MLPR Introduction

Machine Learning and Pattern Recognition

Amos Storkey

School of Informatics, University of Edinburgh

Outline

Welcome

to the Machine Learning and Pattern Recognition Course

What is this about?

- Probabilistic Approaches to Supervised Machine Learning

One course among many:

- Introductory Applied Machine Learning
- Probabilistic Modelling and Reasoning
- Information Theory
- Reinforcement Learning
- Data Mining and Exploration
- Neural Information Processing

Just a few courses that are part of...

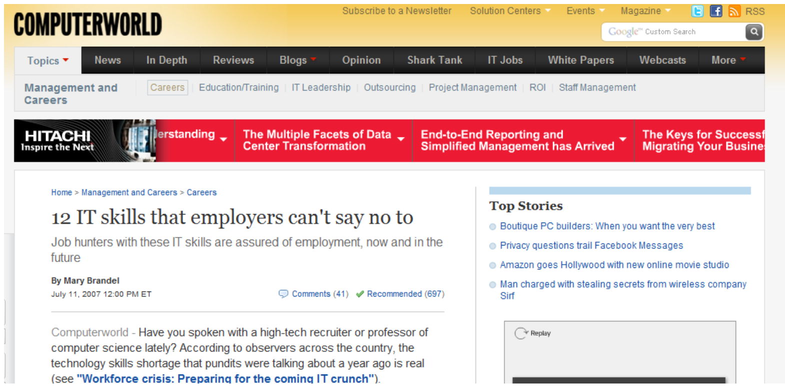
MSc (Master of Science) in Informatics and MSc in Artificial Intelligence at the School of Informatics, University of Edinburgh.

Why?

Exciting area of endeavour

- Coherent.
- Interesting (and some unsolved...) problems.
- Now ubiquitous...
- Relevant to brain science, the scientific method, image analysis and computer vision, language modelling, speech modelling, handwriting recognition, 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.
- In great demand.

12 IT Skills Employers Can't Get Enough Of


COMPUTERWORLD

Home > Management and Careers > Careers

12 IT skills that employers can't say no to

Job hunters with these IT skills are assured of employment, now and in the future

By Mary Brandel

July 11, 2007 12:00 PM ET

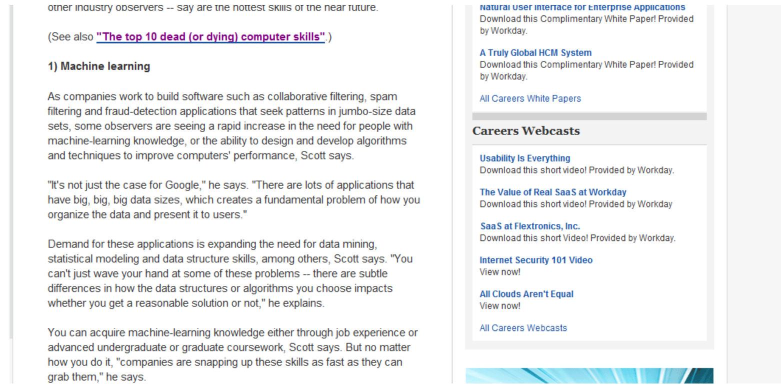
Computerworld - Have you spoken with a high-tech recruiter or professor of computer science lately? According to observers across the country, the technology skills shortage that pundits were talking about a year ago is real (see "Workforce crisis: Preparing for the coming IT crunch").

Top Stories

- Boutique PC builders: When you want the very best
- Privacy questions trail Facebook Messages
- Amazon goes Hollywood with new online movie studio
- Man charged with stealing secrets from wireless company Sirf

So what is number one?

12 IT Skills Employers Can't Get Enough Of


other industry observers -- say are the hottest skills of the near future.

(See also "The top 10 dead (or dying) computer skills".)

1) Machine learning

As companies work to build software such as collaborative filtering, spam filtering and fraud-detection applications that seek patterns in jumbo-size data sets, some observers are seeing a rapid increase in the need for people with machine-learning knowledge, or the ability to design and develop algorithms and techniques to improve computers' performance, Scott says.

"It's not just the case for Google," he says. "There are lots of applications that have big, big, big data sizes, which creates a fundamental problem of how you organize the data and present it to users."

Demand for these applications is expanding the need for data mining, statistical modeling and data structure skills, among others, Scott says. "You can't just wave your hand at some of these problems -- there are subtle differences in how the data structures or algorithms you choose impacts whether you get a reasonable solution or not," he explains.

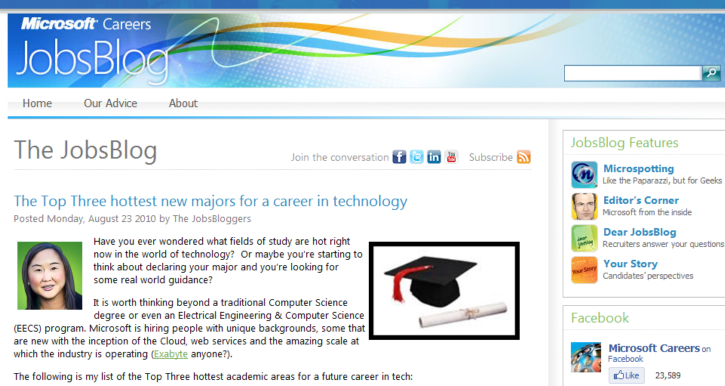
You can acquire machine-learning knowledge either through job experience or advanced undergraduate or graduate coursework, Scott says. But no matter how you do it, "companies are snapping up these skills as fast as they can grab them," he says.

Careers Webcasts

- Usability Is Everything
- The Value of Real SaaS at Workday
- SaaS at Flextronics, Inc.
- Internet Security 101 Video
- All Clouds Aren't Equal

- Machine Learning.
- But that was 2007. What about more recently...

Top 3 Hottest Majors for a Career in Technology


Microsoft Careers JobsBlog

The JobsBlog

The Top Three hottest new majors for a career in technology

Posted Monday, August 23 2010 by The JobsBloggers

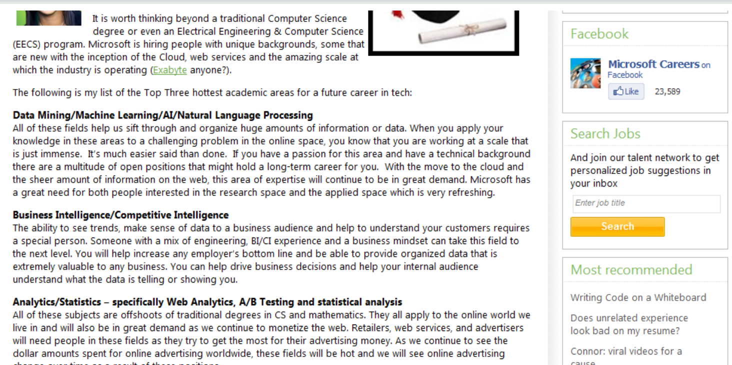
Have you ever wondered what fields of study are hot right now in the world of technology? Or maybe you're starting to think about declaring your major and you're looking for some real world guidance?

It is worth thinking beyond a traditional Computer Science degree or even an Electrical Engineering & Computer Science (EECS) program. Microsoft is hiring people with unique backgrounds, some that are new with the inception of the Cloud, web services and the amazing scale at which the industry is operating (Exabyte anyone?).

The following is my list of the Top Three hottest academic areas for a future career in tech:

Now we are in summer 2010. The Top 3 hottest majors for a career in technology are...

Top 3 Hottest Majors for a Career in Technology


The following is my list of the Top Three hottest academic areas for a future career in tech:

Data Mining/Machine Learning/AI/Natural Language Processing

All of these fields help us sift through and organize huge amounts of information or data. When you apply your knowledge in these areas to a challenging problem in the online space, you know that you are working at a scale that is just immense. It's much easier said than done. If you have a passion for this area and have a technical background there are a multitude of open positions that might hold a long-term career for you. With the move to the cloud and the sheer amount of information on the web, this area of expertise will continue to be in great demand. Microsoft has a great need for both people interested in the research space and the applied space which is very refreshing.

Business Intelligence/Competitive Intelligence

The ability to see trends, make sense of data to a business audience and help to understand your customers requires a special person. Someone with a mix of engineering, BI/CI experience and a business mindset can take this field to the next level. You will help increase any employer's bottom line and be able to provide organized data that is extremely valuable to any business. You can help drive business decisions and help your internal audience understand what the data is telling or showing you.

Analytics/Statistics – specifically Web Analytics, A/B Testing and statistical analysis

All of these subjects are offshoots of traditional degrees in CS and mathematics. They all apply to the online world we live in and will also be in great demand as we continue to monetize the web. Retailers, web services, and advertisers will need people in these fields as they try to get the most for their advertising money. As we continue to see the dollar amounts spent for online advertising worldwide, these fields will be hot and we will see online advertising change over time as a result of these positions.

- No surprises there[1].
- Machine learning skills *are* in high demand.

[1](Selection bias caveat)

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.

So What is Machine Learning?

A one liner?

Problem 1

Problem 2

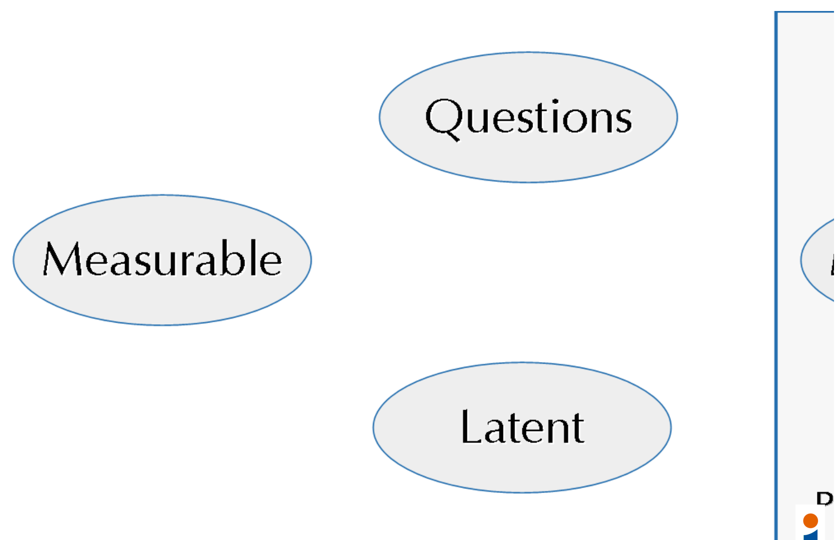
Problem 3

Knock, knock. Whos there? Amos. Amos who? A mosquito.

So What is Machine Learning?

- One Answer...
  - Machine learning is performing tasks that are difficult to program explicitly by:
  - Defining the problem task.
  - Providing a general model for the problem.
  - Using historic data for which the task is done to help refine the model to enable better performance at the task.
  - Using methods to go from the refined model to making predictions in new situations.

Probabilistic Machine Learning



Example: Single Image Super-Resolution

Example

- Given lots of example images...
- Learn how to fill in a high resolution image given a single low resolution one.
- State of the art is hard to beat, and appears to be in widespread use.
- However state of art technology appears to be restricted to a small region in Los Angeles area.
- Technology used unknown but seems particularly pertinent at discovering reflections of people in windows, and is usually accompanied by series of short beeps in quick procession.

Example: Single Image Super-Resolution

Let's Enhance

./Figures/LetsEnhance.mp4

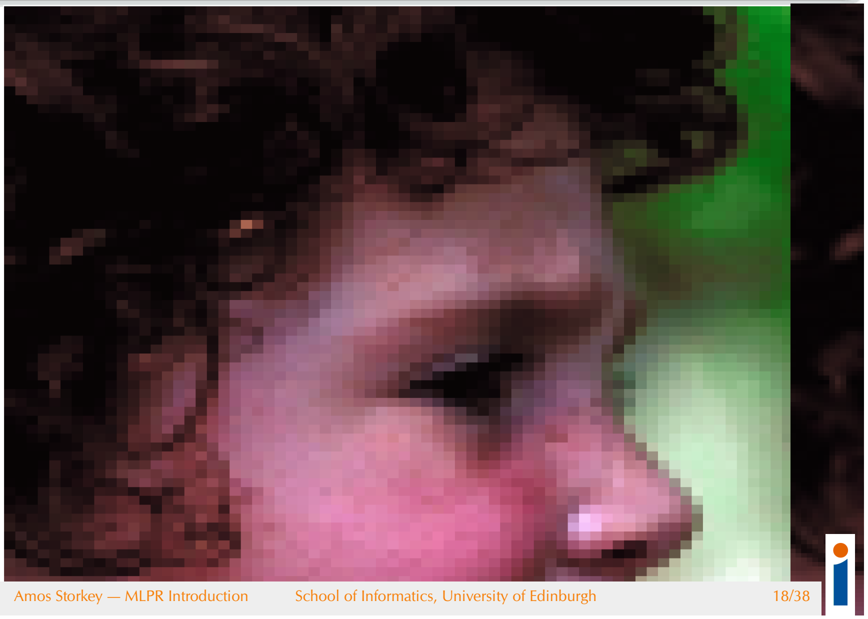
Anncknowledgements: Duncan Robson

Example: Single Image Super-Resolution

Example

- Rest of us stuck with more standard approaches.
- One Approach:
- Build model for image patches. Build model for map from high to low dimensional space.
- Refine model to fit example data.
- Invert model to give high dimensional region for low dimensional data.

Example: Single Image Super-Resolution

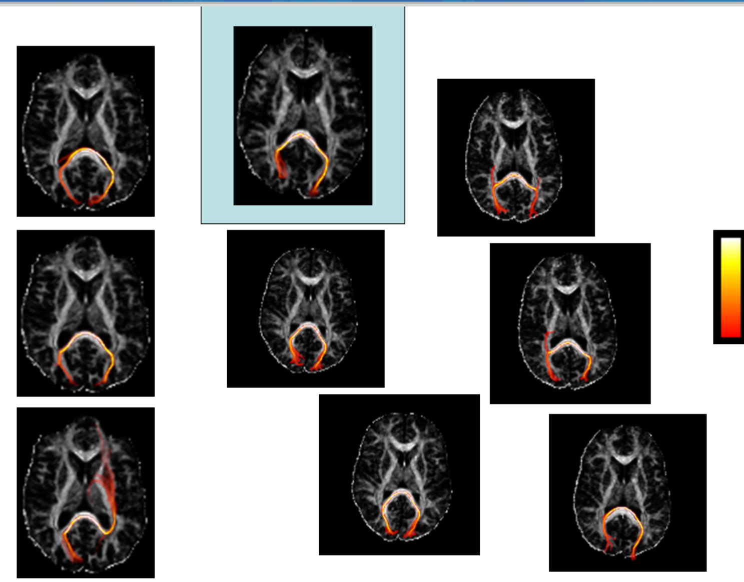


Example 2: Discovering Matches

Example

- Suppose we have to work out if two things match.
- Images of faces... but many other applications:
  - e.g. Tracts in Brains
  - or Authors of papers
  - How can we do this?
- Use example data to learn models for matched and unmatched objects pairs. Combine with prior. Use to score.

Example 2: Discovering Matches



Thinking about Data

- Computer science and algorithms:

Algorithms

- Simplifying in the extreme: View computer science as algorithm generation.
- If the algorithm works it is good. If it doesn't it is bad.

Machine Learning: the Algorithm and the Model

- Model encodes understanding about the data. Process of learning from data.
- Algorithm comes from the model (and a bit of maths describing the process).
- Different algorithms give different approximations.

Example

Example

3 Boolean variables. Data set

$$\begin{array}{ccc} 1 & 0 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & x \end{array}$$

- What is x?

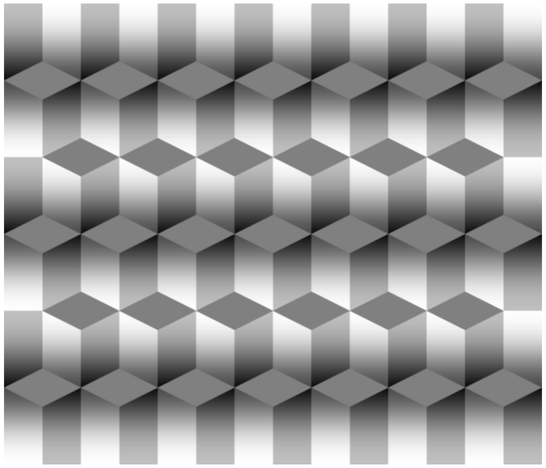
Thinking about Data

- Machine learning does not give us something for nothing.
- Prior beliefs and model + data $\rightarrow$ posterior beliefs.
- Can do nothing without some a priori model - no connection between data and question.
- A priori model sometimes called the inductive bias.

Illusions

- Logvinenko illusion
- Inverted mask illusion

Logvinenko Illusion



Inverted Mask Illusion

Dragon Movie

./Figures/dragonmovie.mpg

Example

Example

3 Boolean variables. Data set

1	0	1
1	1	1
1	1	0
0	1	0
0	0	x

- What is x?
- We cannot say. We have no information at all about how any of these data items is connected.
- "No free lunch"

No Free Lunch

Example

You wish to predict binary variable C, which you believe is deterministically dependent on binary A and binary B. Consider all the possible hypotheses.

1. What is the total size of the hypothesis space?
2. Suppose you observe $C = 1$ for $A = 0$ and $B = 1$, and $C = 0$ for $A = 1$ and $B = 1$. What is the size of the space of remaining valid hypotheses.
3. Given the above data, purely on the basis of the proportion of supporting hypotheses, what would you predict for C given $A = 1$ and $B = 0$?
4. Assume now that you know the true hypothesis must be one of ($C = A$ OR B) OR ($C = A$ AND B) OR ($C = A$ XOR B). What would you now predict for C given $A = 1$ and $B = 0$?

No Free Lunch

Problem

- Predict $C \in \{0, 1\}$ from $A, B \in \{0, 1\}$.
- No noise - given A,B then C is always the same.

Hypotheses

Possible hypotheses: represent each by set S of (A, B) pairs for which $C = 1$. For the rest $C = 0$.

Example

One hypothesis is $\{(1, 1), (0, 1), (1, 0)\}$ (i.e. $C = A$ OR B). Here $(0, 1)$ means $A = 0$, $B = 1$.

No Free Lunch contd.

- No bias implies ${}^4C_0 + {}^4C_1 + {}^4C_2 + {}^4C_3 + {}^4C_4 = 16$ equally possible hypotheses - the hypothesis space
- Each datum reduces the size of the hypothesis space

Predict C given unseen values of A,B?

Number of hypotheses predicting $C = 1$ equals the number predicting $C = 0$.

Example

- Data: $(A = 1, B = 1, C = 0)$, $(A = 0, B = 1, C = 1)$. What is hypothesis space now?

No Free Lunch contd.

Example

- Data: $(A=1, B=1, C=0)$, $(A=0, B=1, C=1)$. Hypothesis space is now: $\{(0,1)\}$, $\{(0,1),(1,0)\}$, $\{(0,1),(0,0)\}$, $\{(0,1),(1,0),(0,0)\}$.
- If we query $(A=1, B=0)$, of valid hypotheses, two predict $C=1$, and two predict $C=0$.
- Suppose we now see data $(a=0, b=0, c=0)$. Hypothesis space is $\{(0,1)\}$, $\{(0,1),(1,0)\}$.
- One of the remaining hypotheses predicts $C=1$, and the other predicts $C=0$.
- Regardless of data or (unseen) query the number of hypotheses predicting each will be the same.

Prior Assumption

- Prior assumption: say something about the ways C is allowed to relate to A, B.

Example

$(C = A \text{ OR } B)$ OR $(C = A \text{ AND } B)$ OR $(C = A \text{ XOR } B)$.

- Data $(A=1, B=0, C=1)$, $(A=0, B=1, C=1)$
- So either OR or XOR. Can predict $(a=0, b=0, c=?)$.

Suspect Terms

- Model Free.
- Bias Free. Unbiased. (In a general not statistical sense).
- No prior assumption.
- Generally applicable.

Machine Learning and Probability

- Prior assumptions are encoded probabilistically via
  - The model
  - The prior distribution (more on this later)
- Non-trivial:
  - Hard to really understand the full implications of a probability distribution.
  - Hard to accurately represent your prior beliefs, and represent them in a way that is amenable to computation.

Supervised Learning

- Supervised Learning (Unsupervised Learning)
- Describes the type of *task*

Supervised Learning

Given dataset $\mathcal{D} = \{(\mathbf{x}^n, \mathbf{y}^n), n = 1, 2, \ldots, N\}$, the supervised learning task is to learn, from that dataset, about the unknown relationship between $\mathbf{x}$ and $\mathbf{y}$, such that given some new $\mathbf{x}^*$ you can make some useful and accurate statement about the the associated $\mathbf{y}^*$ under that relationship.

- Useful and accurate? That's a bit vague?

Inference and Decision

The statement is typically about the probability of $\mathbf{y}^*$ or about the (expected) utility of taking some action dependent on the value of $\mathbf{y}^*$.

Summary

- Course primarily on Supervised Learning.
- Ubiquitous and useful.
- Theoretical grounding is key.
- No free lunch.
- Models not algorithms.
- Probability.

To Do

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 Barber.

Preparatory Reading

Bishop Chapter 1.

Extra Reading

Cox's Axioms. Subjective Probability.