outside Baltimore's city limits. It might just merit consideration in among the league's best-ever playoff games. (On what certainly felt like one of the best-ever weekends.) At 4hrs 11mins, it was the longest NFL game since 1987 and there were not a lot of dull moments in that time.

Label:	Politics	Sports

Example 1: Categorizing Documents

- Make a list of sport terms?
- What about this:

  Supreme court ruling: Medicaid expansion becomes political football

  States may opt out of a programme offering health coverage to 16 million of America's poorest after supreme court ruling

  Chris McGreal and **Richard Adams**
  guardian.co.uk, Friday 29 June 2012 14.23 BST
  Jump to comments (3)

  Protesters outside the supreme court in Washington DC. Photograph: Mark Wilson/Getty Images

  While the supreme court's decision to uphold the Obama administration's individual mandate took the headlines, the court's

- Even this simple task is a bit more complicated.

Example 2: Handwriting Recognition

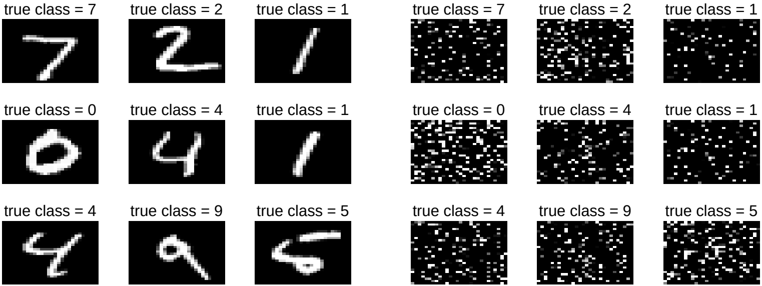


Figure credit: Murphy Fig. 1.5

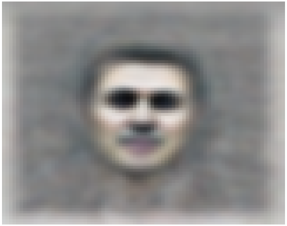
- This is deployed! All cheques and handwritten envelopes are scanned automatically
- Lots of other computer vision problems raise similar issues (but are harder to solve)

Example 3: Find Some Patterns

Input: 10 million images from YouTube videos

Researchers from Stanford and Google (Le et al, ICML 2012)

Output: Find something interesting.

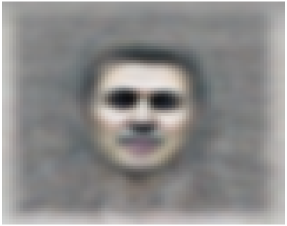


Example 3: Find Some Patterns

Input: 10 million images from YouTube videos

Researchers from Stanford and Google (Le et al, ICML 2012)

Output: Find something interesting.



Cats!

And More Applications ...

- Computer vision: Face detection, object recognition, scene understanding
- Speech processing and generation
- Collaborative filtering: Predict how much I will like a book / movie
- Computational advertising: Predict whether I will click an ad
- Bioinformatics: Identify which regions of DNA encode proteins
- Scientific Applications: Find galaxies in images, Model cellular chemical processes
- Robotics: Learn a map of a building as a robot explores it
- Natural language processing: Syntactic parsing, building a database from text, Web search

Why Machine Learning?

Exciting area of endeavour

- Data is everywhere, and growing.
- ML combines (some) theoretical foundations with (many) practical problems
- Ubiquitous in AI problems (computer vision, language modelling, speech modelling, handwriting recognition)
- Growing demand outside of AI (risk management, characterising historical artefacts, medical imaging, web analytics, recommender engines, computer games engines, financial modelling, geoinformational systems, intelligent management, operational research, etc. etc. etc.)
- Machine learning skills are in high demand
- Buzzwords: “big data”, “analytics”, “data science”

Outline

- Different Types of Learning Problems
- The Model and the Algorithm
- Probabilities in Machine Learning
- Feature Vectors
- The need for assumptions/models
- Course Outline

Different Types of Learning Problems

- Supervised learning
  - Classification
  - Regression
- Unsupervised Learning
  - Clustering
  - Discovering latent factors
  - many others (see Chapter 1, Murphy)

Supervised Learning

Given dataset $\mathcal{D} = \{(\mathbf{x}_i, \mathbf{y}_i), i = 1, 2, \ldots, N\}$, learn a predictor that given a new $\mathbf{x}^*$ makes a useful statement about the associated $\mathbf{y}^*$.

Unsupervised Learning

Given dataset $\mathcal{D} = \{\mathbf{x}_i, i = 1, 2, \ldots, N\}$, find some interesting patterns in the data set.

Examples of unsupervised learning methods:

- Clustering
- Dimensionality reduction (will explain this later)
- Association rule learning (won't explain this; take DME)

“Principled” Machine Learning

- IAML gives you a toolbag of algorithms
- This course focuses on a *principled* and *probabilistic* view of ML
- What does it means to have principles (in ML)?

By “principles” I mean a theoretical framework that helps you to

- Understand what assumptions a learning algorithm makes
- Understand similarities and differences between algorithms
- Derive custom models and algorithms for a new learning task

The Model and the Algorithm

- Model encodes understanding about the data. Process of learning from data. (e.g., a set of probability distributions $p(y|\mathbf{x})$)
- Algorithm comes from the model, causing us to select a distribution from the set. Or multiple distributions!
- Different algorithms give different approximations

Probabilities in Machine Learning

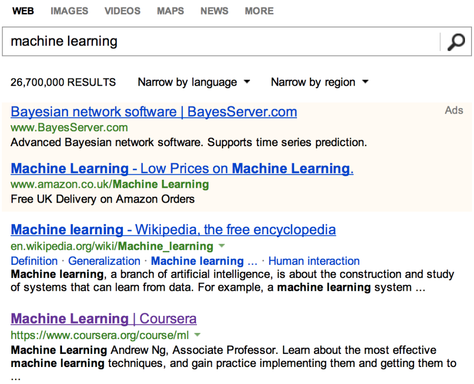
- Consider document classification again. Let $\mathbf{x}$ denote the document, and y the label. $y \in \{$"Sports", "Politics"$\}$
- You write a function f in Java that takes $\mathbf{x}$ and returns y
- Suppose I pay you £1000 for every politics article you get right, and £1M for every sports article you get right.[1] How do you modify f?

[1] Important clarification: I am not actually going to do this

Probabilities in Machine Learning

- Consider document classification again. Let $\mathbf{x}$ denote the document, and y the label. $y \in \{$"Sports", "Politics"$\}$
- You write a function f in Java that takes $\mathbf{x}$ and returns y
- Suppose I pay you £1000 for every politics article you get right, and £1M for every sports article you get right.[1] How do you modify f?
- Or to make things more complicated, suppose I also charge you £10000 for every one you get wrong. Now what do you do?
- One answer: Don't write a function. Specify a probability distribution $p(y|\mathbf{x})$ Then you can make decisions by maximizing the expected profit
- This situation happens in real life ...

[1] Important clarification: I am not actually going to do this



To choose ads

- Estimate clickthrough rate
- Look up what advertisers have bid
- Show ads with high expected value

If you want to act based on your predictions, it helps to know uncertainty.

Feature Vectors

We (usually) represent the input as a vector $\mathbf{x} \in \mathbb{R}^D$.
This is called a *feature vector*.
Each element x_i for $i \in \{1 \dots D\}$ is called a feature.
Examples:

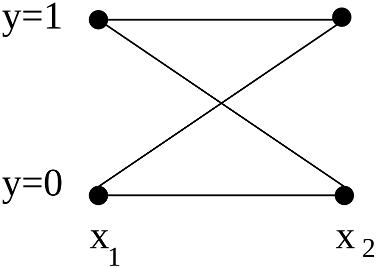
Documents

Let $(w_1, w_2, \dots w_V)$ be a dictionary of English,
e.g, $w_1 =$ "aardvark", $w_2 =$ "apple".
x_i the number of times that word w_i appears in document (bag of words representation)

Images

Suppose the image is $m \times m$ pixels, black and white.
Let $D = m^2$. Order pixels from 1 to D (e.g. raster scan).
Let $x_i \in \{0, 1, \dots 255\}$ be the greyscale value of the pixel i

The Need for Assumptions/Models



- Two input locations, x_1 and x_2, binary classification problem
- Suppose we know $y(x_1) = 1$, what does this tell us about $y(x_2)$?
- With no assumptions it tells us nothing ...
- "A learner that makes no a priori assumptions regarding the target concept has no rational basis for classifying any unseen instances" (Mitchell, 1997)

- Assumptions are sometimes known as inductive bias
- No free lunch theorem (Wolpert, 1996): there is no universally best model
- All learning algorithms make prior assumptions. Anyone who tells you otherwise is selling you something.

Course Outline

- Statistical Fundamentals
  - Probability, Data and models, Bayesian methods, maximum likelihood, exponential family
- Supervised Learning
  - Linear and nonlinear regression, logistic regression, neural networks
- Unsupervised Learning
  - Dimensionality reduction, expectation maximization
- Computational Issues in Probability Distributions
  - Optimization, Variational inference, Markov chain Monte Carlo
- Advanced Topics (if time)
  - Deep learning, Gaussian processes

What is the Point of Studying this Course?

What should you be able to do after this course?

- Understand why and how it is possible to do machine learning
- Understand how the wide set of machine learning methods fit into an overall framework
- Know how to use and justify these methods
- Be able to create your own machine learning methods
- Learn to think in terms of probabilistic models

Summary

- Machine learning is ubiquitous and useful
- Theoretical grounding helps us understand algorithms and generate new ones.
- No free lunch
- Models not algorithms
- Probabilistic view

Actions

Attending lectures is no substitute for working through the material! Lectures will motivate the methods and approaches. Only by study of the notes and bookwork will the details be clear. If you do not understand the notes then discuss them with one another. Ask your tutors.

Reading

These lecture slides. Chapter 1 of Murphy.