



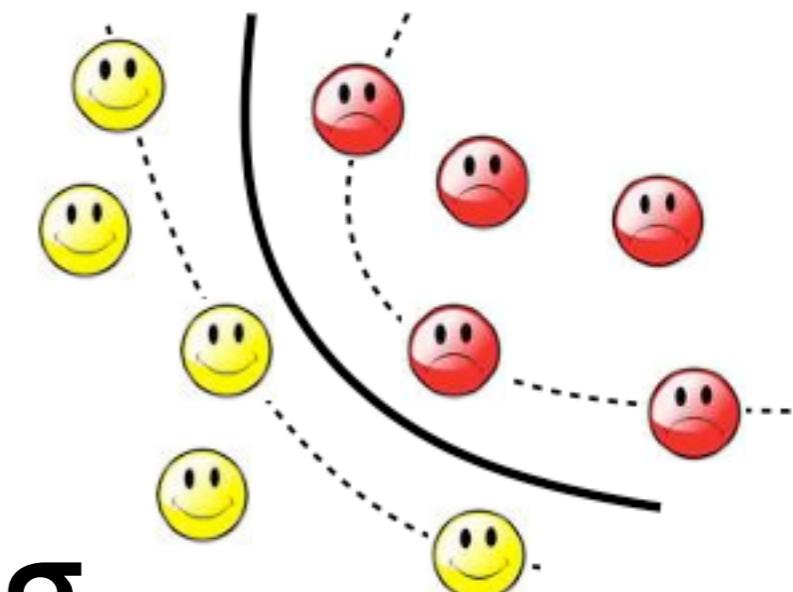
THE UNIVERSITY OF  
**SYDNEY**

# Machine Learning and Data Mining

(COMP 5318)

School of Information Technologies

A/Prof Fabio Ramos  
Dr Lionel Ott  
Dr Roman Marchant





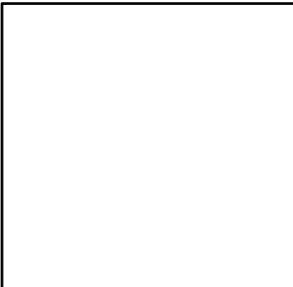
THE UNIVERSITY OF  
**SYDNEY**

# WHS INDUCTION

School of Information Technologies



THE UNIVERSITY OF  
SYDNEY



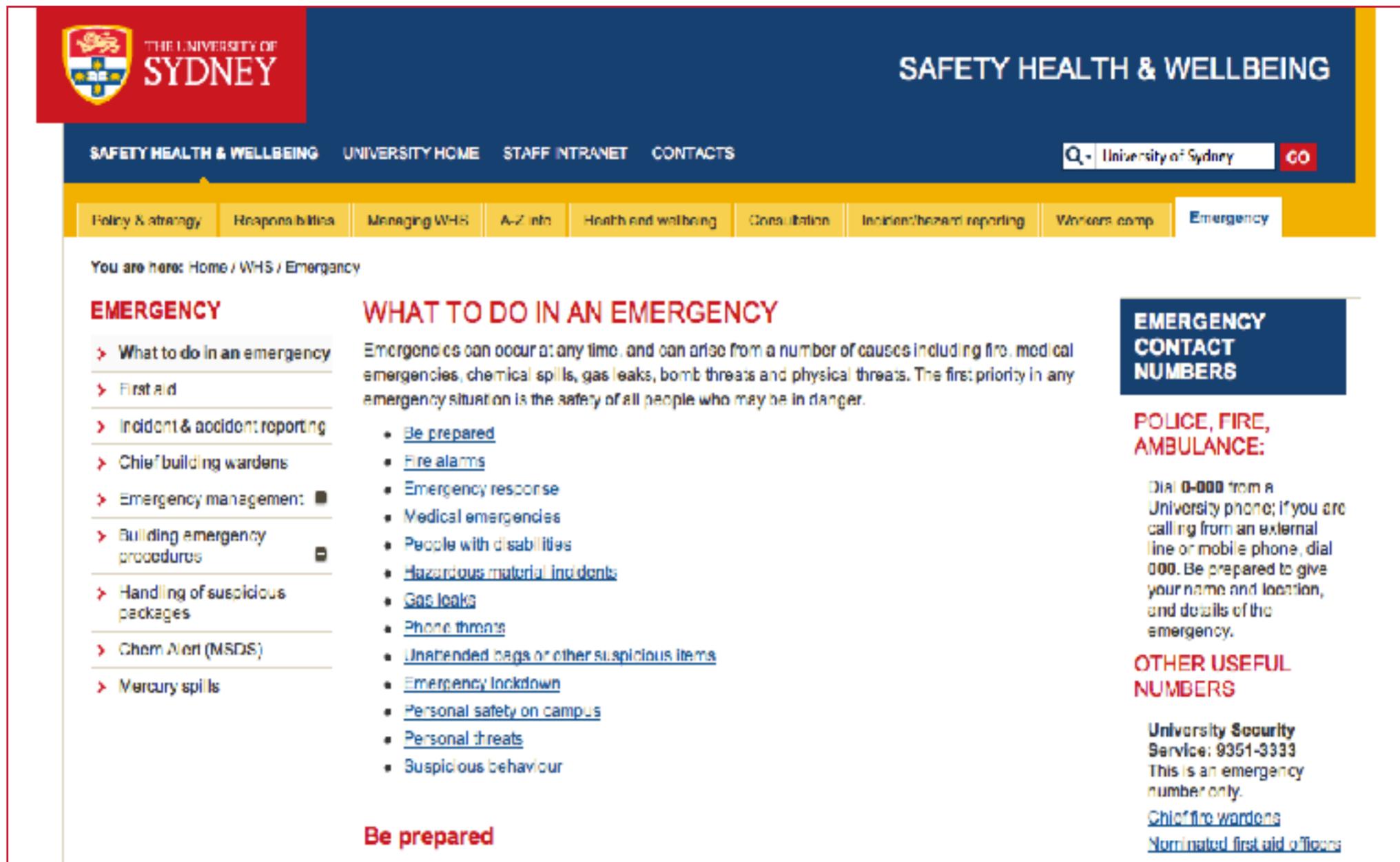
# General Housekeeping – Use of Labs

- › Keep work area clean and orderly
- › Remove trip hazards around desk area
- › No food and drink near machines
- › No smoking permitted within University buildings
- › Do not unplug or move equipment without permission



# EMERGENCIES – Be prepared

→ [www.sydney.edu.au/whs/emergency](http://www.sydney.edu.au/whs/emergency)



The screenshot shows the University of Sydney's Safety Health & Wellbeing website. The top navigation bar includes links for SAFETY HEALTH & WELLBEING, UNIVERSITY HOME, STAFF INTRANET, and CONTACTS. A search bar and a red 'GO' button are also present. Below the navigation, a secondary menu bar features links for Policy & strategy, Responsibilities, Managing WHS, A-Z info, Health and wellbeing, Consultation, Incident/hazard reporting, Workers comp, and Emergency. The main content area displays the following sections:

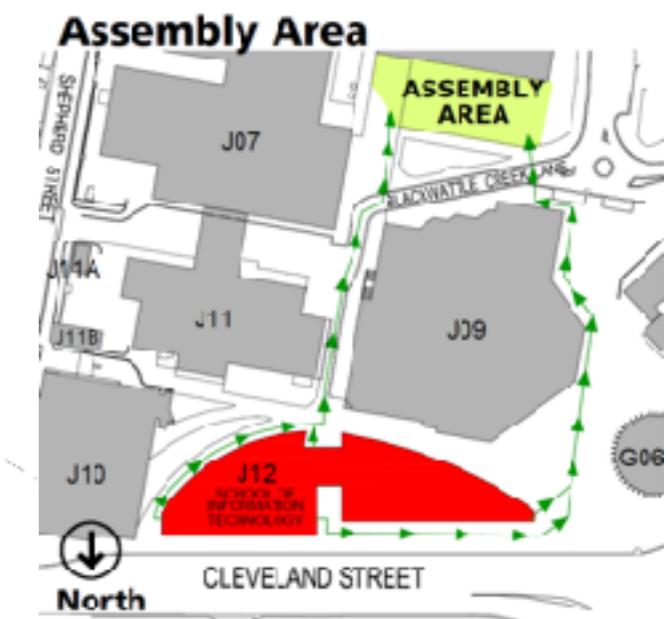
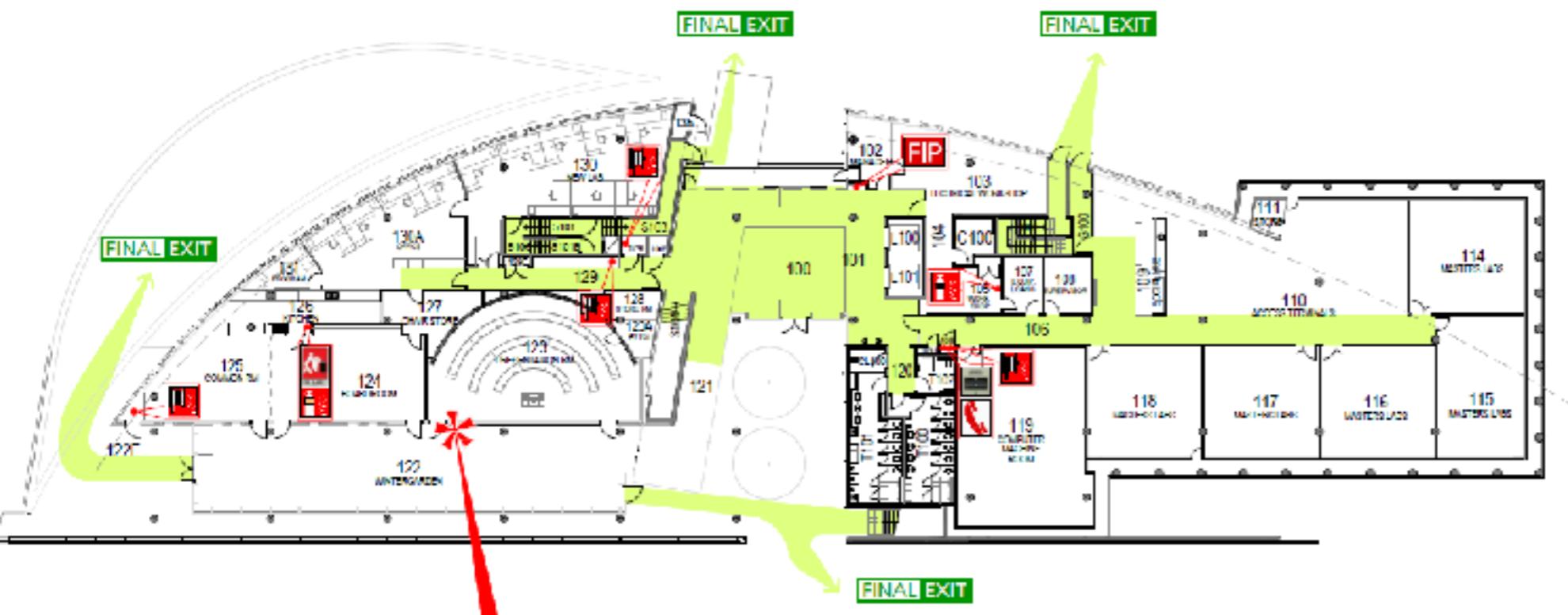
- EMERGENCY** (Left sidebar):
  - What to do in an emergency
  - First aid
  - Incident & accident reporting
  - Chief building wardens
  - Emergency management
  - Building emergency procedures
  - Handling of suspicious packages
  - Chem Alert (MSDS)
  - Mercury spills
- WHAT TO DO IN AN EMERGENCY** (Main content):

Emergencies can occur at any time, and can arise from a number of causes including fire, medical emergencies, chemical spills, gas leaks, bomb threats and physical threats. The first priority in any emergency situation is the safety of all people who may be in danger.

  - Be prepared
  - Fire alarms
  - Emergency response
  - Medical emergencies
  - People with disabilities
  - Hazardous material incidents
  - Gas leaks
  - Phone threats
  - Unattended bags or other suspicious items
  - Emergency lockdown
  - Personal safety on campus
  - Personal threats
  - Suspicious behaviour
- EMERGENCY CONTACT NUMBERS** (Right sidebar):
  - POLICE, FIRE, AMBULANCE:** Dial 0-000 from a University phone; if you are calling from an external line or mobile phone, dial 000. Be prepared to give your name and location, and details of the emergency.
  - OTHER USEFUL NUMBERS**:
    - University Security Service: 9351-3333 (This is an emergency number only.)
    - Chief fire wardens
    - Nominated first aid officers



# WHERE IS YOUR CLOSEST SAFE EXIT ?



## Evacuation Procedures

### ALARMS



**BEEP... BEEP...**

Prepare to evacuate

1. Check for any signs of immediate danger.
2. Shut Down equipment / processes.
3. Collect any nearby personal items.



**WHOOP... WHOOP...**

Evacuate the building

1. Follow the **EXIT** exit signs.
2. Escort visitors & those who require assistance.
3. DO NOT use lifts.
4. Proceed to the assembly area.

### EMERGENCY RESPONSE

1. Warn anyone in immediate danger.
2. Fight the fire or contain the emergency, if safe & trained to do so.  
**If necessary...**
3. Close the door, if safe to do so.
4. Activate the **'Break Glass'** Alarm  or 
5. Evacuate via your closest safe exit. 
6. Report the emergency to 0-000 & 9351-3333

› If a person is seriously ill/injured:

1. **call an ambulance 0-000**
2. **notify the closest Nominated First Aid Officer**

If unconscious— send for Automated External Defibrillator (AED)

## AED locations.

NEAREST to SIT Building (J12)

- Electrical Engineering Building, L2 (ground) near lifts
- Seymour Centre, left of box office
- Carried by all Security Patrol vehicles

3. **call Security - 9351-3333**

4. **Facilitate the arrival of Ambulance Staff (via Security)**



### Nearest Medical Facility

University Health Service in Level 3, Wentworth Building

### First Aid kit – SIT Building (J12)

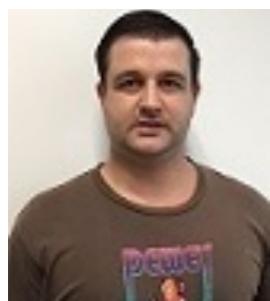
kitchen area adjacent to Lab 110

## CHIEF WARDEN

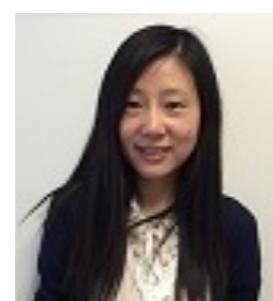
Name: Greg Ryan  
Mobile: : +61 411 406 322



## FIRST AID OFFICERS



Name: Will Calleja  
Location: 1 West  
Phone: 9036 9706



Name: Katie Yang  
Location: 2E-227  
Phone: 9351 4918

**Orally REPORT all  
INCIDENTS  
& HAZARDS  
to your SUPERVISOR**

OR

Undergraduates: to Katie Yang  
9351 4918  
  
Coursework  
Postgraduates: to Cecille Faraizi  
9351 6060

SIT School Manager: Shari Lee  
9351 4158



# Places

- Lecture:
  - Every Monday, 6 to 8pm, in New Law School Lecture Theatre 101
- Labs (**from Week 2**):
  - Every Monday, 8 to 9pm
    - New Law (Seminar rooms 100,105,107)
    - Madsen (Computer Labs 211, 206)
  - Every Tuesday, 5 to 6pm
    - SIT (Labs 115,116,118)
- Do not miss classes, except for illness, emergencies, etc
- Get help from staff if you feel you are falling behind



# Team

- Lecturers:
  - A/Prof Fabio Ramos (coordinator)
  - Dr Roman Marchant (lecturer)
  - Dr Lionel Ott (lecturer)
- Teaching Assistant:
  - Niku Gorji
- Tutors:
  - Philippe Morere, Rafael de Oliveira, Kelvin Hsu, Dongang Wang, Harrison Nguyen, Anthony Tompkins, Nicholas James, Mr XXX



# Resources

- Canvas
  - Login using Unikey and password
  - CUSP: Official schedule, list of learning outcomes, etc
  - Copies of slides
  - Lab instructions
  - Assignment instructions
  - Lecture videos
    - We intend to record the lectures, but the technology is not reliable
  - *Submit official assignments in Canvas;*
  - See your grades; etc



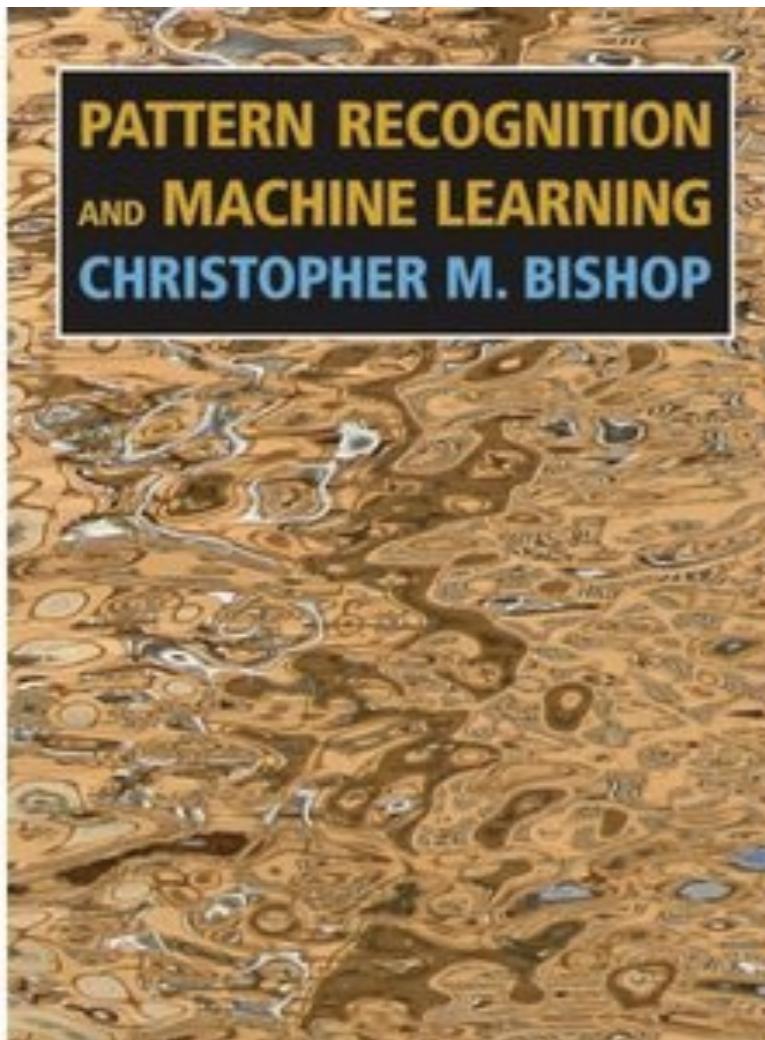
# Resources

- Two Discussion boards on Canvas:
- **General and Technical**  
[https://canvas.sydney.edu.au/courses/2441/discussion\\_topics](https://canvas.sydney.edu.au/courses/2441/discussion_topics)
- Technical questions mailbox:  
[comp5318.tutors@sydney.edu.au](mailto:comp5318.tutors@sydney.edu.au)
- Administration/General questions:  
[niku.gorji@sydney.edu.au](mailto:niku.gorji@sydney.edu.au)

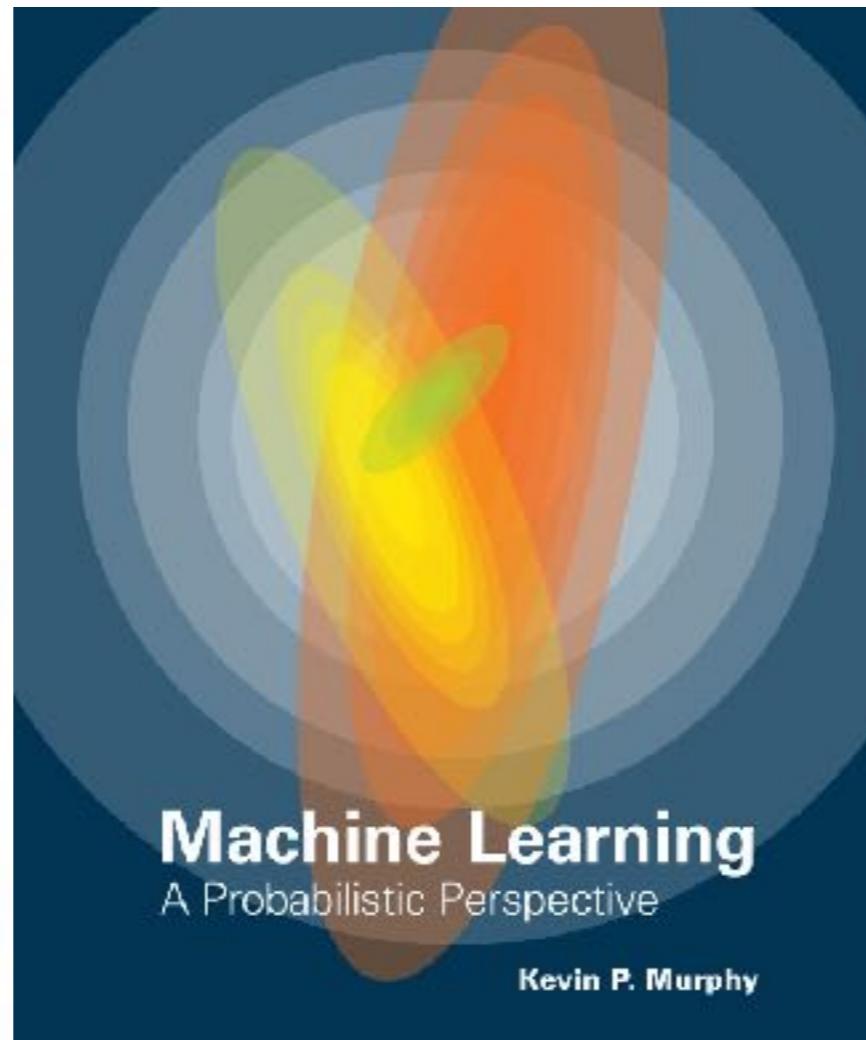


THE UNIVERSITY OF  
SYDNEY

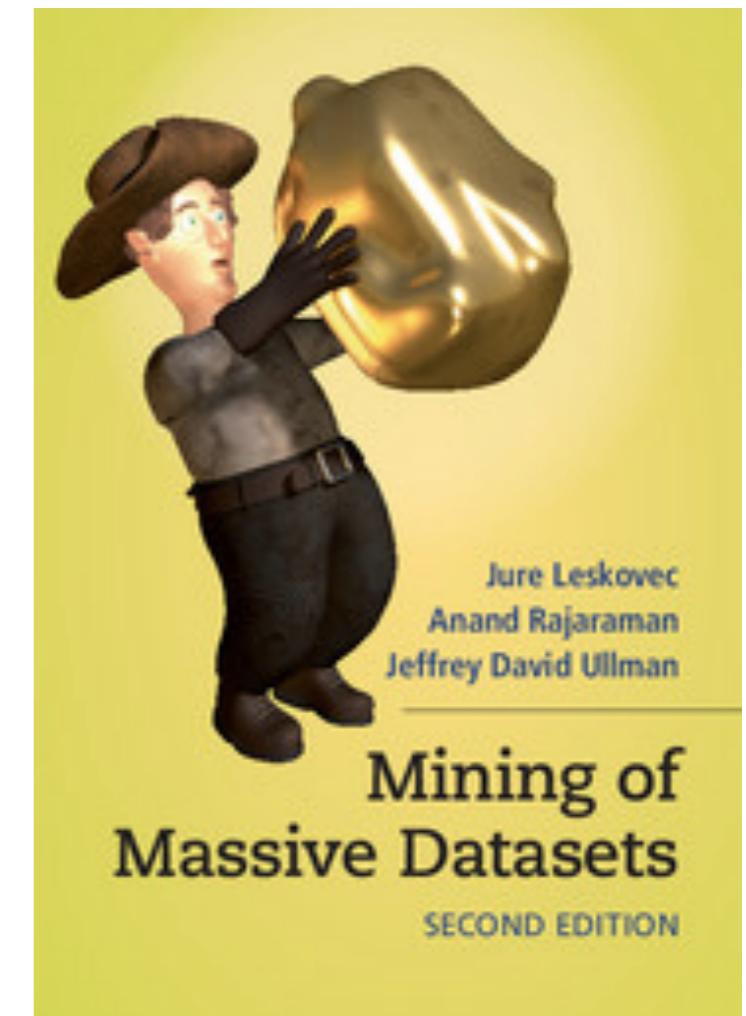
# Best books available



Textbook



Textbook



Extra



# Assessment overview

- Quiz: 10%
  - Week 4 (26/03) in-class, 6-8pm
  - Individual
  - Linear algebra + probability theory
- Assignment 1: 20%
  - 7 May 2017, 5:00pm
  - Groups of 2 or 3 students
  - Classification task
- Assignment 2: 20%
  - 4 June 2017, 5:00pm
  - Groups of 2 or 3 students
  - Method comparison and analysis
- Final exam: 50%
  - June 2017 (date to be defined)



THE UNIVERSITY OF  
SYDNEY

# Assessment overview

In order to pass the course, the School requires at least 40% in the written exam, at least 40% in the other assessment components together and an overall final mark of 50 or more. This means that students who score less than 40% in the exam will fail the course regardless of their marks during the semester.



# Expectations

- Students attend scheduled classes, and devote an extra 6-9 hrs per week
  - doing assessments
  - preparing and reviewing for classes
  - revising and integrating the ideas
  - practice and self-assess
- Students are responsible learners
  - Participate in classes, constructively
    - Respect for one another (criticise ideas, not people)
    - Humility: none of us knows it all; each of us knows valuable things
  - Check eLearning site at least once a week!
  - Notify academics whenever there are difficulties
  - Notify group partners honestly and promptly about difficulties



# Topics

Week	Date	Topic
1	5 March	Introduction, Admin matters, Research
2	12 March	Basic matrix analysis, SVD
3	19 March	Basic probability theory and Bayes' Rule
4	26 March	In-class Quiz
5	9 April	Classification and ROC curve
6	16 April	Logistic Regression and SVMs
7	24 April	Linear Regression
8	30 April	Bayesian inference / Gaussian processes
9	7 May	Latent Linear Models
10	14 May	Clustering and EM
11	21 May	Neural Nets and large scale ML
12	28 May	Deep Learning
13	4 June	Industry talk and review



THE UNIVERSITY OF  
SYDNEY

# Format of the lectures

- 10-15 min review from previous week
- 1h-1h15min of new content
- 5-10 min of examples
- 15-30 min of research topics



THE UNIVERSITY OF  
SYDNEY

# Assumed knowledge

- Linear algebra, calculus
- Basics of probability theory
- Programming skills



THE UNIVERSITY OF  
SYDNEY

If you think you are not ready...

Please consider enrolling in the second semester

Unit Code INFO5011



# Labs: Python

- Python is a high-level programming language designed to enforce good coding practices.
- Interactive and very natural to use.
- Extremely versatile and excellent for prototyping.
- Great libraries for machine learning eg. scikit-learn, TensorFlow, Keras, Edward

[www.python.org](http://www.python.org)



THE UNIVERSITY OF  
SYDNEY

# Tutorial I (homework)

- Available on canvas
- Introduction to Python
- Bring questions to your tutor next week or post them on canvas



# Special Consideration (University Policy)

- If your performance on assessments is affected by illness or misadventure
- Follow proper bureaucratic procedures
  - Have professional practitioner sign special USyd form
  - Submit application for special consideration online, upload scans
  - Note you have only a quite short deadline for applying (3 days)
  - [http://sydney.edu.au/current\\_students/special\\_consideration/](http://sydney.edu.au/current_students/special_consideration/)
- Also, notify the TA by email as soon as *anything begins to go wrong*
- There is a similar process if you need special arrangements eg for religious observance, military service, representative sports



# Academic Integrity (University Policy)

- Please read the University policy on Academic Honesty carefully:
- [http://sydney.edu.au/elearning/student/EI/academic\\_honesty.shtml](http://sydney.edu.au/elearning/student/EI/academic_honesty.shtml)
- All cases of academic dishonesty and plagiarism will be investigated
- There is a new process and a centralised University system and database
- Three types of offences:
  - **Plagiarism** – when you copy from another student, website or other source. This includes copying the whole assignment or only a part of it.
  - **Academic dishonesty** – when you make your work available to another student to copy (the whole assignment or a part of it). There are other examples of academic dishonesty.
  - **Misconduct** - when you engage another person to complete your assignment (or a part of it), for payment or not. This is a **very serious** matter and the Policy requires that your case is forwarded to the University Registrar for investigation.



# Academic Integrity (University Policy)

- The penalties are **severe** and include:
  - 1) a permanent record of academic dishonesty, plagiarism and misconduct in the University database and on your student file
  - 2) mark deduction, ranging from 0 for the assignment to Fail for the course
  - 3) expulsion from the University and cancelling of your student visa
- **Do not confuse legitimate co-operation and cheating!** You can discuss the assignment with another student, this is a legitimate collaboration, but you cannot complete the assignment together – everyone must write their own code or report, unless the assignment is group work.
- When there is copying between students, note that **both students are penalised** – the student who copies and the student who makes his/her work available for copying



# Academic Integrity (University Policy)

- We will use the similarity detection software TurnItIn and MOSS to compare your assignments with these of other students (current and previous) and the Internet
  - Turnitin is for text documents: [http://www.turnitin.com/en\\_us/higher-education](http://www.turnitin.com/en_us/higher-education)
  - MOSS is for programming code: <https://theory.stanford.edu/~aiken/moss/>
- These tools are **extremely good!**
  - e.g. MOSS cannot be fooled by changing the names of the variables or changing the order of the conditions in if-else statements
- Examples of plagiarism in programming code:
  - [http://www.upenn.edu/academicintegrity/ai\\_computercode.html](http://www.upenn.edu/academicintegrity/ai_computercode.html)



# Academic Integrity (University Policy)

- All these are cases of **plagiarism** and **academic dishonesty** we have seen in our school and the student excuses are not acceptable:
  - *I sat the test and then posted the questions and solutions to my friends whose test was later in the week. I only wanted to help them understand the concepts that are examinable.*
  - *I posted parts of my code on my web page (group discussion forum) because my solution was cool (or I wanted to help them). I didn't expect them to copy it.*
  - *I tried to do the assignment on my own but I had problems with the extension part that I couldn't fix, so I submitted my core part and his extension part. I didn't cheat.*
  - *I finished my assignment but my friend had family problems. I felt sorry for her, so I gave her my assignment as an example. She said she only wanted to have a look and promised not to copy it.*
  - *The test has finished but the tutor hasn't collected the papers yet. I showed my answer to my friend. I didn't expect him to copy it.*
  - *He is my best friend. I had no choice but to let him copy my assignment.*



# Academic Integrity (University Policy)

- Plagiarism and any form of academic dishonesty will be dealt with, and the penalties are severe
- We use plagiarism detection systems such as MOSS and TurnItIn that are extremely good. If you cheat, the chances you will be caught are very high.
- If someone asks you to see or copy your assignment, or to complete the assignment instead of them, just say: *I can't do this. This is against the University policy. I will not risk my future by doing this.*

**Be smart and don't risk your future by engaging in plagiarism and academic dishonesty!**



# Self-test

- When is the first assessment due?
- What do you do if you get sick during semester?
- What help can you use when answering assessments?
- How do you find out about assignment instructions?
- How do you submit your work?
- What is Turnitin?
- What language will you be coding in?

# Do you have a disability?

You may not think of yourself as having a 'disability' but the definition under the **Disability Discrimination Act (1992)** is broad and includes temporary or chronic medical conditions, physical or sensory disabilities, psychological conditions and learning disabilities.

The types of disabilities we see include:

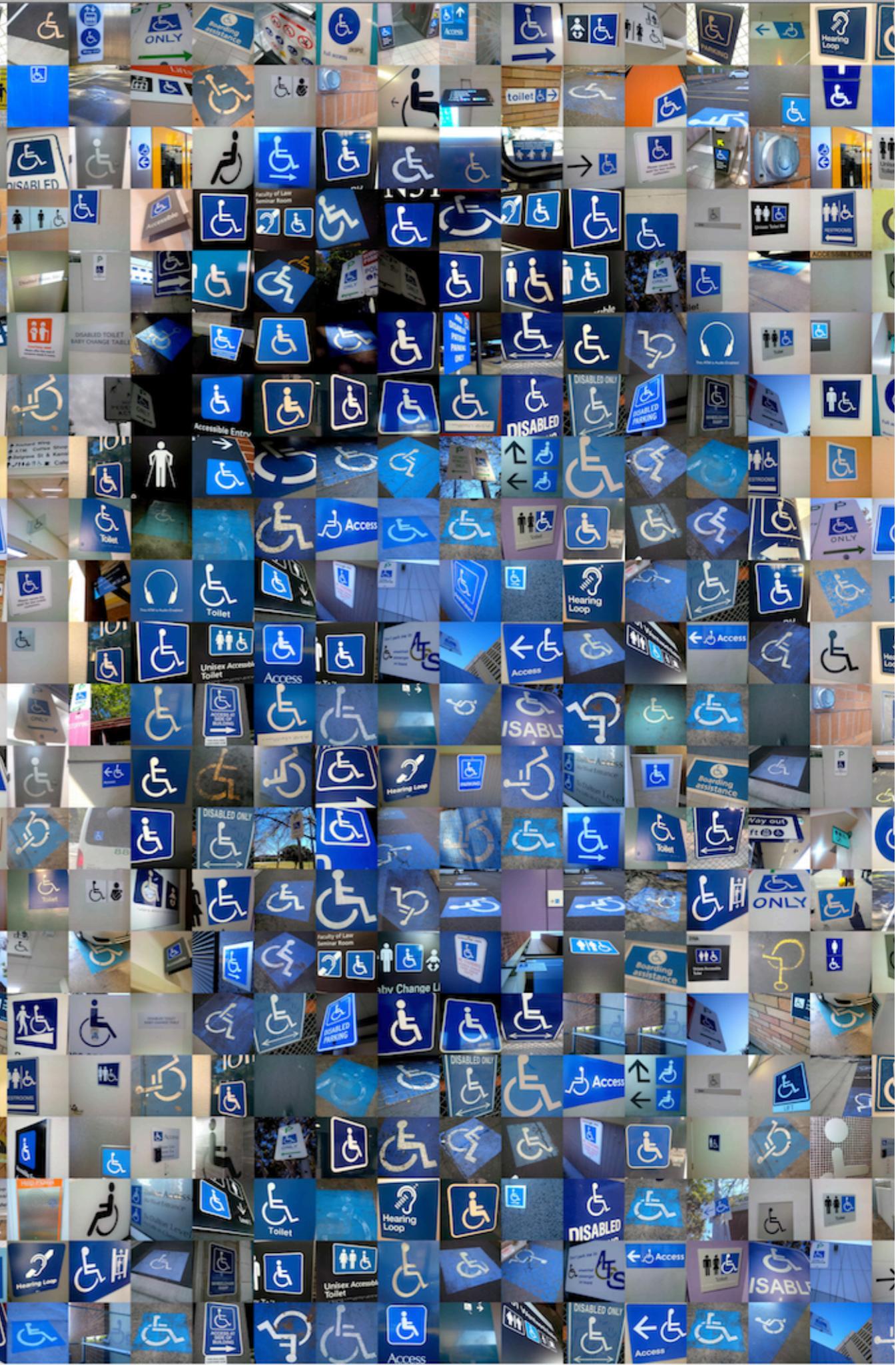
Anxiety // Arthritis // Asthma // Autism // ADHD  
Bipolar disorder // Broken bones // Cancer  
Cerebral palsy // Chronic fatigue syndrome  
Crohn's disease // Cystic fibrosis // Depression  
Diabetes // Dyslexia // Epilepsy // Hearing impairment // Learning disability // Mobility impairment // Multiple sclerosis // Post-traumatic stress // Schizophrenia // Vision impairment  
and much more.

Students needing assistance must register with Disability Services. It is advisable to do this as early as possible. Please contact us or review our website to find out more.



THE UNIVERSITY OF  
**SYDNEY**

**Disability Services Office**  
[sydney.edu.au/disability](http://sydney.edu.au/disability)  
02-8627-8422





# Other support

- Learning support
  - <http://sydney.edu.au/study/academic-support/learning-support.html>
- International students
  - <http://sydney.edu.au/study/academic-support/support-for-international-students.html>
- Aboriginal and Torres Strait Islanders
  - <http://sydney.edu.au/study/academic-support/aboriginal-and-torres-strait-islander-support.html>
- Student organisation (can represent you in academic appeals etc)
  - <http://srcusyd.net.au/> or <http://www.supra.net.au/>
- Please make contact, and get help
- You are not required to tell anyone else about this
- If you are willing to inform the unit coordinator, they may be able to work with other support to reduce the impact on this unit
  - eg provide advice on which tasks are most significant

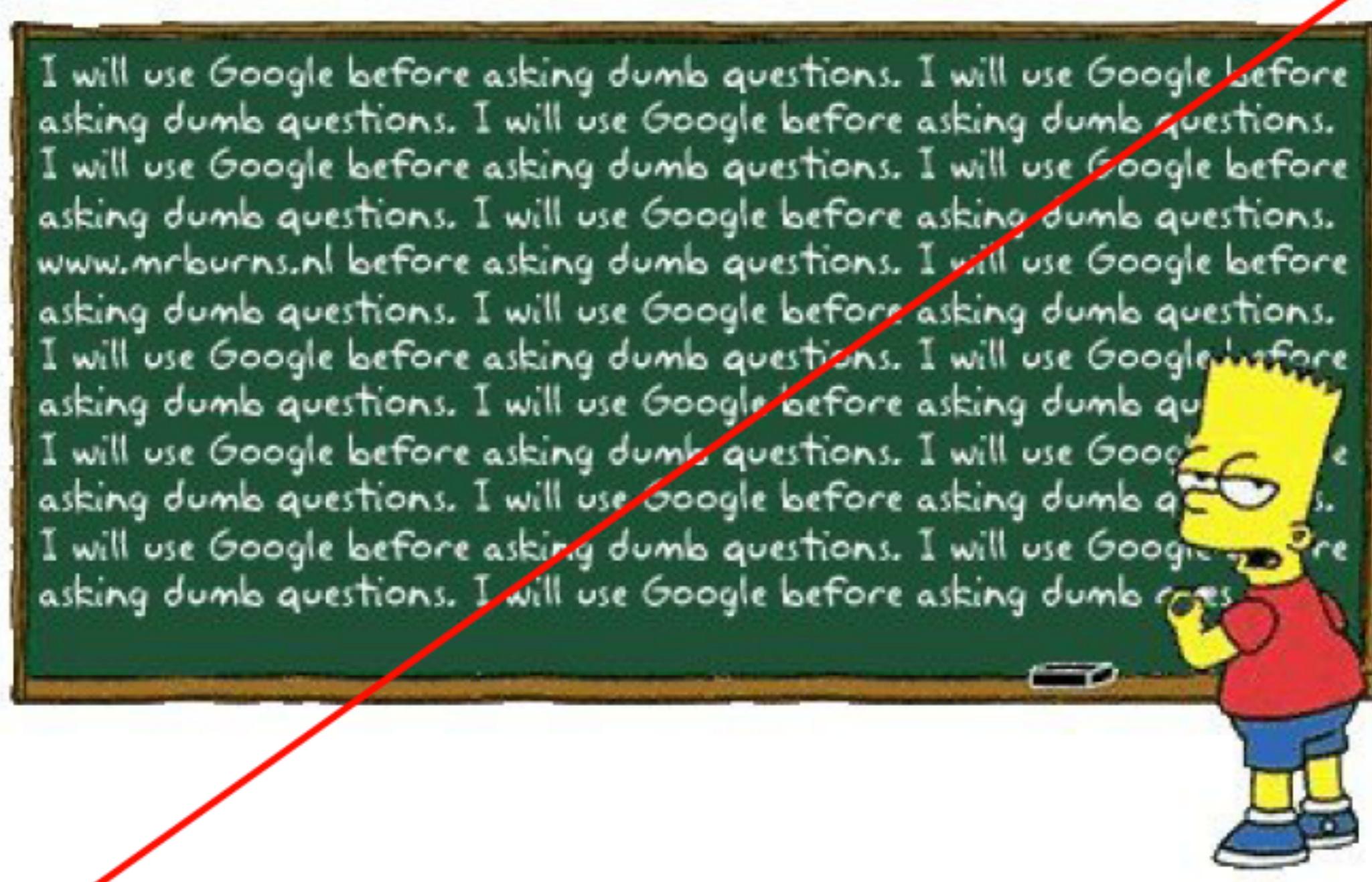


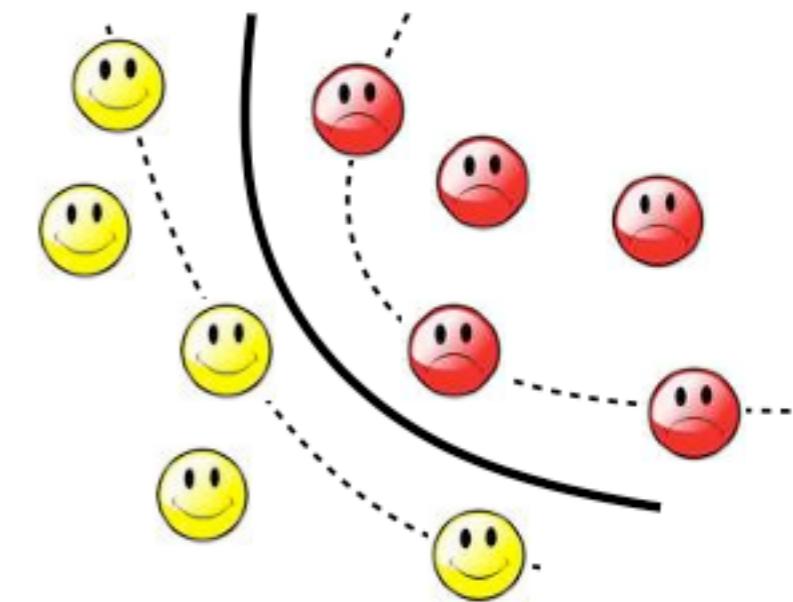
# Advice

- Metacognition
  - Pay attention to the learning outcomes in CUSP
  - Self-check that you are achieving each one
  - Think how each assessment task relates to these
- Time management
  - Watch the due dates
  - Start work early, submit early
- Networking and community-formation
  - Make friends and discuss ideas with them
  - Know your TA, tutor, lecturer, coordinator
  - Keep them informed, especially if you fall behind
    - Don't wait to get help
- Enjoy the learning!



# Ask questions sooner!



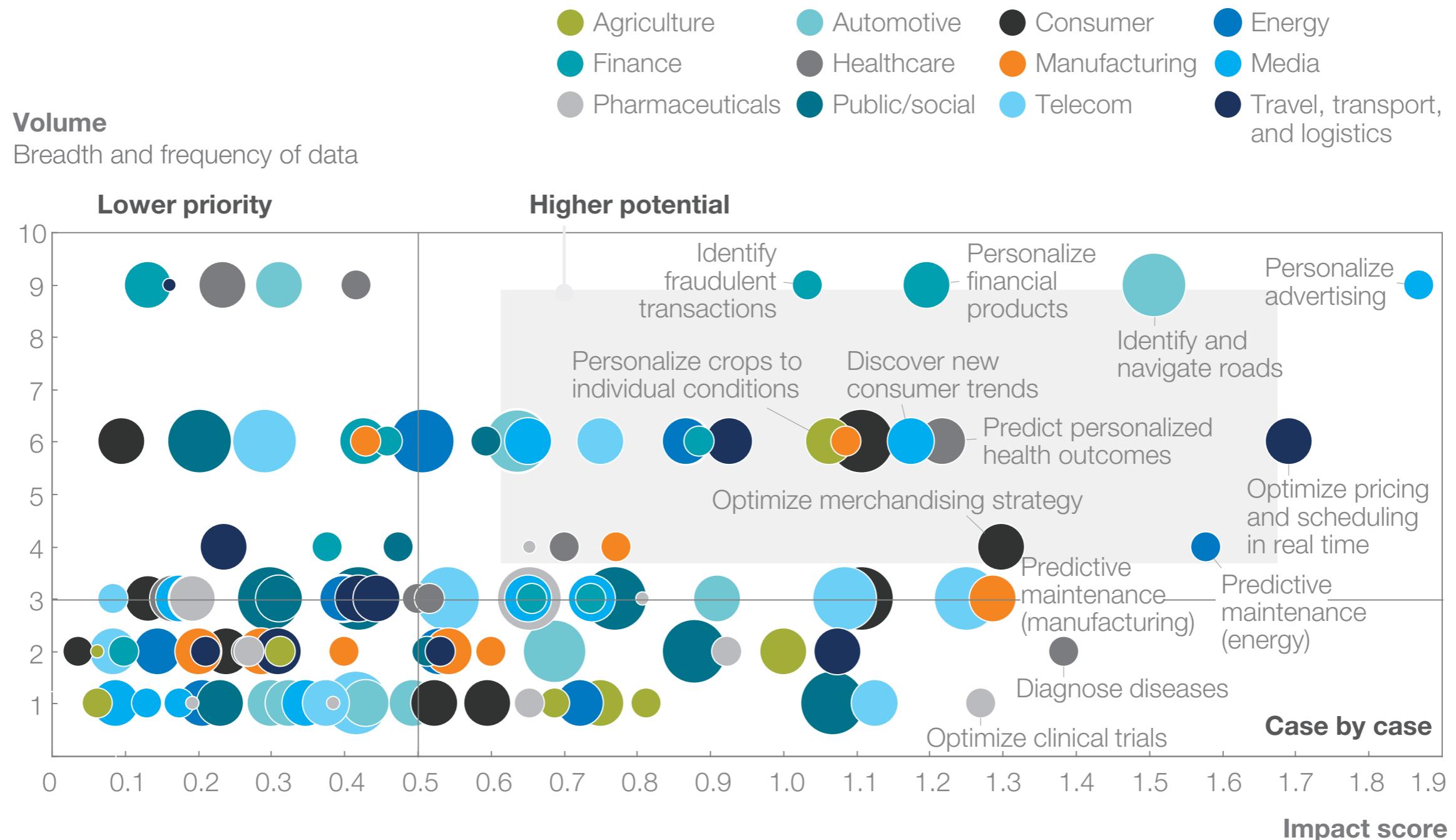


Finally let's talk about  
Machine Learning



i am ai

NVIDIAから生まれた

**Exhibit 4:** Machine learning has broad potential across industries and use cases


Source: McKinsey Global Institute



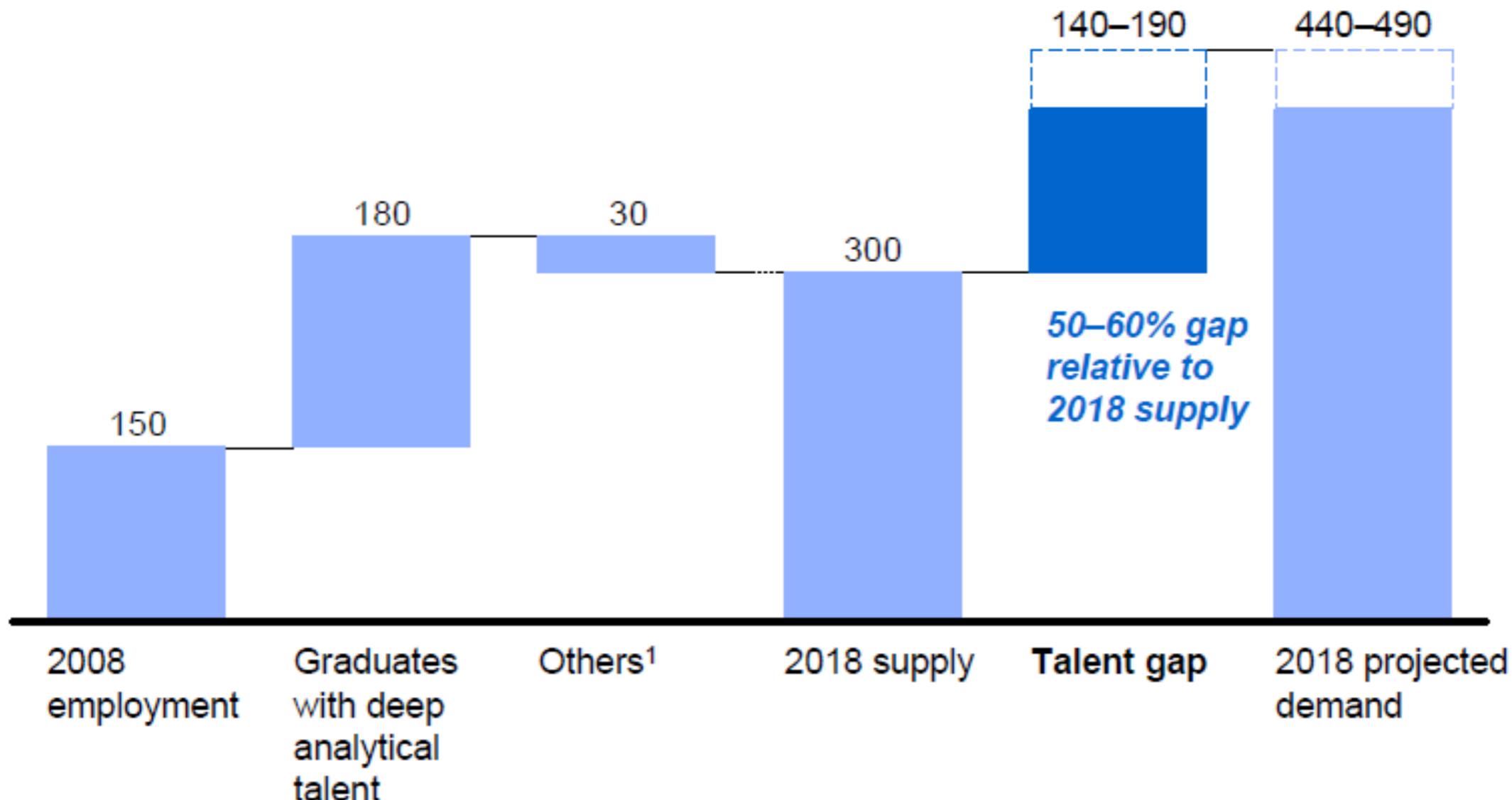
# Good news: Demand for ML

THE UNIVERSITY OF  
**SYDNEY**

**Demand for deep analytical talent in the United States could be 50 to 60 percent greater than its projected supply by 2018**

Supply and demand of deep analytical talent by 2018

Thousand people



<sup>1</sup> Other supply drivers include attrition (-), immigration (+), and reemploying previously unemployed deep analytical talent (+).

SOURCE: US Bureau of Labor Statistics; US Census; Dun & Bradstreet; company interviews; McKinsey Global Institute analysis

J. Leskovec, A. Rajaraman, J. Ullman: Mining of Massive Datasets, <http://www.mmds.org>

**Google**™

**amazon.com**®



**TOYOTA**

**SONY**

**SIEMENS**

**IBM**

**Microsoft®**

**Research**

**Research**

**facebook.**

**YAHOO!**®



**i n v e n t**

**HRI**  
Europe  
Honda Research Institute

**CommonwealthBank**



**RioTinto**



THE UNIVERSITY OF  
SYDNEY

# What is Machine Learning?

Informally: Making predictions from data

Formally: The construction of a statistical model that is an underlying distribution from which the data is drawn from.



# ML / DM

- **But to extract the knowledge data needs to be**
  - **Stored**
  - **Managed**
  - **And ANALYSED ← this course**

**Data Mining ≈ Big Data ≈ Statistics  
≈ Machine Learning ≈ Data Science**

**\$600** to buy a disk drive that can store all of the world's music

**5 billion** mobile phones in use in 2010

**30 billion** pieces of content shared on Facebook every month

**40%** projected growth in global data generated per year vs.

**5%** growth in global IT spending

**\$5 million vs. \$400**

Price of the fastest supercomputer in 1975<sup>1</sup> and an iPhone 4 with equal performance

**235** terabytes data collected by the US Library of Congress by April 2011

**15 out of 17** sectors in the United States have more data stored per company than the US Library of Congress<sup>42</sup>



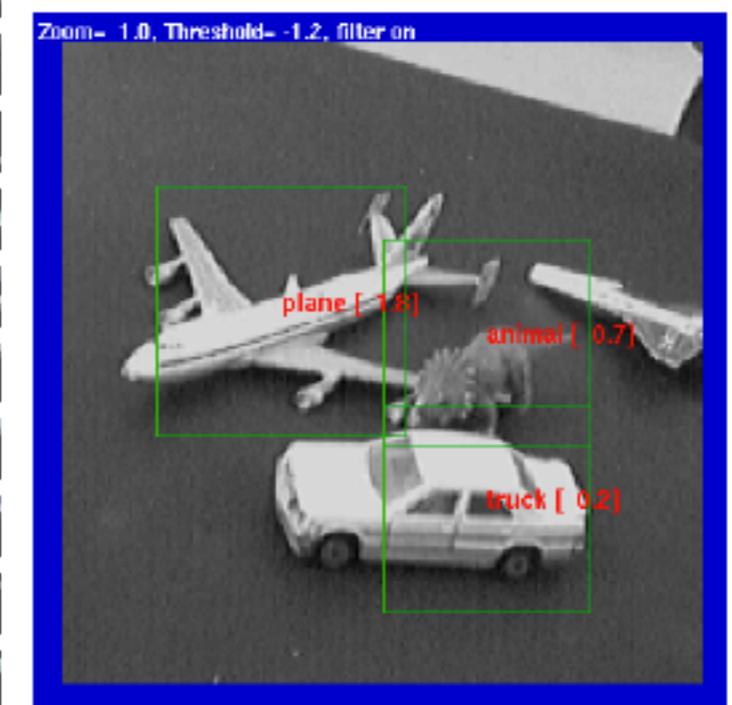
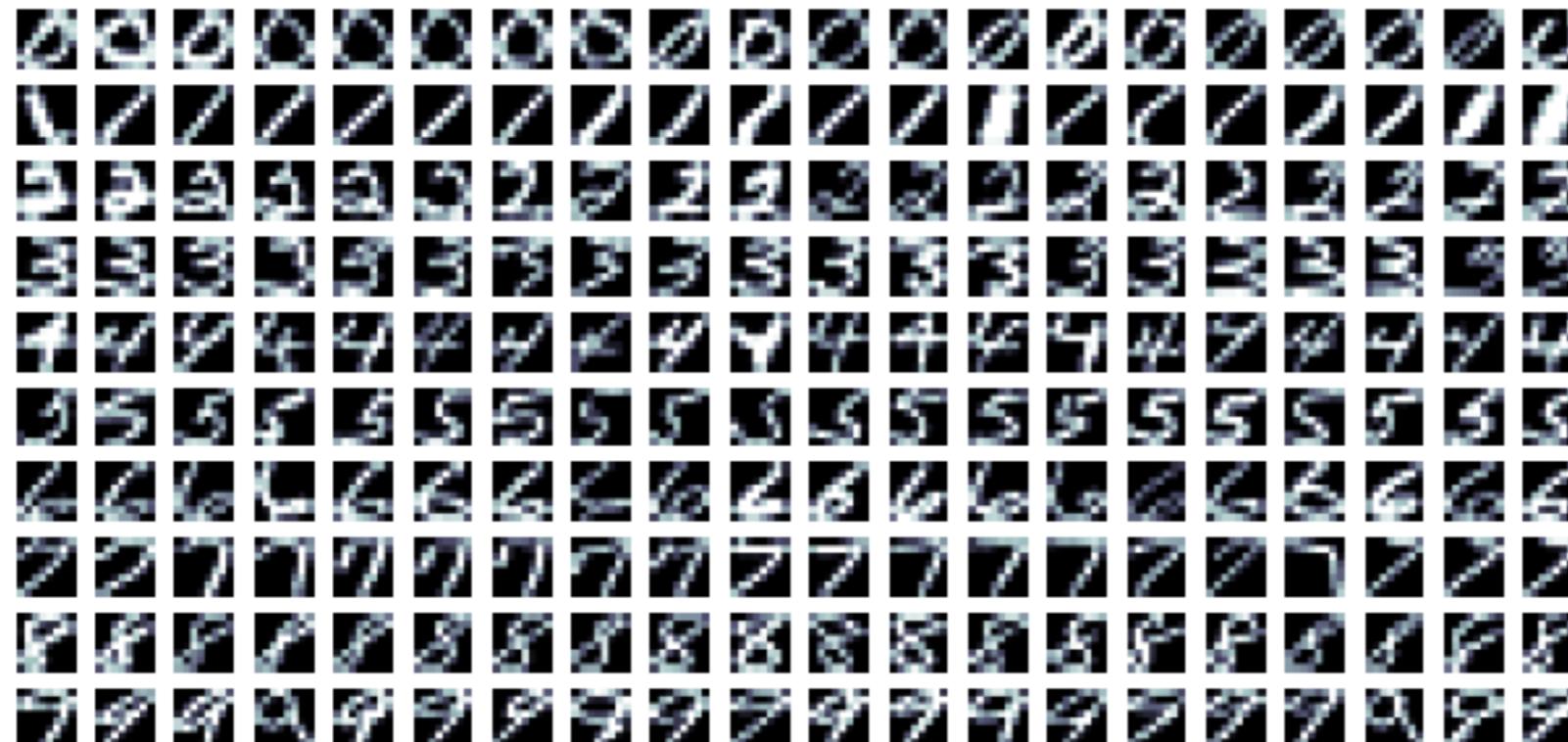
THE UNIVERSITY OF  
SYDNEY

# Speech recognition





THE UNIVERSITY OF  
SYDNEY



(NORB image from Yann LeCun)



# Information retrieval

Google Search: Unsupervised Learning <http://www.google.com/search?q=Unsupervised+Learning&esrc=s&tbo=...>

Web Images Groups News Books more... [Advanced Search](#) [Feedback](#)

Web Results 1 - 10 of about 160,000 for [Unsupervised Learning](#) (0.27 seconds)

Mixture modelling, Clustering, Intrinsic classification... Mixture Modelling page. Welcome to David Dowsen's clustering, mixture modelling and unsupervised learning page. Mixture modelling (or ... [www.csse.monash.edu.au/~dcl/mixture/modelling/pgs.html](http://www.csse.monash.edu.au/~dcl/mixture/modelling/pgs.html) - 26k - 4 Oct 2004 - Cached - Similar pages

ACL'99 Workshop -- Unsupervised Learning in Natural Language ... PROGRAM: ACL'99 Workshop Unsupervised Learning in Natural Language Processing. University of Maryland June 21, 1999. Sponsored by SIGNLL ... [www.aList.com/~kohler/unsup-ad-99.html](http://www.aList.com/~kohler/unsup-ad-99.html) - 5k - Cached - Similar pages

Unsupervised learning and Clustering [cgm.cs.mcgill.ca/~scs/cos644/projects/wtjhe/](http://cgm.cs.mcgill.ca/~scs/cos644/projects/wtjhe/) - 1k - Cached - Similar pages

NIPS'98 Workshop - Integrating Supervised and Unsupervised... NIPS'98 Workshop "Integrating Supervised and Unsupervised Learning" Friday, December 4, 1998 ... 4:45-5:30, Théâtre of Unsupervised Learning and Missing Values... [www-2.cs.cmu.edu/~mcallum/supunsup/](http://www-2.cs.cmu.edu/~mcallum/supunsup/) - 7k - Cached - Similar pages

NIPS Tutorial 1999 Probabilistic Models for Unsupervised Learning Tutorial presented at the 1999 NIPS Conference by Zoubin Ghahramani and Sam Roweis ... [www.gatsby.ucl.ac.uk/~zoubin/NIPStutorial.html](http://www.gatsby.ucl.ac.uk/~zoubin/NIPStutorial.html) - 4k - Cached - Similar pages

Gatsby Course: Unsupervised Learning : Homepage Unsupervised Learning (Fall 2000) ... syllabus (resources page); 10/10 1 - Introduction to Unsupervised Learning Geoff project (ps, pdf)... [www.gatsby.ucl.ac.uk/~quaid/course/](http://www.gatsby.ucl.ac.uk/~quaid/course/) - 19k - Cached - Similar pages [More results from [www.gatsby.ucl.ac.uk/](http://www.gatsby.ucl.ac.uk/)]

pdf] Unsupervised Learning of the Morphology of a Natural Language File Format: PDF/Adobe Acrobat - View as HTML Page 1. Page 2. Page 3. Page 4. Page 5. Page 6. Page 7. Page 8. Page 9. Page 10. Page 11. Page 12. Page 13. Page 14. Page 15. Page 16. Page 17. Page 18. Page 19... [ad1.cs.psu.edu/JJ/J01/J01-2001.pdf](http://ad1.cs.psu.edu/JJ/J01/J01-2001.pdf) - Similar pages

Unsupervised Learning - The MIT Press ... From Bradford Books: Unsupervised Learning Foundations of Neural Computation Edited by Geoffrey Hinton and Terrence J. Sejnowski Since its founding in 1989 by ... [mitpress.mit.edu/book-home.tcl?isbn=026258168X](http://mitpress.mit.edu/book-home.tcl?isbn=026258168X) - 13k - Cached - Similar pages

(ps) Unsupervised Learning of Disambiguation Rules for Part of... File Format: Adobe PostScript - View as Text Unsupervised Learning of Disambiguation Rules for Part of Speech Tagging- Eric Brill. 1. ... It is possible to use unsupervised learning to train stochastic... [www.cs.jhu.edu/~brill/col-wkshop.ps](http://www.cs.jhu.edu/~brill/col-wkshop.ps) - Similar pages

The Unsupervised Learning Group (ULG) at UT Austin ... The Unsupervised Learning Group (ULG). What ? The Unsupervised Learning Group (ULG) is a group of graduate students from the Computer... [www.cs.utexas.edu/ulg/](http://www.cs.utexas.edu/ulg/) - 14k - Cached - Similar pages

Result Page: 1 2 3 4 5 6 7 8 9 10 [Next](#)

1 of 2 06/10/04 15:44

## Web Pages

Retrieval  
Categorisation  
Clustering  
Relations between pages



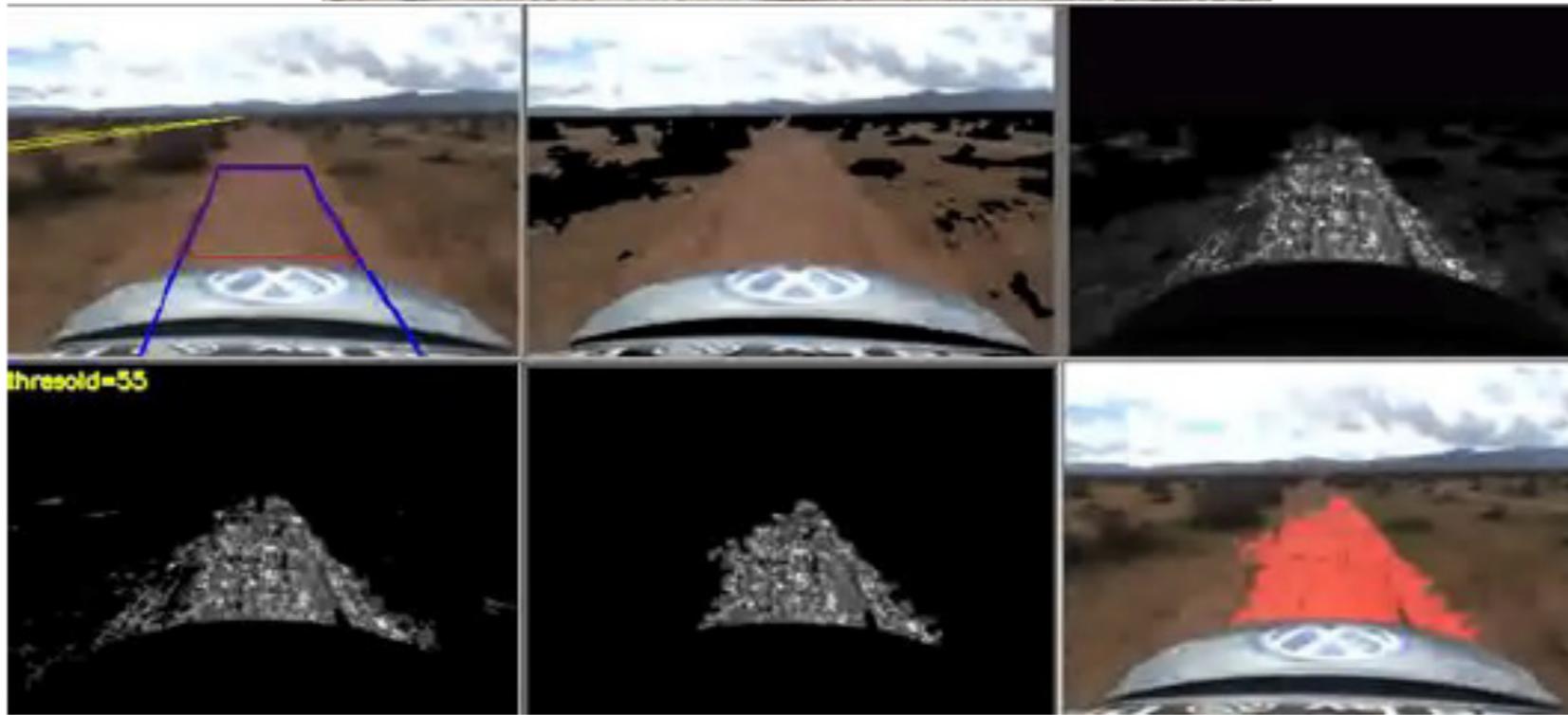
# Financial prediction





THE UNIVERSITY OF  
SYDNEY

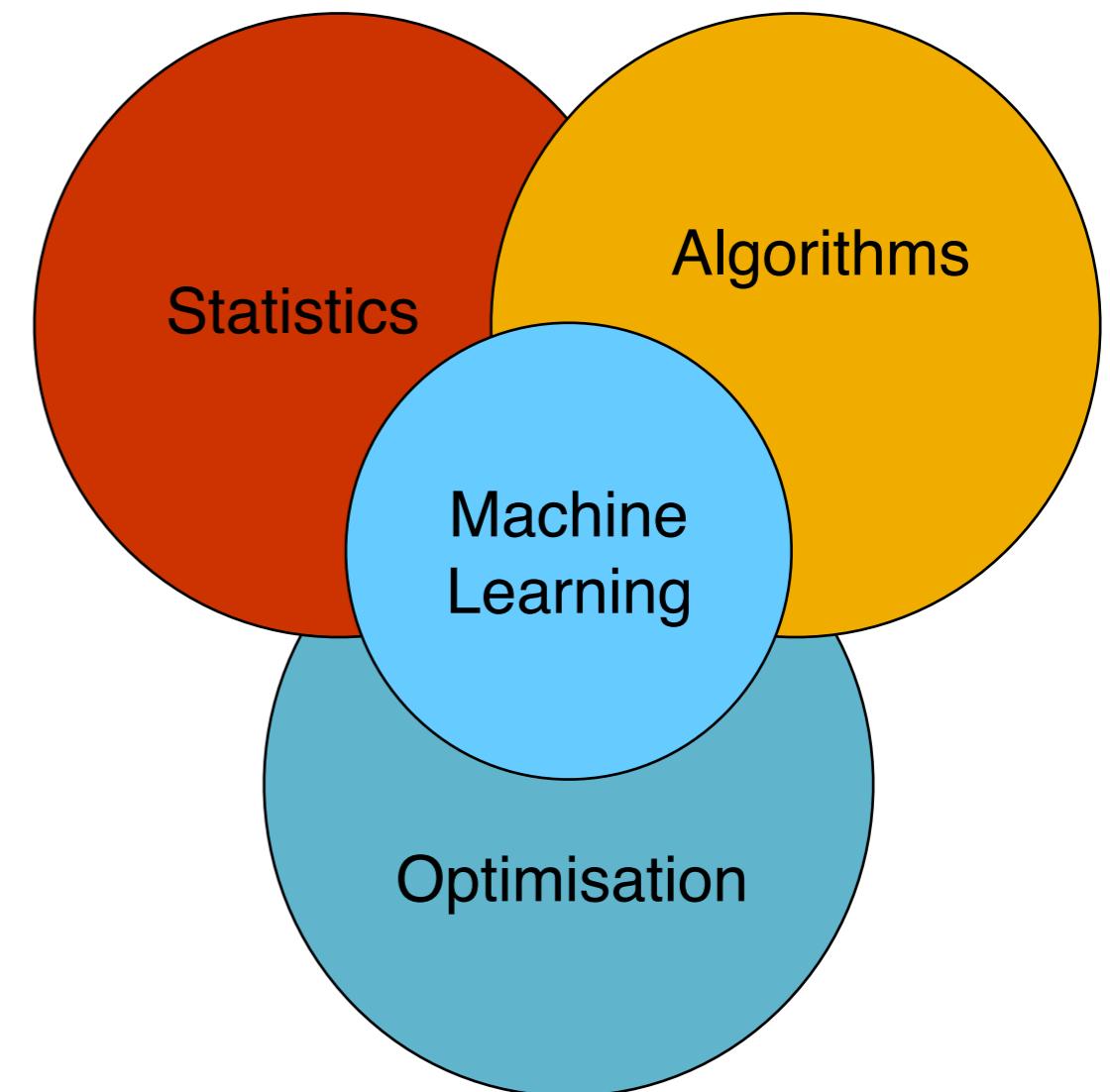
# Robotics





# This Course: COMP 5318

- This course overlaps with statistics, artificial intelligence, databases but more stress on
  - Algorithms
  - Mathematical modelling
  - Automation for handling large data





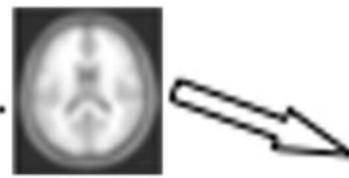
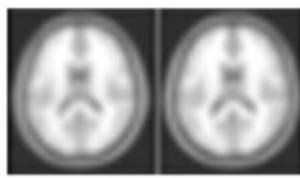
# Machine Learning Problems

- Prediction
  - Classification and Regression
- Clustering, segmentation and summarisation
  - Find patterns in the data
- Outlier/anomaly detection
  - Find unusual patterns
- Reinforcement learning
  - Learn from rewards (like babies)

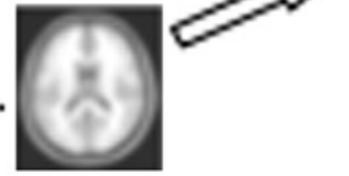
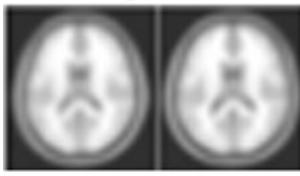


# Elements of Machine Learning

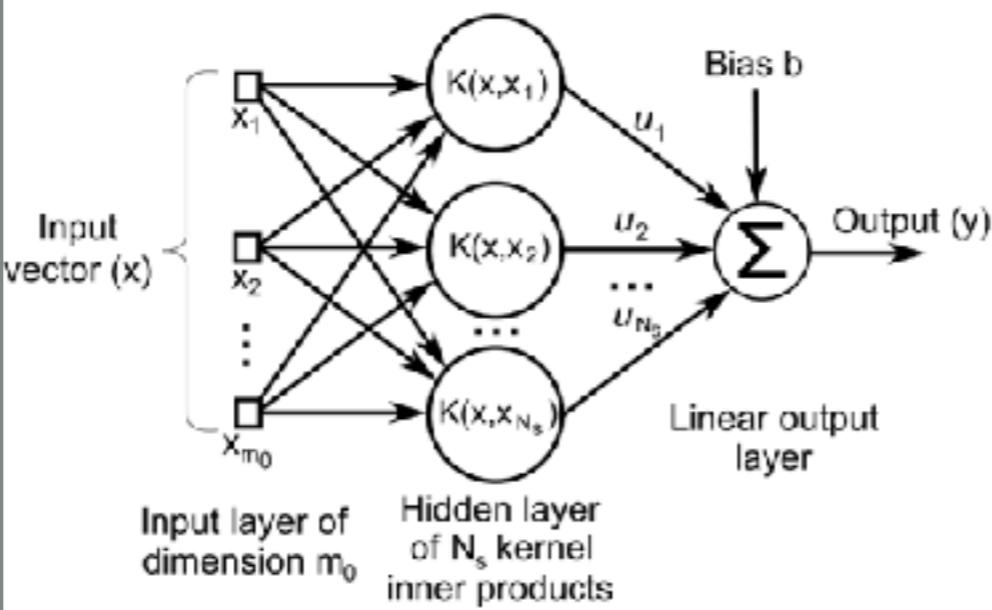
Group 1



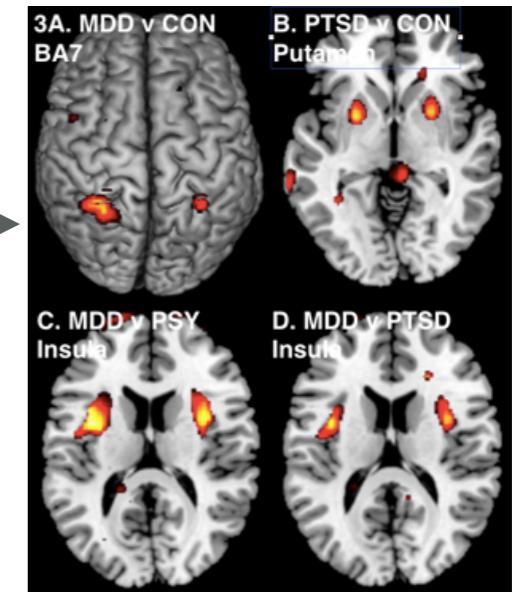
Group 2



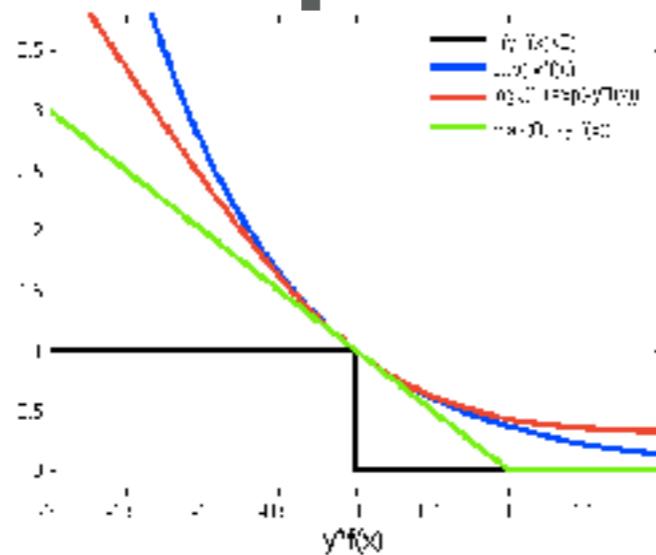
## Mathematical Model



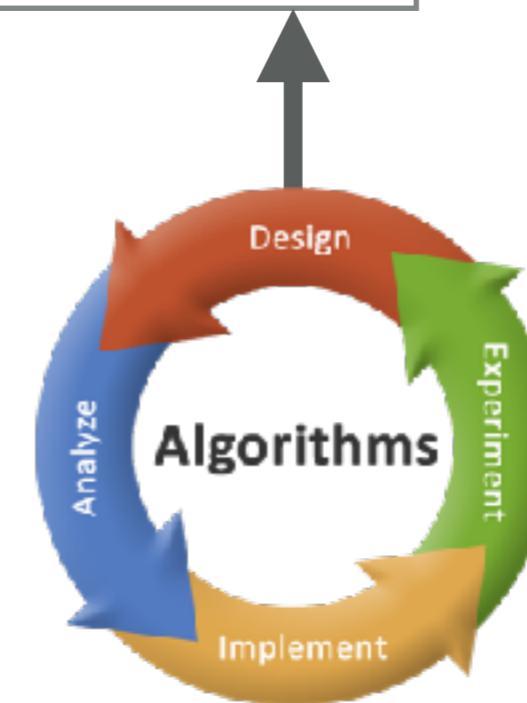
## Predictions/Patterns



Data

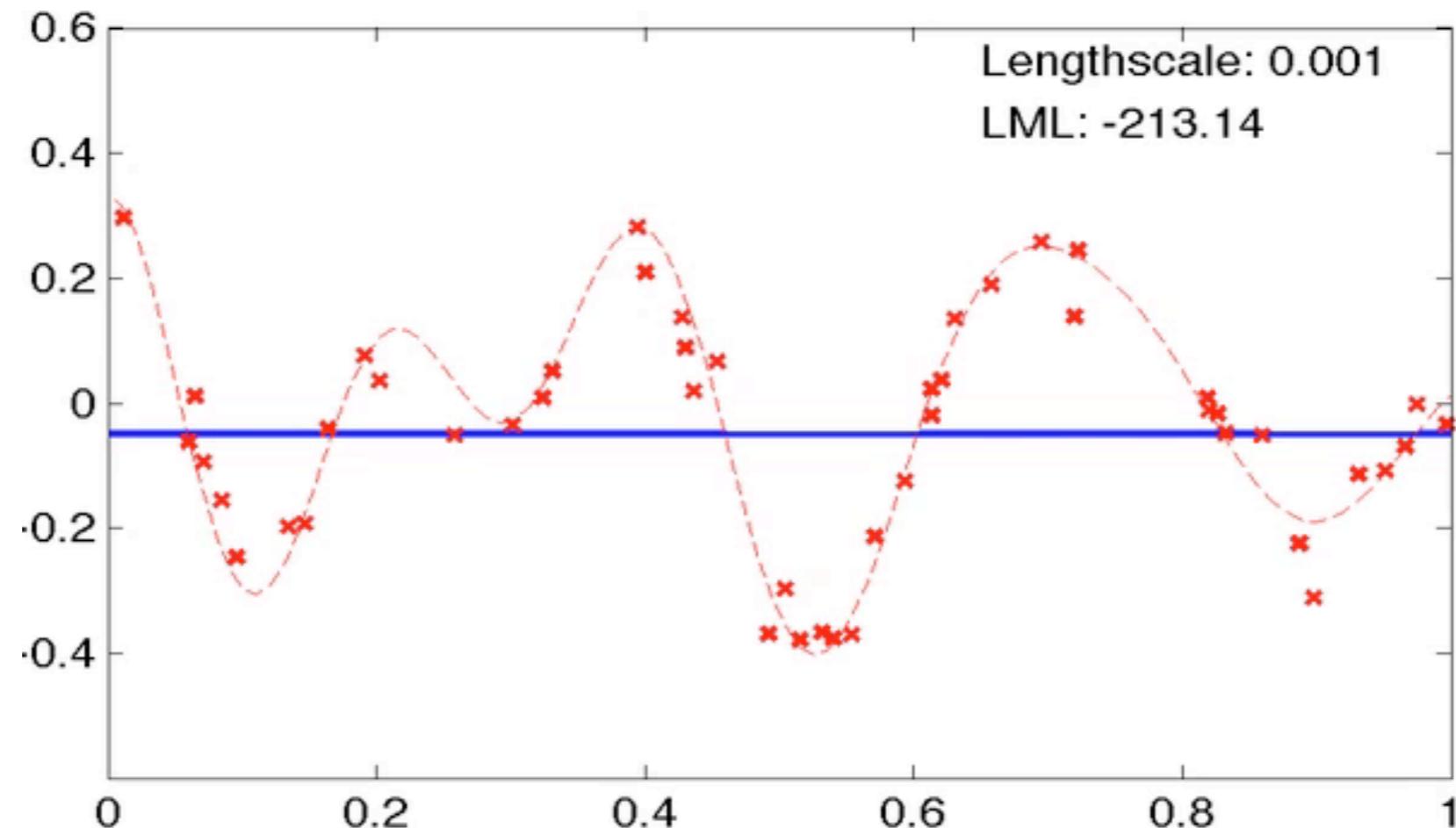


Objective function

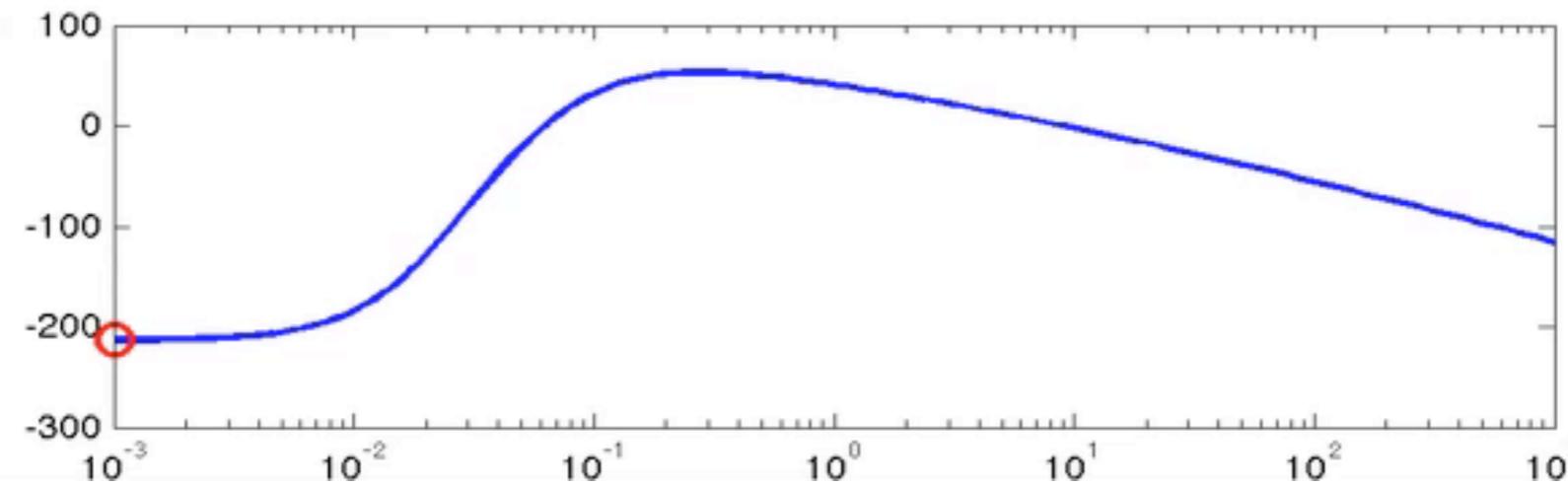


# Regression

Problem



Objective





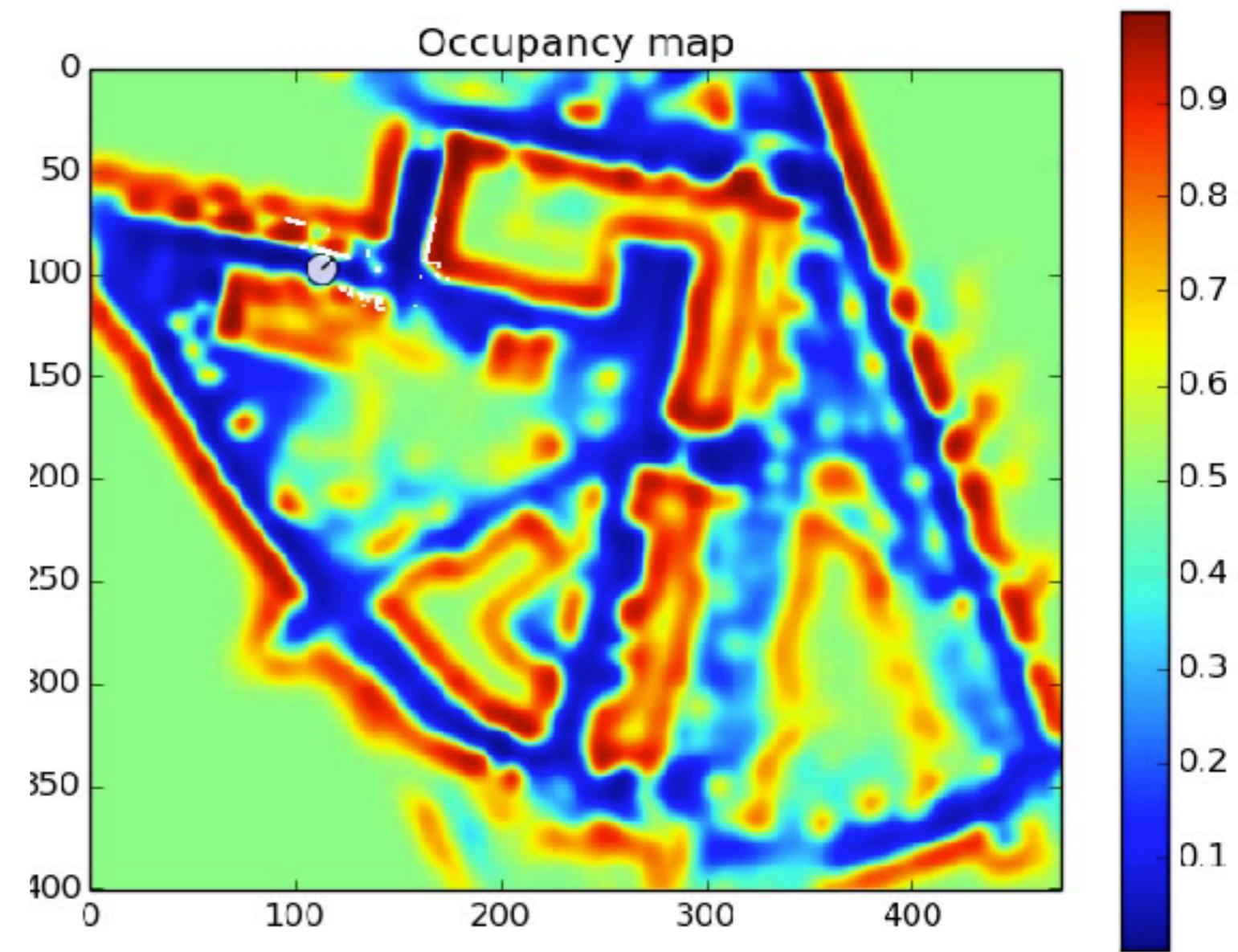
THE UNIVERSITY OF  
SYDNEY

# Classification for auto-driving

Aerial View

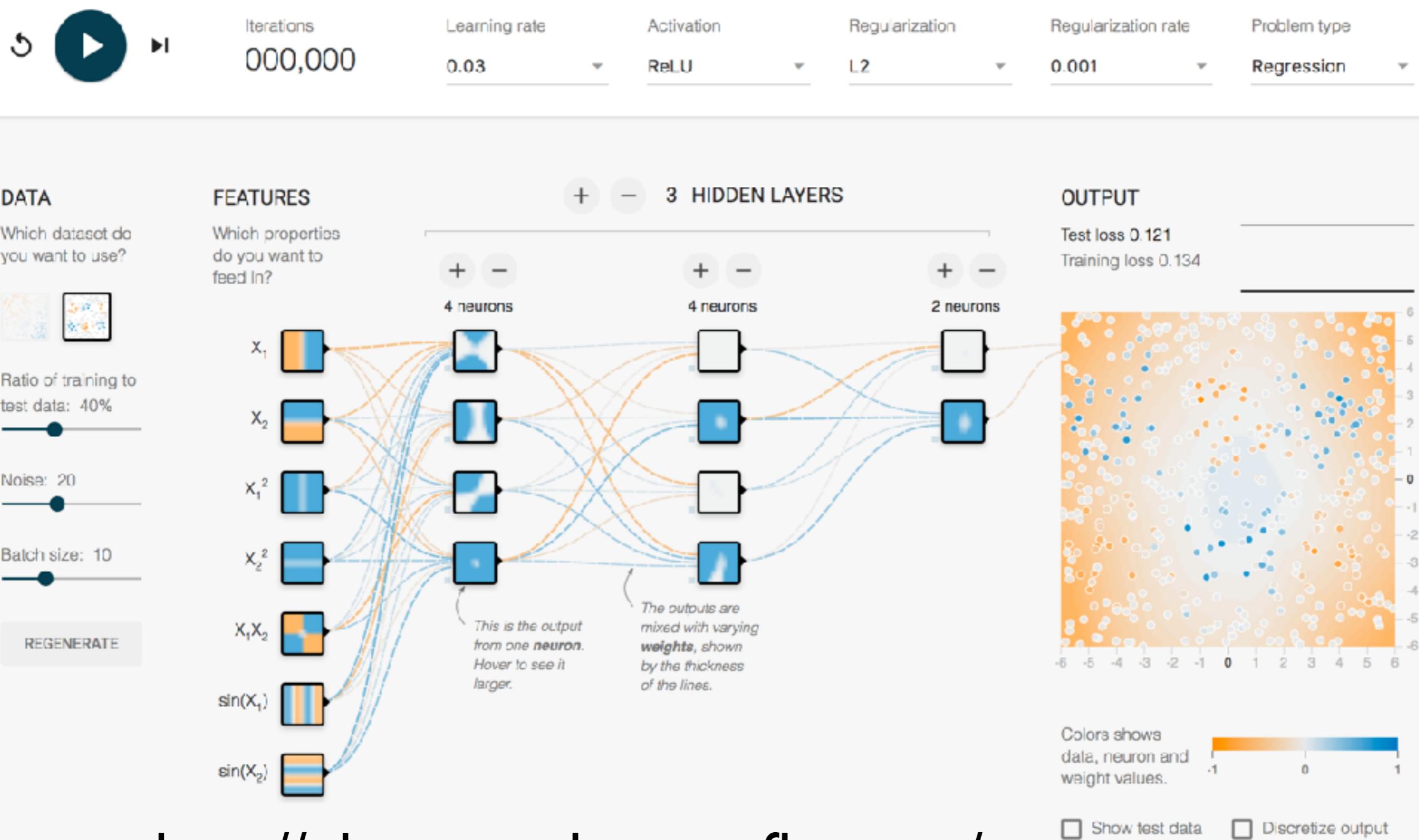


Occupancy map





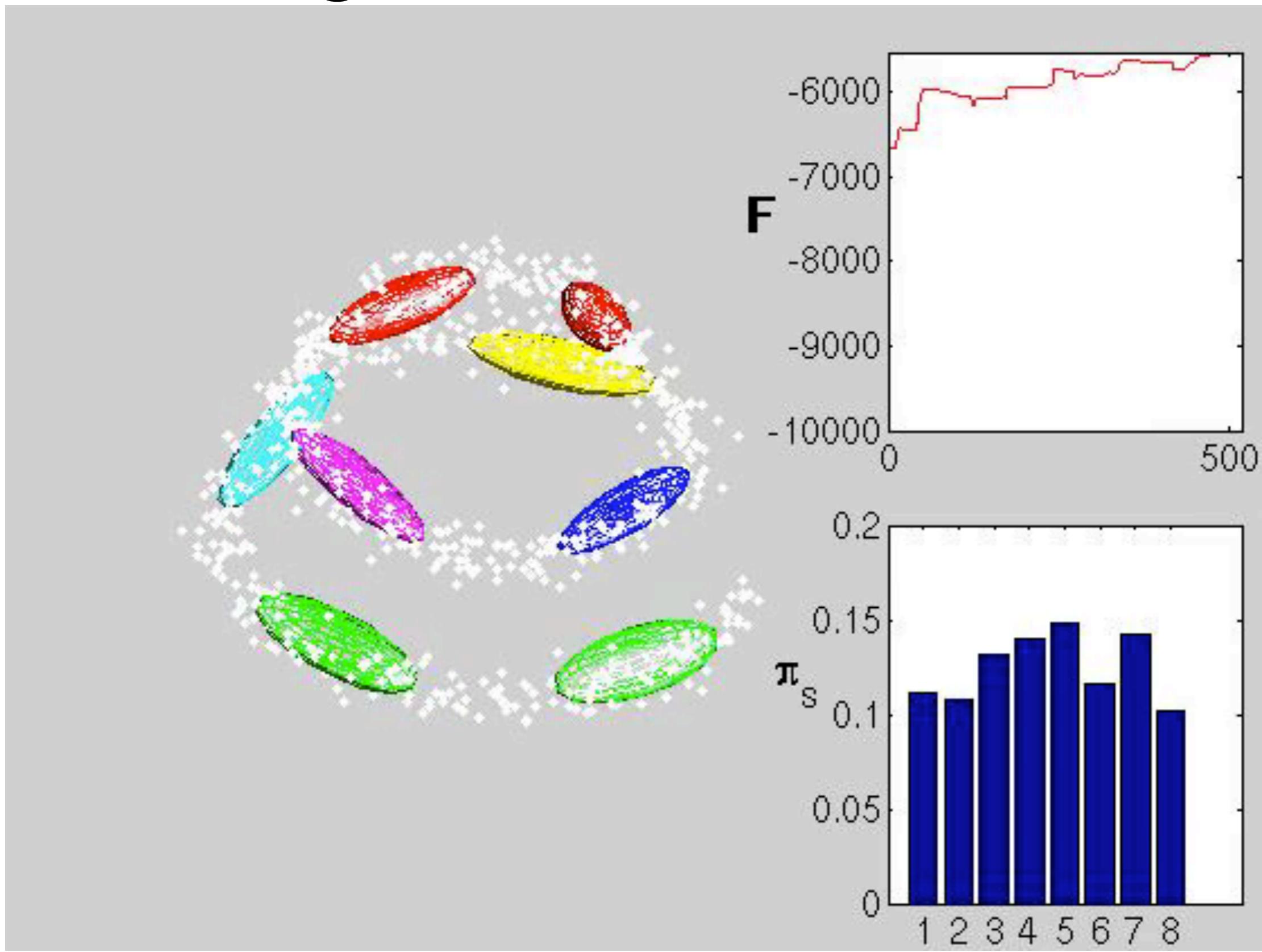
THE UNIVERSITY OF  
SYDNEY



Source: <http://playground.tensorflow.org/>



THE UNIVERSITY OF  
SYDNEY





# Common representation

IMAGE/  
VIDEO

TEXT/  
COMMENT

TIME  
SERIES

SYSTEM  
LOGS

NETWORK

TABULAR/  
RATING

Is there a common way to represent data  
of different modalities ?



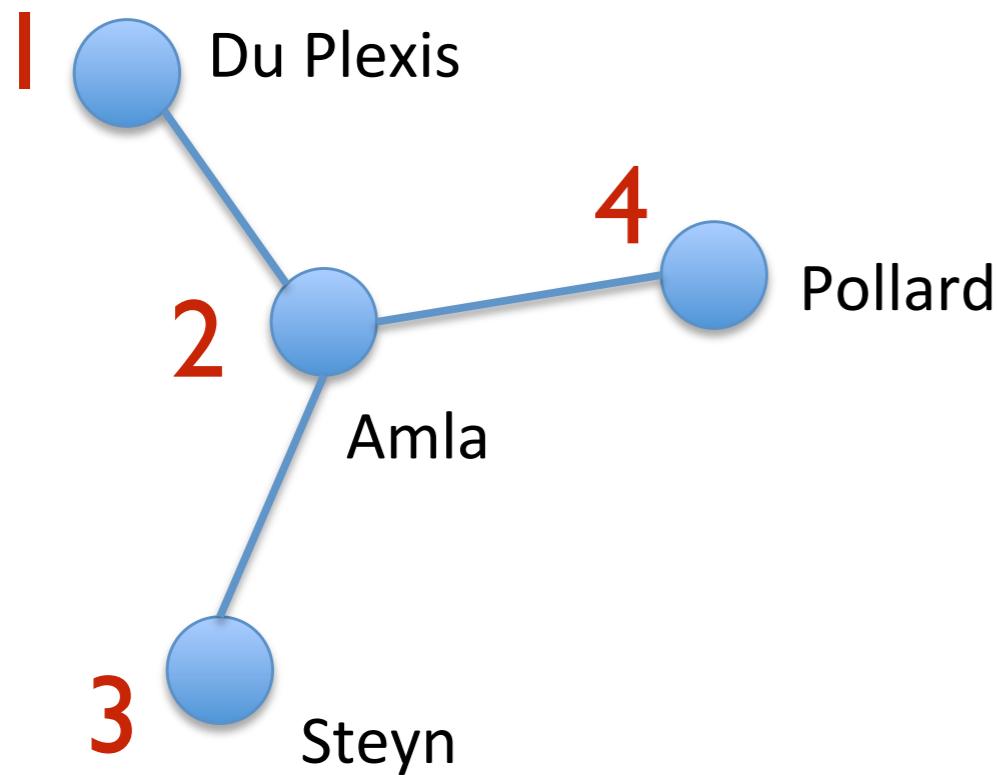
# Text to matrix

- Document- Word Matrix
- Document 1: “AACCBBAAA”
- Document 2: “CCAABBDD”

$$\begin{bmatrix} A & B & C & D \\ 5 & 2 & 2 & 0 \\ 2 & 2 & 2 & 2 \end{bmatrix}$$



# Network data



Nodes	Nodes	Nodes	Nodes
0	1	0	0
1	0	1	1
0	1	0	0
0	1	0	0



# Image data



[www.sydney.visitorsbureau.com.au](http://www.sydney.visitorsbureau.com.au)



700 x 500

4	45	6
6	12	33
22	17	44



4	45	6	6	12	33	22	17	44
---	----	---	---	----	----	----	----	----



THE UNIVERSITY OF  
SYDNEY

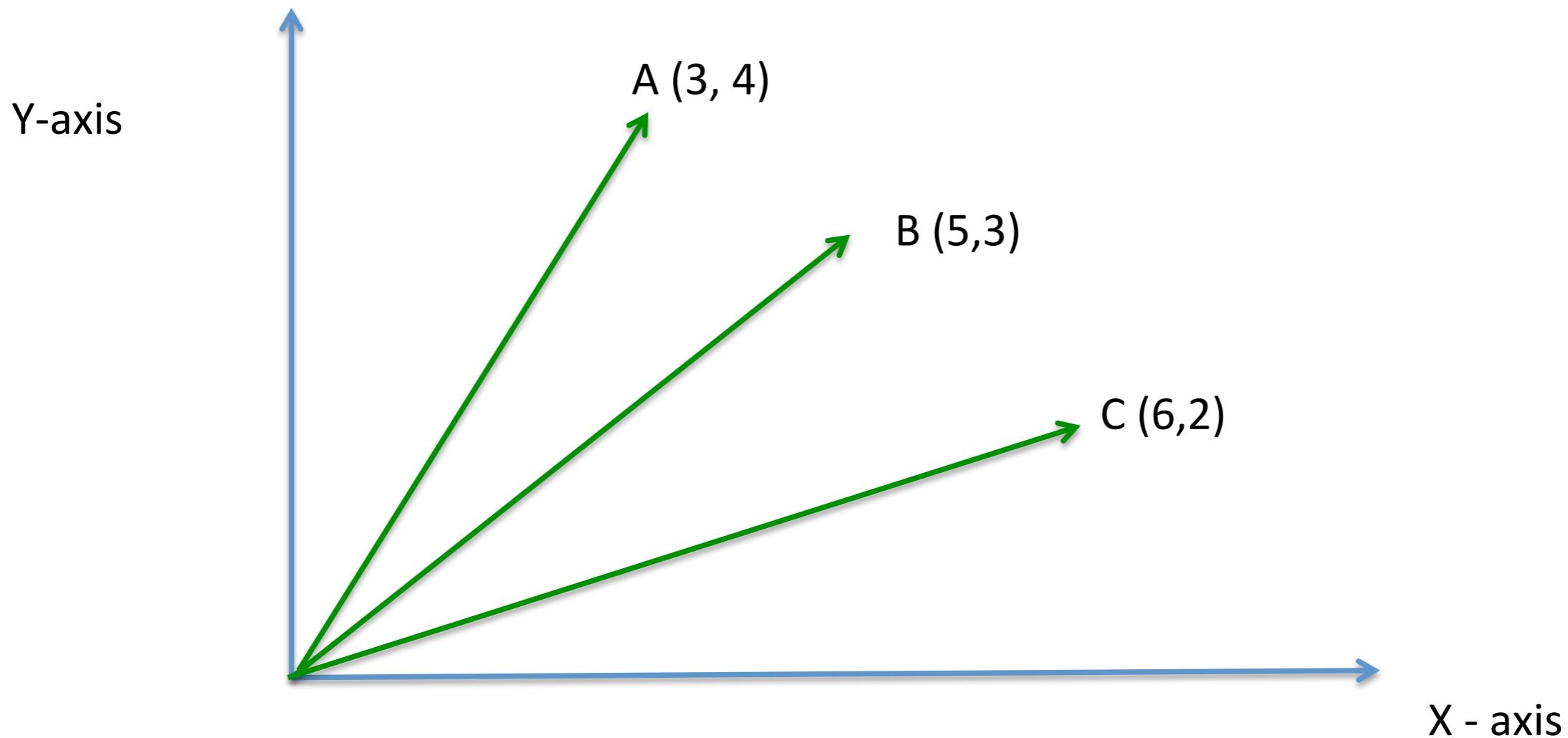
# Similarity Computation

- We can now represent most data types as a matrix.
- A special case of a matrix is a vector.
- Now lets compute similarities with these objects.



# Similarity Computation

How can we quantify similarity between A, B and C ?





# Similarity Computation

- Dot product

$$x = (x_1, x_2, \dots, x_n); \quad y = (y_1, y_2, \dots, y_n);$$

$$x.y = (x_1y_1 + x_2y_2 + \dots + x_ny_n);$$

- Norm (length) of a vector

$$\|x\| = (x.x)^{1/2} = (x_1.x_1 + x_2.x_2 + x_n.x_n)^{1/2}$$



THE UNIVERSITY OF  
SYDNEY

# Similarity Computation

- The similarity between two vectors  $x$  and  $y$  is given by

$$sim(x, y) = x \cdot y / (\|x\| \|y\|)$$



# Example

- Let  $x = \langle 3, 1, 2, 4 \rangle$ ,  $y = \langle 1, 2, 1, 2 \rangle$

- Step 1: Compute the dot-product

$$x \cdot y = 3 \cdot 1 + 1 \cdot 2 + 2 \cdot 1 + 4 \cdot 2 = 15$$

- Step 2: Compute length of x vector

$$\|x\| = (3^2 + 1^2 + 2^2 + 4^2)^{0.5} = 5.477$$

$$\|y\| = 3.162$$

$$sim(x, y) = x \cdot y / (\|x\| \|y\|) = 0.8660$$



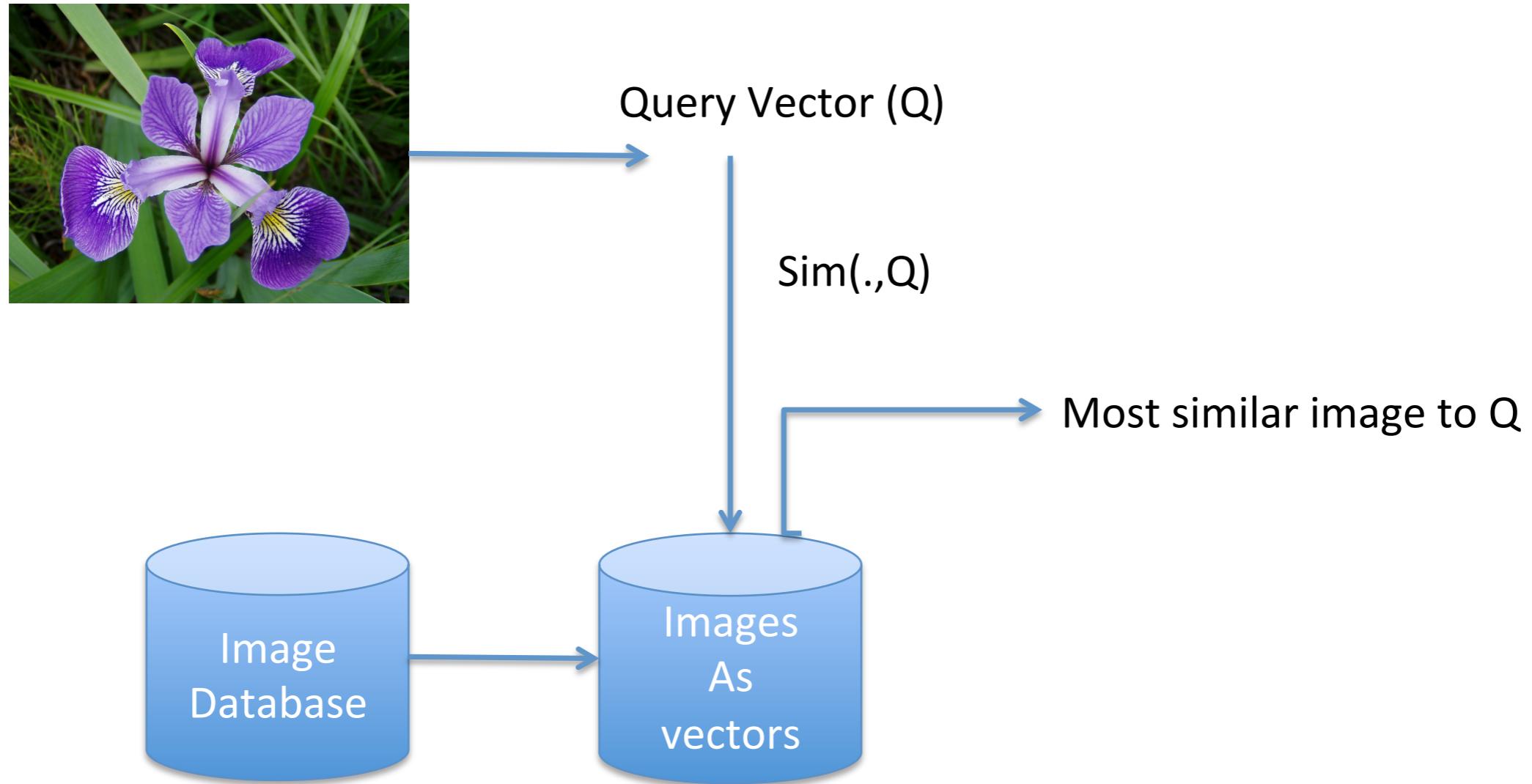
# Properties

- When is  $\text{sim}(x, y) = 0$  ?
- When is  $\text{sim}(x, y) = 1$  ?
- Can  $\text{sim}(x, y) < 0$  ?
- Can  $\text{sim}(x, y) > 1$  ?



THE UNIVERSITY OF  
SYDNEY

# Image search engine



# Object recognition

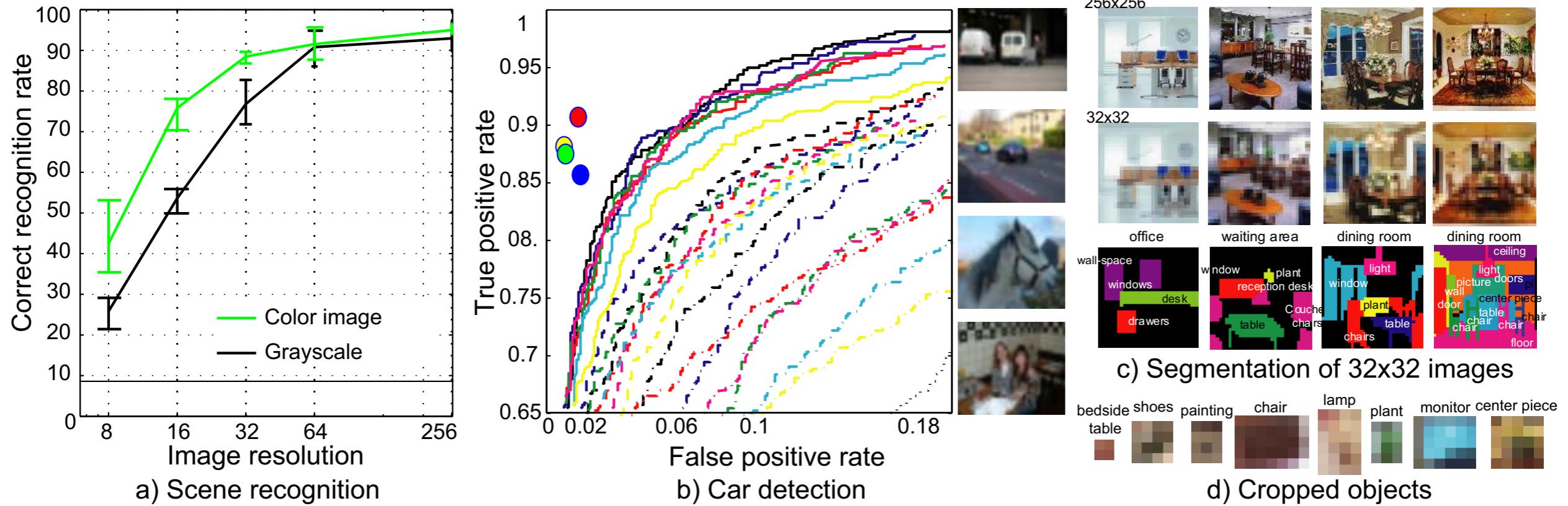
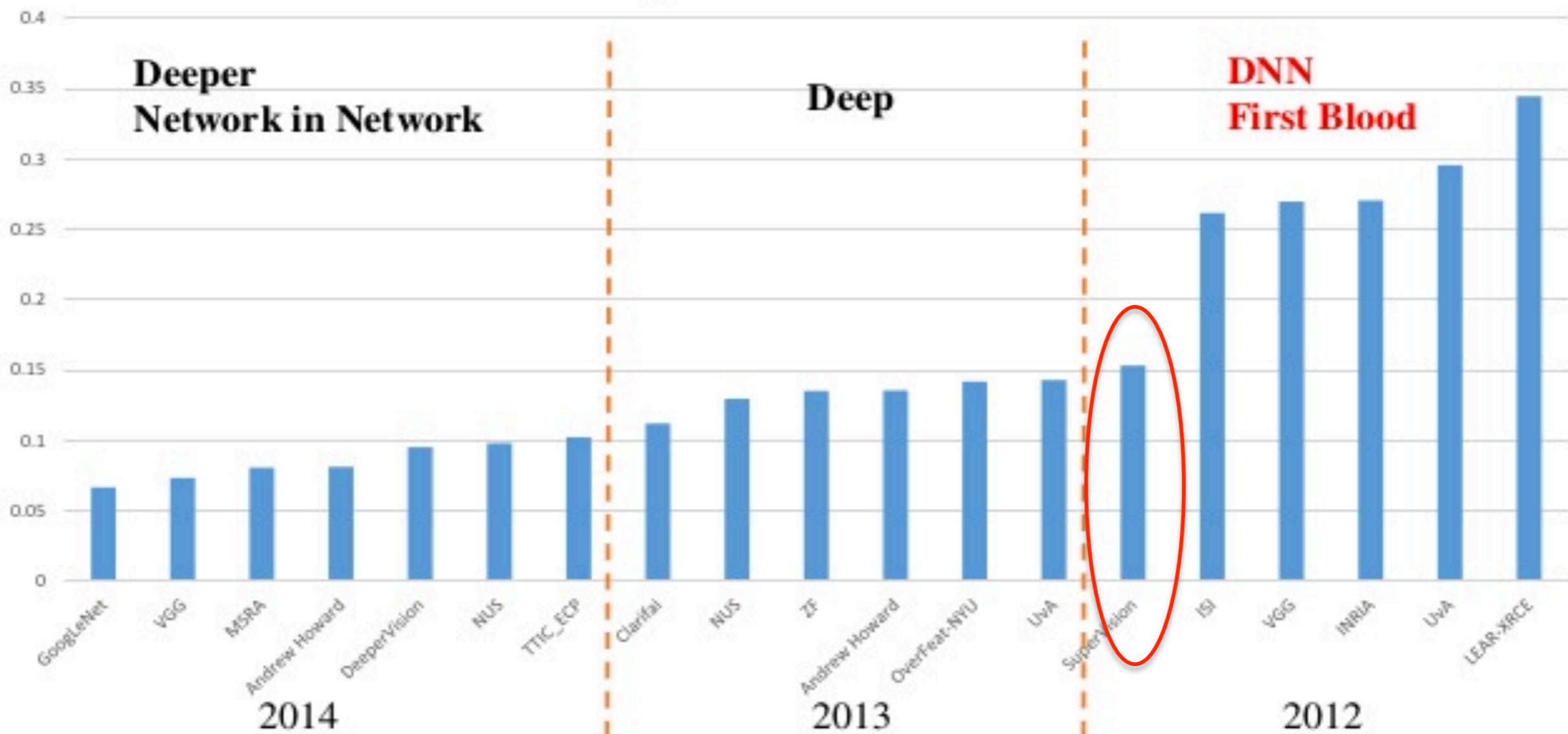


Fig. 1. a) Human performance on scene recognition as a function of resolution. The green and black curves show the performance on color and gray-scale images respectively. For color  $32 \times 32$  images the performance only drops by 7% relative to full resolution, despite having 1/64th of the pixels. b) Car detection task on the PASCAL 2006 test dataset. The colored dots show the performance of four human subjects classifying tiny versions of the test data. The ROC curves of the best vision algorithms (running on full resolution images) are shown for comparison. All lie below the performance of humans on the tiny images, which rely on none of the high-resolution cues exploited by the computer vision algorithms. c) Humans can correctly recognize and segment objects at very low resolutions, even when the objects in isolation can not be recognized (d).

Torralba et al. 80 million tiny images: a large dataset for non-parametric object and scene recognition, PAMI 2008

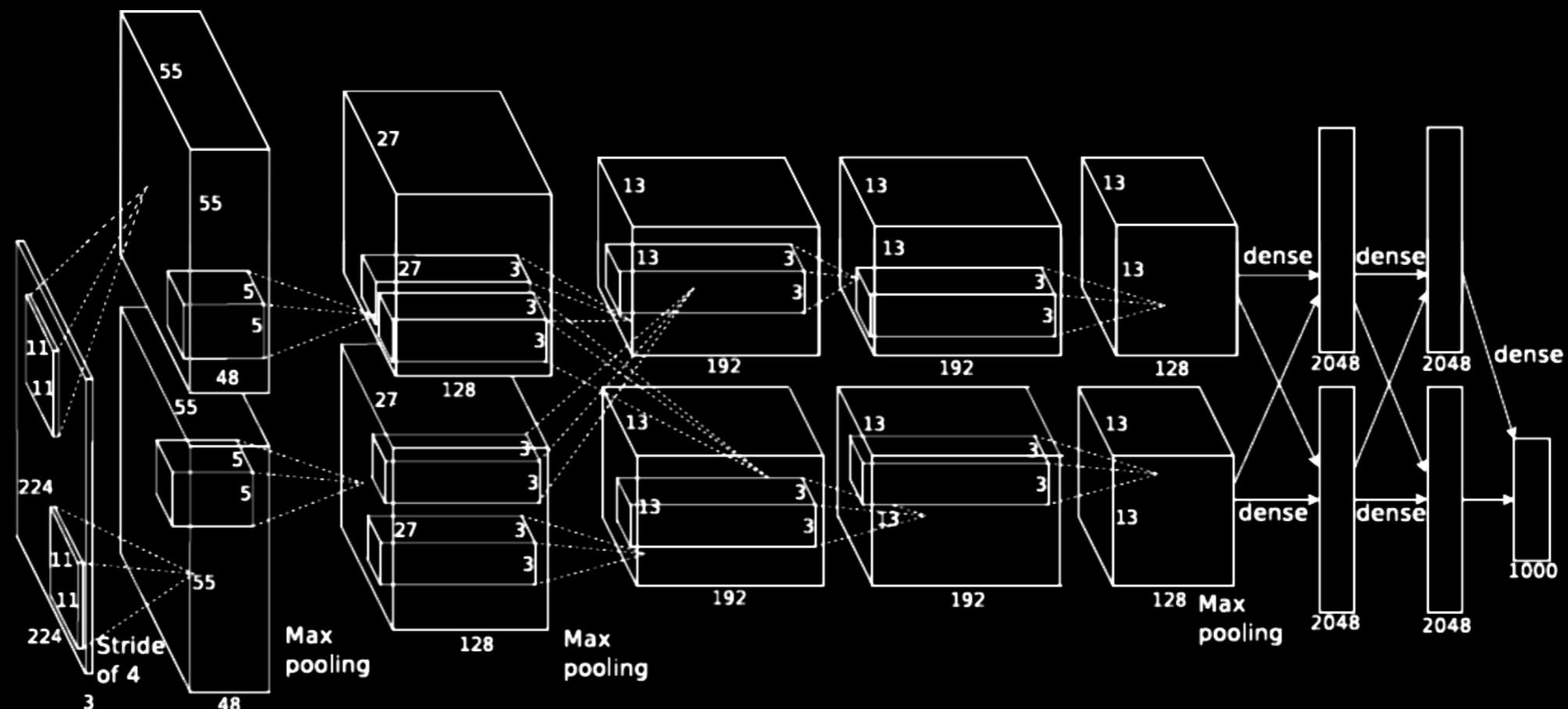
# ILSVRC

ImageNet Classification error throughout years and groups



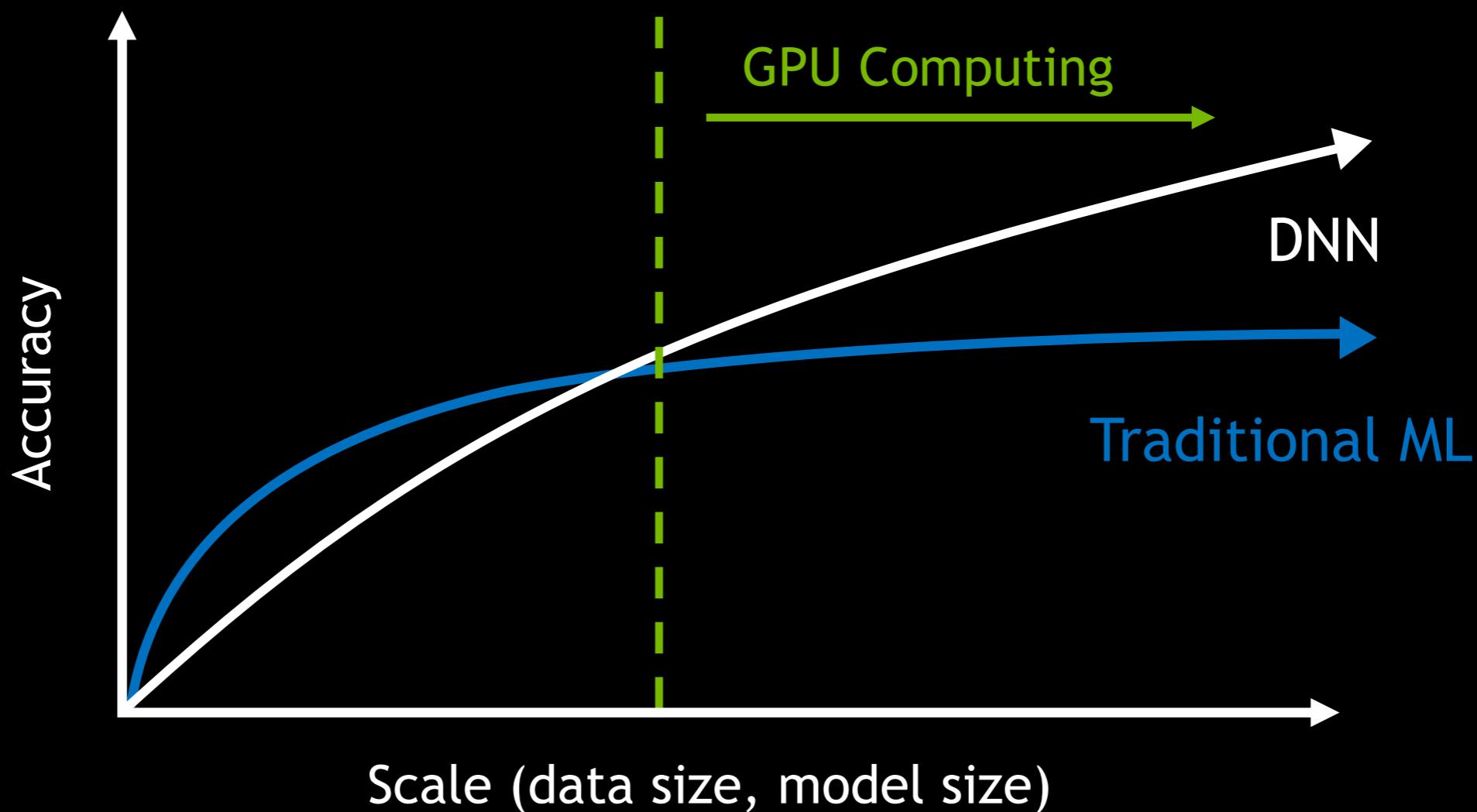
Li Fei-Fei: ImageNet Large Scale Visual Recognition Challenge, 2014 <http://image-net.org/>

# Deep Convolutional Neural Networks



AlexNet 2012

# The rise of Deep Learning



<https://blog.statsbot.co/deep-learning-achievements-4c563e034257>



THE UNIVERSITY OF  
SYDNEY

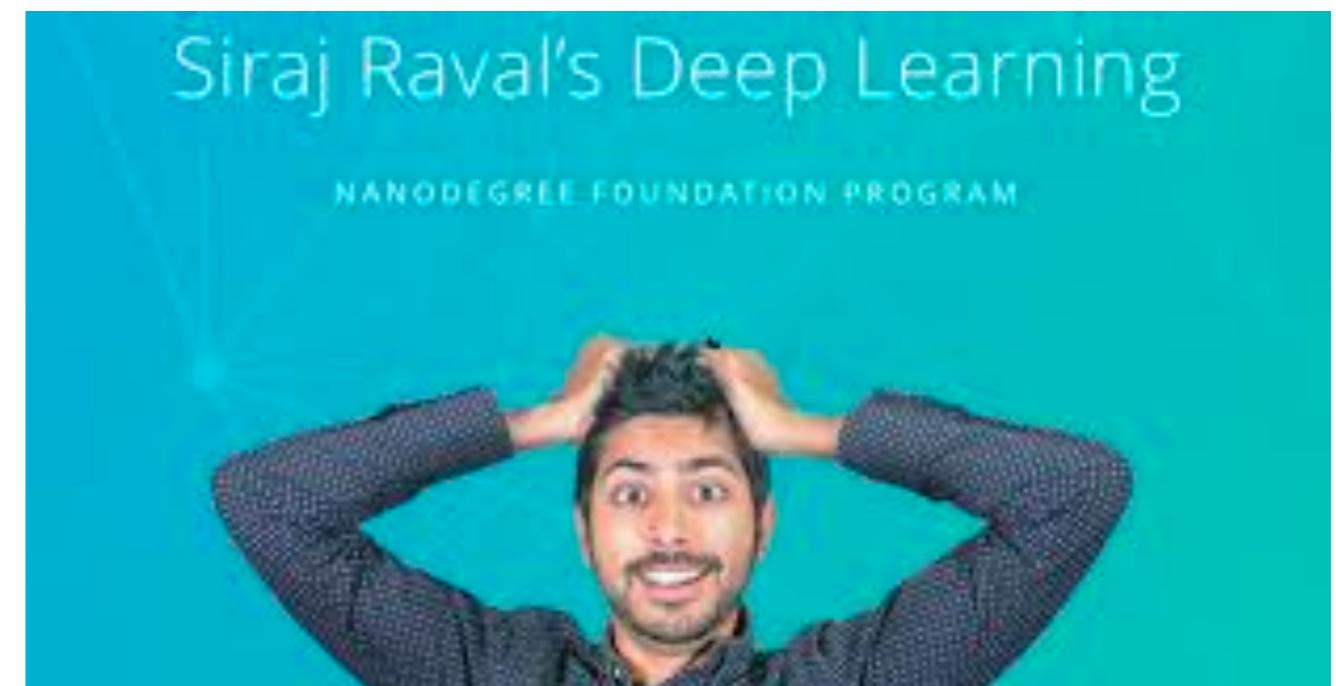
# Chatbots

Source: BBC Radio 4 - Analysis Podcast



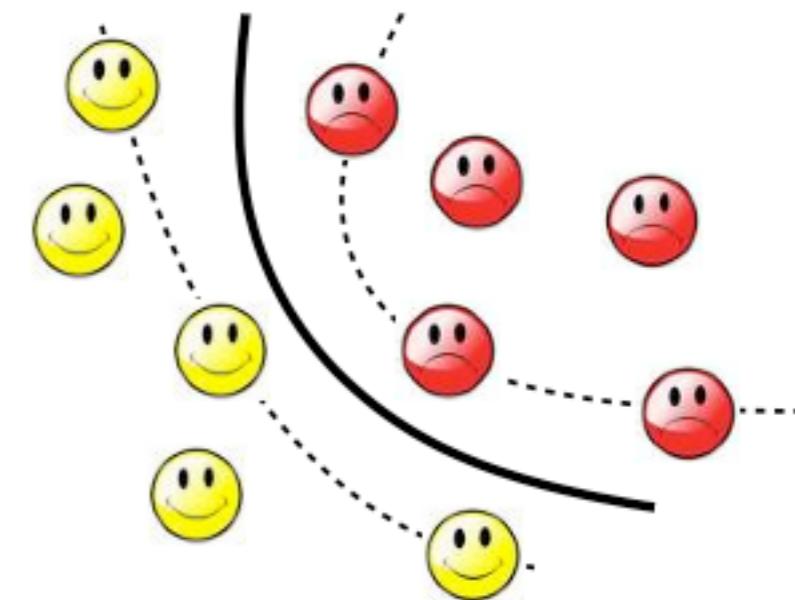
THE UNIVERSITY OF  
SYDNEY

# Youtube channels





THE UNIVERSITY OF  
SYDNEY



# My research