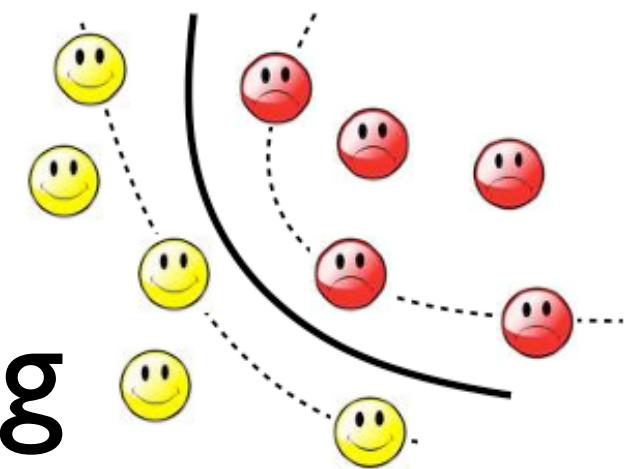


THE UNIVERSITY OF  
**SYDNEY**

# Machine Learning and Data Mining

## COMP 5318



School of Information Technologies

Fabio Ramos  
Lionel Ott  
Roman Marchant



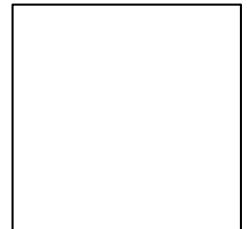
THE UNIVERSITY OF  
**SYDNEY**

# 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



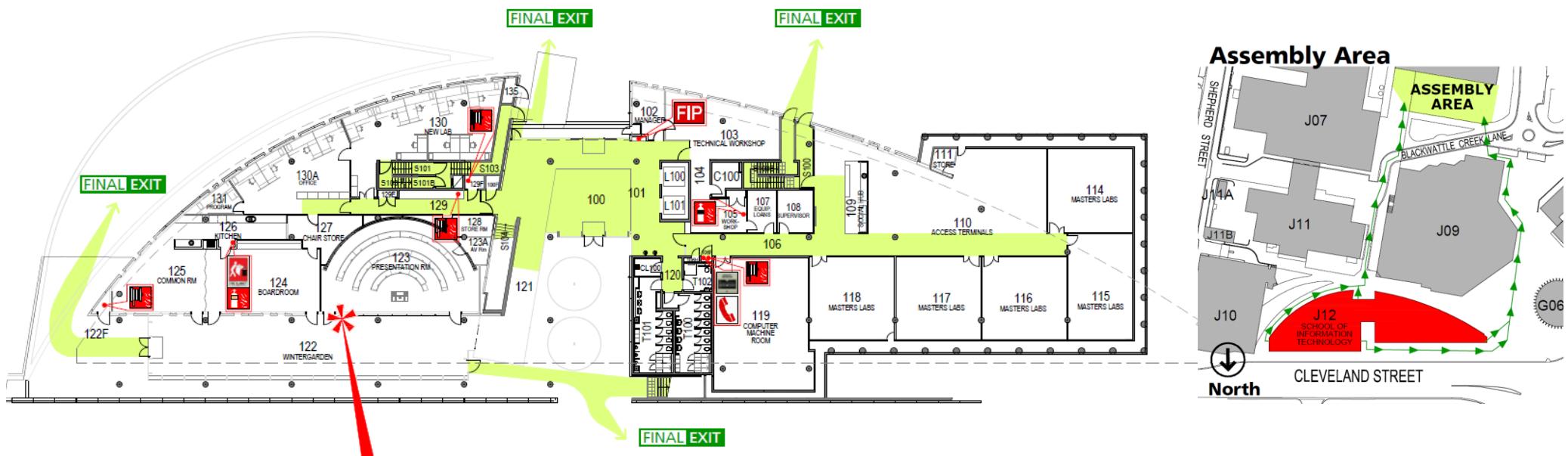
[www.sydney.edu.au/whs/emergency](http://www.sydney.edu.au/whs/emergency)

The screenshot shows the University of Sydney's Safety Health & Wellbeing website. At the top, there is a red header bar with the university crest and the text 'THE UNIVERSITY OF SYDNEY'. Below this is a blue header bar with the text 'SAFETY HEALTH & WELLBEING' and a search bar. A red arrow points to the URL 'www.sydney.edu.au/whs/emergency'.

The main content area has a white background. On the left, there is a sidebar titled 'EMERGENCY' with a list of links: '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)', and 'Mercury spills'. The main content area has a title 'WHAT TO DO IN AN EMERGENCY' and a paragraph about emergencies. It also lists various emergency situations with corresponding links: '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', and 'Suspicious behaviour'. At the bottom of this section is a link 'Be prepared'. On the right, there are two boxes: 'EMERGENCY CONTACT NUMBERS' which includes a section for 'POLICE, FIRE, AMBULANCE' with a note about dialing 0-000, and 'OTHER USEFUL NUMBERS' which includes links for 'University Security Service: 9351-3333', 'Chief fire wardens', and '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. **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



# School of IT Safety Contacts

## CHIEF WARDEN

Name: Greg Ryan  
Mobile:



## 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

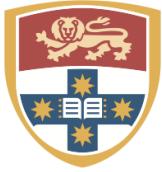
Postgraduates: to Cecille Faraizi  
9351 6060

SIT School Manager: Shari Lee  
9351 4158



# Places

- Lecture: every Monday, 6 to 8pm, in Institute Lecture Theatre 1 (H03)
- Labs: every Monday, 8 to 9pm, in SIT (Labs 115, 116, 117, 130b)  
every Tuesday, 5 to 6pm, in SIT (Labs 114, 115, 116, 118)
  - More labs to open soon
- 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)

[fabio.ramos@sydney.edu.au](mailto:fabio.ramos@sydney.edu.au)

Dr Lionel Ott (lecturer)

Dr Roman Marchant (lecturer)

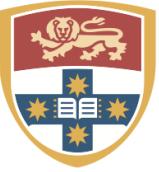
Tutors: Philippe Morere (TA) [philippe.morere@gmail.com](mailto:philippe.morere@gmail.com)

Rafael de Oliveira, Gilad Francis, Sheila

Carceres, Kelvin Hsu, Prasad Cheema,

Dongang Wang

Consultation: Monday, 2-3pm, SIT, Level 5, 547



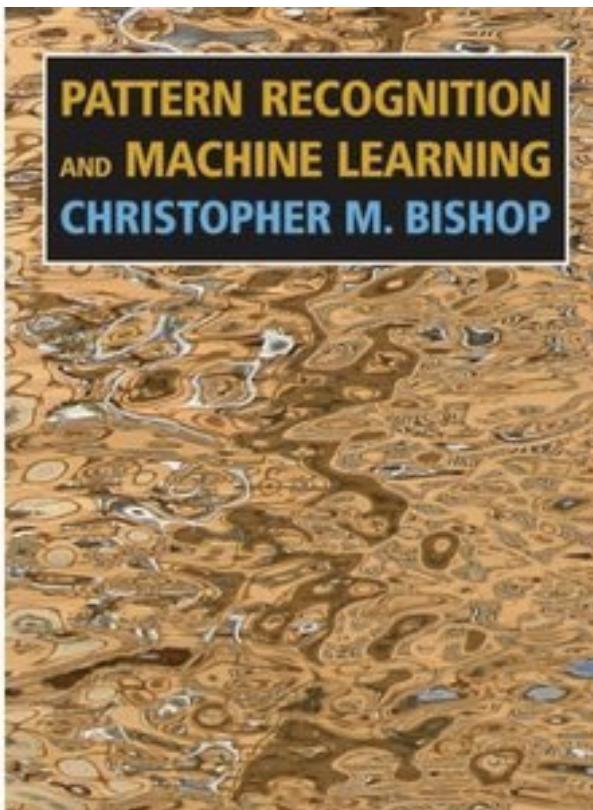
# Resources

- eLearning
  - 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 eLearning;*
  - see your grades; etc
  - Discussion forum: on edstem, link in eLearning site

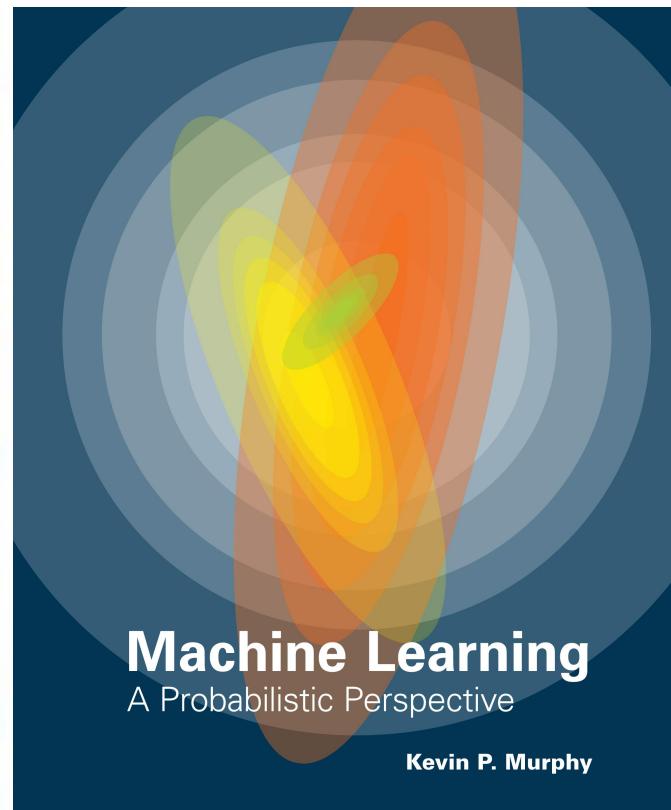


THE UNIVERSITY OF  
SYDNEY

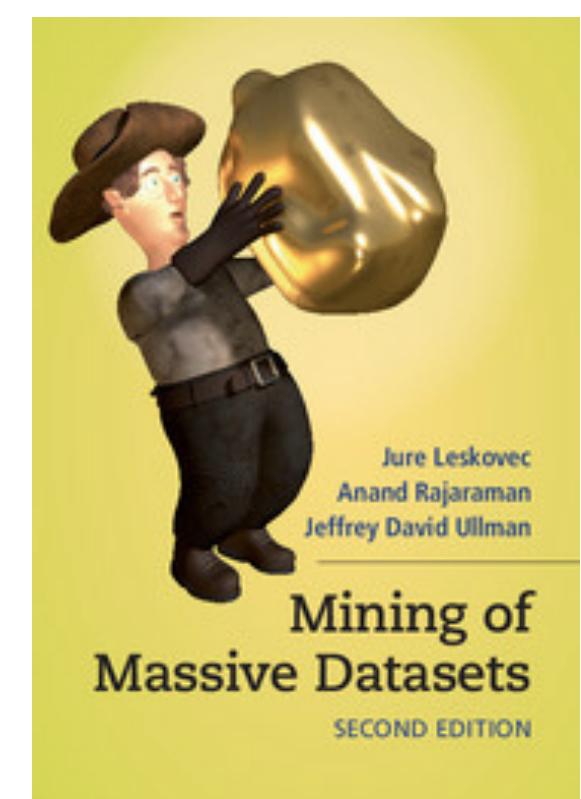
# Best books available



Textbook



Textbook



Extra



# Assessment overview

- Quiz: **10%**
  - Week 4 (27/03) in-class, 6-8pm
  - Individual
  - Linear algebra + probability theory
- Assignment 1: **20%**
  - **8 May 2017**, 5:00pm
  - Groups of 2 or 3 students
  - Classification task
- Assignment 2: **20%**
  - **5 June 2017**, 5:00pm
  - Groups of 2 or 3 students
  - Method comparison and analysis
- Final exam: **50%**
  - **June 2017 (date to be defined)**



# 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	6 March	Introduction, Admin matters, Research
2	13 March	Basic matrix analysis, SVD
3	20 March	Basic probability theory and Bayes' Rule
4	27 March	In-class Quiz
5	3 April	Classification and ROC curve
6	10 April	Logistic Regression and SVMs
7	24 April	Linear Regression
8	1 May	Clustering and EM
9	8 May	Latent Linear Models
10	15 May	Outlier detection
11	22 May	Neural Nets and large scale ML
12	29 May	Deep Learning
13	5 June	Industry talk and review



# 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



# Assumed knowledge

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



# 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)



# 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 centralized University system and database
- Three types of offenses:
  - **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!**



# Assistance

- There are a wide range of support services available for students
- 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 or TA, 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



# Disability Services

## Do you have a disability?

- You may not think of yourself as having a ‘disability’ but the definition under the **Disability Discrimination Act** 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, asperger's disorder, 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.
  - <http://sydney.edu.au/study/academic-support/disability-support.html>



# 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



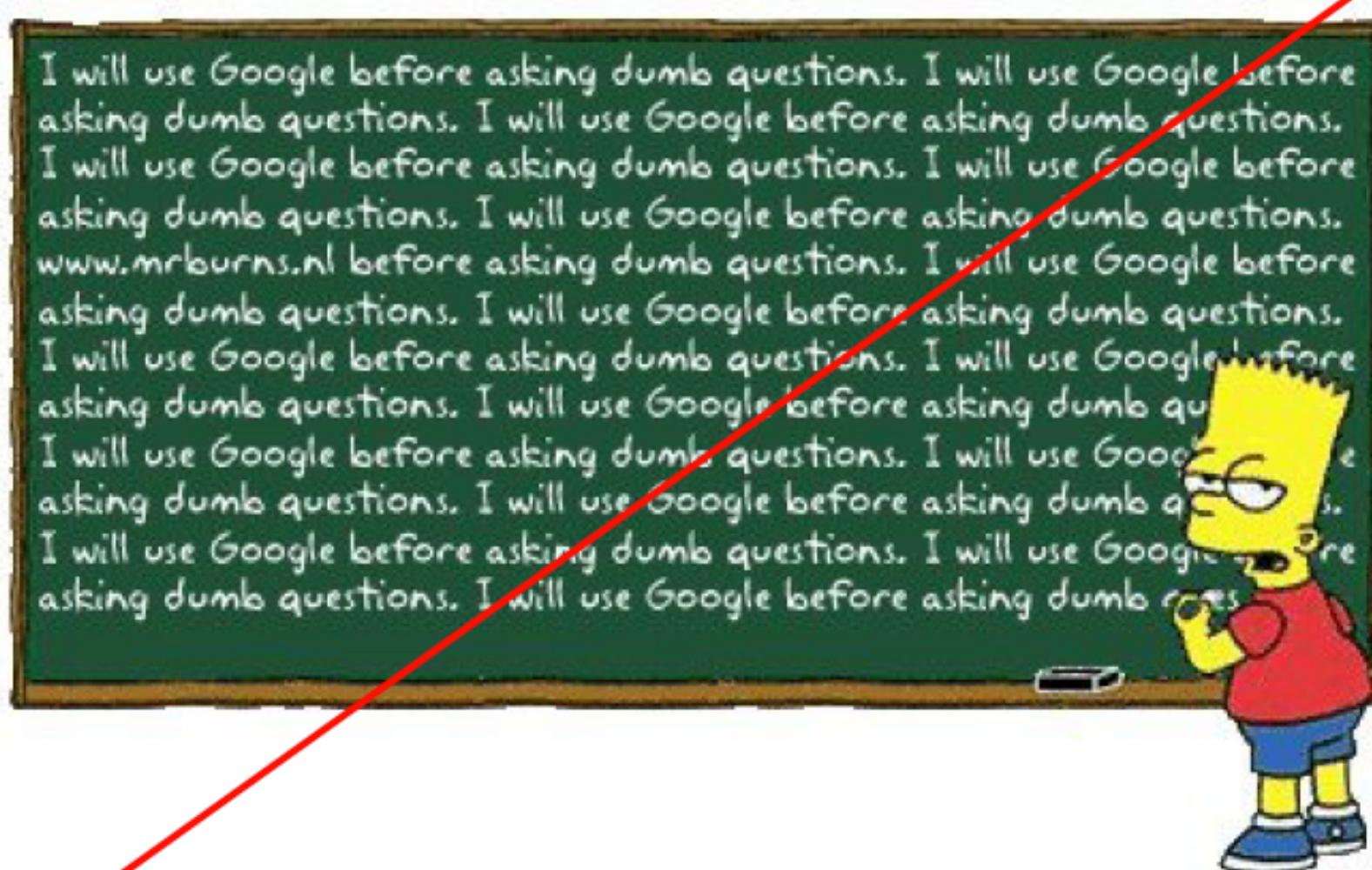
# Advice

- 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!



THE UNIVERSITY OF  
**SYDNEY**

# Ask questions sooner!





# What is Machine Learning?

**Informally:** Knowledge discovery from data

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

**\$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



THE UNIVERSITY OF  
SYDNEY



Data contains value and knowledge



# 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**

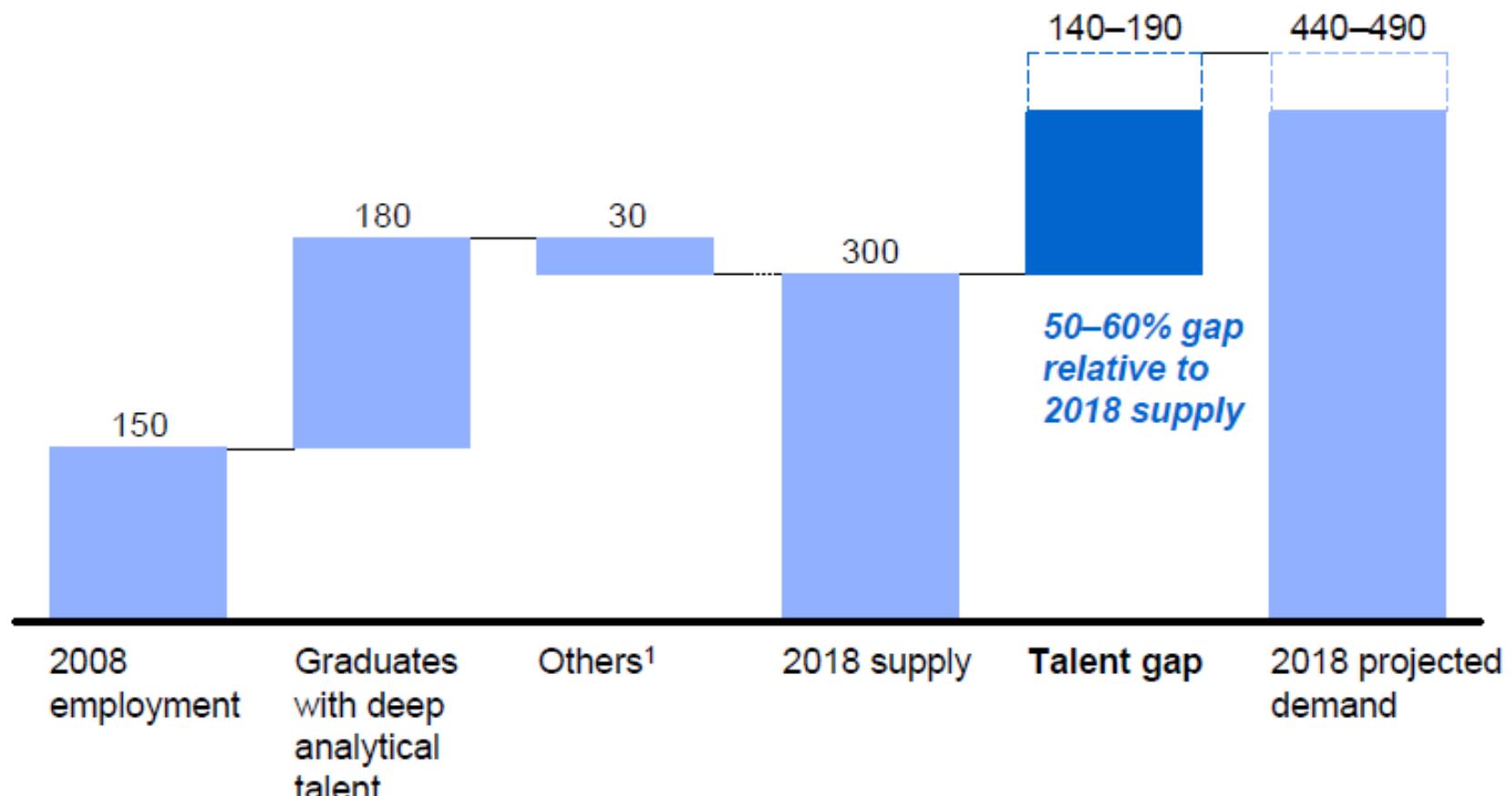


# Good news: Demand for Data Mining

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

**Google**™

**amazon.com**®



**SONY**

**SIEMENS**

**IBM**

Microsoft®

**Research**

**Research**

**facebook**®

**YAHOO!**®



**HRI**  
Honda Research Institute

**Commonwealth**Bank



**RioTinto**



THE UNIVERSITY OF  
SYDNEY

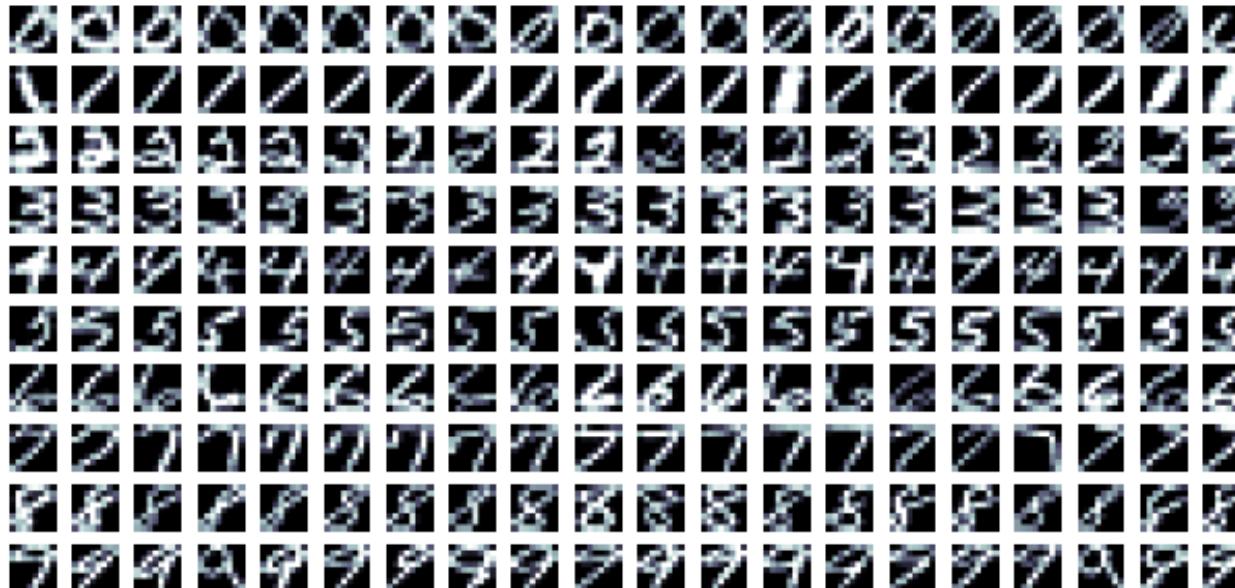
# Speech recognition





THE UNIVERSITY OF  
SYDNEY

# Object and handwriting recognition



(NORB image from Yann LeCun)



# Information retrieval

Google Search: Unsupervised Learning      http://www.google.com/search?q=Unsupervised+Learning&sourceid=f...

Web Images Groups News Books more...      Advanced Search Help

Google      Unsupervised Learning      Search      Advanced Search Help

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 Lowe's clustering, mixture modelling and unsupervised learning page. Mixture modelling (or ... www.csse.monash.edu.au/~dlm/mixture/modelling\_pg.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 SIGNL ... www.aclweb.org/~kohlenz/unsup-acl-99.html - 5k - Cached - Similar pages)

Unsupervised learning and Clustering  
cgm.cs.mcgill.ca/~soosi/cs644/projects/wlha/ - 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, Theories of Unsupervised Learning and Missing Values. ... www-2.cs.cmu.edu/~mcollum/nipsupsup/ - 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 - 4k - Cached - Similar pages)

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

ICPCL Unsupervised Learning of the Morphology of a Natural Language  
File Format: PDF/Mobi/Word/HTML/RTF/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, Page 20, Page 21, Page 22, Page 23, Page 24, Page 25, Page 26, Page 27, Page 28, Page 29, Page 30, Page 31, Page 32, Page 33, Page 34, Page 35, Page 36, Page 37, Page 38, Page 39, Page 40, Page 41, Page 42, Page 43, Page 44, Page 45, Page 46, Page 47, Page 48, Page 49, Page 50, Page 51, Page 52, Page 53, Page 54, Page 55, Page 56, Page 57, Page 58, Page 59, Page 60, Page 61, Page 62, Page 63, Page 64, Page 65, Page 66, Page 67, Page 68, Page 69, Page 70, Page 71, Page 72, Page 73, Page 74, Page 75, Page 76, Page 77, Page 78, Page 79, Page 80, Page 81, Page 82, Page 83, Page 84, Page 85, Page 86, Page 87, Page 88, Page 89, Page 90, Page 91, Page 92, Page 93, Page 94, Page 95, Page 96, Page 97, Page 98, Page 99, Page 100, Page 101, Page 102, Page 103, Page 104, Page 105, Page 106, Page 107, Page 108, Page 109, Page 110, Page 111, Page 112, Page 113, Page 114, Page 115, Page 116, Page 117, Page 118, Page 119, Page 120, Page 121, Page 122, Page 123, Page 124, Page 125, Page 126, Page 127, Page 128, Page 129, Page 130, Page 131, Page 132, Page 133, Page 134, Page 135, Page 136, Page 137, Page 138, Page 139, Page 140, Page 141, Page 142, Page 143, Page 144, Page 145, Page 146, Page 147, Page 148, Page 149, Page 150, Page 151, Page 152, Page 153, Page 154, Page 155, Page 156, Page 157, Page 158, Page 159, Page 160, Page 161, Page 162, Page 163, Page 164, Page 165, Page 166, Page 167, Page 168, Page 169, Page 170, Page 171, Page 172, Page 173, Page 174, Page 175, Page 176, Page 177, Page 178, Page 179, Page 180, Page 181, Page 182, Page 183, Page 184, Page 185, Page 186, Page 187, Page 188, Page 189, Page 190, Page 191, Page 192, Page 193, Page 194, Page 195, Page 196, Page 197, Page 198, Page 199, Page 200, Page 201, Page 202, Page 203, Page 204, Page 205, Page 206, Page 207, Page 208, Page 209, Page 210, Page 211, Page 212, Page 213, Page 214, Page 215, Page 216, Page 217, Page 218, Page 219, Page 220, Page 221, Page 222, Page 223, Page 224, Page 225, Page 226, Page 227, Page 228, Page 229, Page 230, Page 231, Page 232, Page 233, Page 234, Page 235, Page 236, Page 237, Page 238, Page 239, Page 240, Page 241, Page 242, Page 243, Page 244, Page 245, Page 246, Page 247, Page 248, Page 249, Page 250, Page 251, Page 252, Page 253, Page 254, Page 255, Page 256, Page 257, Page 258, Page 259, Page 259, Page 260, Page 261, Page 262, Page 263, Page 264, Page 265, Page 266, Page 267, Page 268, Page 269, Page 270, Page 271, Page 272, Page 273, Page 274, Page 275, Page 276, Page 277, Page 278, Page 279, Page 280, Page 281, Page 282, Page 283, Page 284, Page 285, Page 286, Page 287, Page 288, Page 289, Page 290, Page 291, Page 292, Page 293, Page 294, Page 295, Page 296, Page 297, Page 298, Page 299, Page 299, Page 300, Page 301, Page 302, Page 303, Page 304, Page 305, Page 306, Page 307, Page 308, Page 309, Page 310, Page 311, Page 312, Page 313, Page 314, Page 315, Page 316, Page 317, Page 318, Page 319, Page 320, Page 321, Page 322, Page 323, Page 324, Page 325, Page 326, Page 327, Page 328, Page 329, Page 330, Page 331, Page 332, Page 333, Page 334, Page 335, Page 336, Page 337, Page 338, Page 339, Page 340, Page 341, Page 342, Page 343, Page 344, Page 345, Page 346, Page 347, Page 348, Page 349, Page 350, Page 351, Page 352, Page 353, Page 354, Page 355, Page 356, Page 357, Page 358, Page 359, Page 359, Page 360, Page 361, Page 362, Page 363, Page 364, Page 365, Page 366, Page 367, Page 368, Page 369, Page 370, Page 371, Page 372, Page 373, Page 374, Page 375, Page 376, Page 377, Page 378, Page 379, Page 380, Page 381, Page 382, Page 383, Page 384, Page 385, Page 386, Page 387, Page 388, Page 389, Page 390, Page 391, Page 392, Page 393, Page 394, Page 395, Page 396, Page 397, Page 398, Page 399, Page 399, Page 400, Page 401, Page 402, Page 403, Page 404, Page 405, Page 406, Page 407, Page 408, Page 409, Page 410, Page 411, Page 412, Page 413, Page 414, Page 415, Page 416, Page 417, Page 418, Page 419, Page 420, Page 421, Page 422, Page 423, Page 424, Page 425, Page 426, Page 427, Page 428, Page 429, Page 430, Page 431, Page 432, Page 433, Page 434, Page 435, Page 436, Page 437, Page 438, Page 439, Page 440, Page 441, Page 442, Page 443, Page 444, Page 445, Page 446, Page 447, Page 448, Page 449, Page 450, Page 451, Page 452, Page 453, Page 454, Page 455, Page 456, Page 457, Page 458, Page 459, Page 459, Page 460, Page 461, Page 462, Page 463, Page 464, Page 465, Page 466, Page 467, Page 468, Page 469, Page 470, Page 471, Page 472, Page 473, Page 474, Page 475, Page 476, Page 477, Page 478, Page 479, Page 480, Page 481, Page 482, Page 483, Page 484, Page 485, Page 486, Page 487, Page 488, Page 489, Page 490, Page 491, Page 492, Page 493, Page 494, Page 495, Page 496, Page 497, Page 498, Page 499, Page 499, Page 500, Page 501, Page 502, Page 503, Page 504, Page 505, Page 506, Page 507, Page 508, Page 509, Page 509, Page 510, Page 511, Page 512, Page 513, Page 514, Page 515, Page 516, Page 517, Page 518, Page 519, Page 520, Page 521, Page 522, Page 523, Page 524, Page 525, Page 526, Page 527, Page 528, Page 529, Page 529, Page 530, Page 531, Page 532, Page 533, Page 534, Page 535, Page 536, Page 537, Page 538, Page 539, Page 539, Page 540, Page 541, Page 542, Page 543, Page 544, Page 545, Page 546, Page 547, Page 548, Page 549, Page 549, Page 550, Page 551, Page 552, Page 553, Page 554, Page 555, Page 556, Page 557, Page 558, Page 559, Page 559, Page 560, Page 561, Page 562, Page 563, Page 564, Page 565, Page 566, Page 567, Page 568, Page 569, Page 569, Page 570, Page 571, Page 572, Page 573, Page 574, Page 575, Page 576, Page 577, Page 578, Page 579, Page 579, Page 580, Page 581, Page 582, Page 583, Page 584, Page 585, Page 586, Page 587, Page 588, Page 589, Page 589, Page 590, Page 591, Page 592, Page 593, Page 594, Page 595, Page 596, Page 597, Page 598, Page 599, Page 599, Page 600, Page 601, Page 602, Page 603, Page 604, Page 605, Page 606, Page 607, Page 608, Page 609, Page 609, Page 610, Page 611, Page 612, Page 613, Page 614, Page 615, Page 616, Page 617, Page 618, Page 619, Page 619, Page 620, Page 621, Page 622, Page 623, Page 624, Page 625, Page 626, Page 627, Page 628, Page 629, Page 629, Page 630, Page 631, Page 632, Page 633, Page 634, Page 635, Page 636, Page 637, Page 638, Page 639, Page 639, Page 640, Page 641, Page 642, Page 643, Page 644, Page 645, Page 646, Page 647, Page 648, Page 649, Page 649, Page 650, Page 651, Page 652, Page 653, Page 654, Page 655, Page 656, Page 657, Page 658, Page 659, Page 659, Page 660, Page 661, Page 662, Page 663, Page 664, Page 665, Page 666, Page 667, Page 668, Page 669, Page 669, Page 670, Page 671, Page 672, Page 673, Page 674, Page 675, Page 676, Page 677, Page 678, Page 679, Page 679, Page 680, Page 681, Page 682, Page 683, Page 684, Page 685, Page 686, Page 687, Page 688, Page 689, Page 689, Page 690, Page 691, Page 692, Page 693, Page 694, Page 695, Page 696, Page 697, Page 698, Page 699, Page 699, Page 700, Page 701, Page 702, Page 703, Page 704, Page 705, Page 706, Page 707, Page 708, Page 709, Page 709, Page 710, Page 711, Page 712, Page 713, Page 714, Page 715, Page 716, Page 717, Page 718, Page 719, Page 719, Page 720, Page 721, Page 722, Page 723, Page 724, Page 725, Page 726, Page 727, Page 728, Page 729, Page 729, Page 730, Page 731, Page 732, Page 733, Page 734, Page 735, Page 736, Page 737, Page 738, Page 739, Page 739, Page 740, Page 741, Page 742, Page 743, Page 744, Page 745, Page 746, Page 747, Page 748, Page 749, Page 749, Page 750, Page 751, Page 752, Page 753, Page 754, Page 755, Page 756, Page 757, Page 758, Page 759, Page 759, Page 760, Page 761, Page 762, Page 763, Page 764, Page 765, Page 766, Page 767, Page 768, Page 769, Page 769, Page 770, Page 771, Page 772, Page 773, Page 774, Page 775, Page 776, Page 777, Page 778, Page 779, Page 779, Page 780, Page 781, Page 782, Page 783, Page 784, Page 785, Page 786, Page 787, Page 788, Page 789, Page 789, Page 790, Page 791, Page 792, Page 793, Page 794, Page 795, Page 796, Page 797, Page 798, Page 799, Page 799, Page 800, Page 801, Page 802, Page 803, Page 804, Page 805, Page 806, Page 807, Page 808, Page 809, Page 809, Page 810, Page 811, Page 812, Page 813, Page 814, Page 815, Page 816, Page 817, Page 818, Page 819, Page 819, Page 820, Page 821, Page 822, Page 823, Page 824, Page 825, Page 826, Page 827, Page 828, Page 829, Page 829, Page 830, Page 831, Page 832, Page 833, Page 834, Page 835, Page 836, Page 837, Page 838, Page 839, Page 839, Page 840, Page 841, Page 842, Page 843, Page 844, Page 845, Page 846, Page 847, Page 848, Page 849, Page 849, Page 850, Page 851, Page 852, Page 853, Page 854, Page 855, Page 856, Page 857, Page 858, Page 859, Page 859, Page 860, Page 861, Page 862, Page 863, Page 864, Page 865, Page 866, Page 867, Page 868, Page 869, Page 869, Page 870, Page 871, Page 872, Page 873, Page 874, Page 875, Page 876, Page 877, Page 878, Page 879, Page 879, Page 880, Page 881, Page 882, Page 883, Page 884, Page 885, Page 886, Page 887, Page 888, Page 889, Page 889, Page 890, Page 891, Page 892, Page 893, Page 894, Page 895, Page 896, Page 897, Page 898, Page 899, Page 899, Page 900, Page 901, Page 902, Page 903, Page 904, Page 905, Page 906, Page 907, Page 908, Page 909, Page 909, Page 910, Page 911, Page 912, Page 913, Page 914, Page 915, Page 916, Page 917, Page 918, Page 919, Page 919, Page 920, Page 921, Page 922, Page 923, Page 924, Page 925, Page 926, Page 927, Page 928, Page 929, Page 929, Page 930, Page 931, Page 932, Page 933, Page 934, Page 935, Page 936, Page 937, Page 938, Page 939, Page 939, Page 940, Page 941, Page 942, Page 943, Page 944, Page 945, Page 946, Page 947, Page 948, Page 949, Page 949, Page 950, Page 951, Page 952, Page 953, Page 954, Page 955, Page 956, Page 957, Page 958, Page 959, Page 959, Page 960, Page 961, Page 962, Page 963, Page 964, Page 965, Page 966, Page 967, Page 968, Page 969, Page 969, Page 970, Page 971, Page 972, Page 973, Page 974, Page 975, Page 976, Page 977, Page 978, Page 979, Page 979, Page 980, Page 981, Page 982, Page 983, Page 984, Page 985, Page 986, Page 987, Page 988, Page 989, Page 989, Page 990, Page 991, Page 992, Page 993, Page 994, Page 995, Page 996, Page 997, Page 998, Page 999, Page 999, Page 1000, Page 1001, Page 1002, Page 1003, Page 1004, Page 1005, Page 1006, Page 1007, Page 1008, Page 1009, Page 1009, Page 1010, Page 1011, Page 1012, Page 1013, Page 1014, Page 1015, Page 1016, Page 1017, Page 1018, Page 1019, Page 1019, Page 1020, Page 1021, Page 1022, Page 1023, Page 1024, Page 1025, Page 1026, Page 1027, Page 1028, Page 1029, Page 1029, Page 1030, Page 1031, Page 1032, Page 1033, Page 1034, Page 1035, Page 1036, Page 1037, Page 1038, Page 1039, Page 1039, Page 1040, Page 1041, Page 1042, Page 1043, Page 1044, Page 1045, Page 1046, Page 1047, Page 1048, Page 1049, Page 1049, Page 1050, Page 1051, Page 1052, Page 1053, Page 1054, Page 1055, Page 1056, Page 1057, Page 1058, Page 1059, Page 1059, Page 1060, Page 1061, Page 1062, Page 1063, Page 1064, Page 1065, Page 1066, Page 1067, Page 1068, Page 1069, Page 1069, Page 1070, Page 1071, Page 1072, Page 1073, Page 1074, Page 1075, Page 1076, Page 1077, Page 1078, Page 1079, Page 1079, Page 1080, Page 1081, Page 1082, Page 1083, Page 1084, Page 1085, Page 1086, Page 1087, Page 1088, Page 1089, Page 1089, Page 1090, Page 1091, Page 1092, Page 1093, Page 1094, Page 1095, Page 1096, Page 1097, Page 1098, Page 1099, Page 1099, Page 1100, Page 1101, Page 1102, Page 1103, Page 1104, Page 1105, Page 1106, Page 1107, Page 1108, Page 1109, Page 1109, Page 1110, Page 1111, Page 1112, Page 1113, Page 1114, Page 1115, Page 1116, Page 1117, Page 1118, Page 1119, Page 1119, Page 1120, Page 1121, Page 1122, Page 1123, Page 1124, Page 1125, Page 1126, Page 1127, Page 1128, Page 1129, Page 1129, Page 1130, Page 1131, Page 1132, Page 1133, Page 1134, Page 1135, Page 1136, Page 1137, Page 1138, Page 1139, Page 1139, Page 1140, Page 1141, Page 1142, Page 1143, Page 1144, Page 1145, Page 1146, Page 1147, Page 1148, Page 1149, Page 1149, Page 1150, Page 1151, Page 1152, Page 1153, Page 1154, Page 1155, Page 1156, Page 1157, Page 1158, Page 1159, Page 1159, Page 1160, Page 1161, Page 1162, Page 1163, Page 1164, Page 1165, Page 1166, Page 1167, Page 1168, Page 1169, Page 1169, Page 1170, Page 1171, Page 1172, Page 1173, Page 1174, Page 1175, Page 1176, Page 1177, Page 1178, Page 1179, Page 1179, Page 1180, Page 1181, Page 1182, Page 1183, Page 1184, Page 1185, Page 1186, Page 1187, Page 1188, Page 1189, Page 1189, Page 1190, Page 1191, Page 1192, Page 1193, Page 1194, Page 1195, Page 1196, Page 1197, Page 1198, Page 1199, Page 1199, Page 1200, Page 1201, Page 1202, Page 1203, Page 1204, Page 1205, Page 1206, Page 1207, Page 1208, Page 1209, Page 1209, Page 1210, Page 1211, Page 1212, Page 1213, Page 1214, Page 1215, Page 1216, Page 1217, Page 1218, Page 1219, Page 1219, Page 1220, Page 1221, Page 1222, Page 1223, Page 1224, Page 1225, Page 1226, Page 1227, Page 1228, Page 1229, Page 1229, Page 1230, Page 1231, Page 1232, Page 1233, Page 1234, Page 1235, Page 1236, Page 1237, Page 1238, Page 1239, Page 1239, Page 1240, Page 1241, Page 1242, Page 1243, Page 1244, Page 1245, Page 1246, Page 1247, Page 1248, Page 1249, Page 1249, Page 1250, Page 1251, Page 1252, Page 1253, Page 1254, Page 1255, Page 1256, Page 1257, Page 1258, Page 1259, Page 1259, Page 1260, Page 1261, Page 1262, Page 1263, Page 1264, Page 1265, Page 1266, Page 1267, Page 1268, Page 1269, Page 1269, Page 1270, Page 1271, Page 1272, Page 1273, Page 1274, Page 1275, Page 1276, Page 1277, Page 1278, Page 1279, Page 1279, Page 1280, Page 1281, Page 1282, Page 1283, Page 1284, Page 1285, Page 1286, Page 1287, Page 1288, Page 1289, Page 1289, Page 1290, Page 1291, Page 1292, Page 1293, Page 1294, Page 1295, Page 1296, Page 1297, Page 1298, Page 1299, Page 1299, Page 1300, Page 1301, Page 1302, Page 1303, Page 1304, Page 1305, Page 1306, Page 1307, Page 1308, Page 1309, Page 1309, Page 1310, Page 1311, Page 1312, Page 1313, Page 1314, Page 1315, Page 1316, Page 1317, Page 1318, Page 1319, Page 1319, Page 1320, Page 1321, Page 1322, Page 1323, Page 1324, Page 1325, Page 1326, Page 1327, Page 1328, Page 1329, Page 1329, Page 1330, Page 1331, Page 1332, Page 1333, Page 1334, Page 1335, Page 1336, Page 1337, Page 1338, Page 1339, Page 1339, Page 1340, Page 1341, Page 1342, Page 1343, Page 1344, Page 1345, Page 1346, Page 1347, Page 1348, Page 1349, Page 1349, Page 1350, Page 1351, Page 1352, Page 1353, Page 1354, Page 1355, Page 1356, Page 1357, Page 1358, Page 1359, Page 1359, Page 1360, Page 1361, Page 1362, Page 1363, Page 1364, Page 1365, Page 1366, Page 1367, Page 1368, Page 1369, Page 1369, Page 1370, Page 1371, Page 1372, Page 1373, Page 1374, Page 1375, Page 1376, Page 1377, Page 1378, Page 1379, Page 1379, Page 1380, Page 1381, Page 1382, Page 1383, Page 1384, Page 1385, Page 1386, Page 1387, Page 1388, Page 1389, Page 1389, Page 1390, Page 1391, Page 1392, Page 1393, Page 1394, Page 1395, Page 1396, Page 1397, Page 1398, Page 1399, Page 1399, Page 1400, Page 1401, Page 1402, Page 1403, Page 1404, Page 1405, Page 1406, Page 1407, Page 1408, Page 1409, Page 1409, Page 1410, Page 1411, Page 1412, Page 1413, Page 1414, Page 1415, Page 1416, Page 1417, Page 1418, Page 1419, Page 1419, Page 1420, Page 1421, Page 1422, Page 1423, Page 1424, Page 1425, Page 1426, Page 1427, Page 1428, Page 1429, Page 1429, Page 1430, Page 1431, Page 1432, Page 1433, Page 1434, Page 1435, Page 1436, Page 1437, Page 1438, Page 1439, Page 1439, Page 1440, Page 1441, Page 1442, Page 1443, Page 1444, Page 1445, Page 1446, Page 1447, Page 1448, Page 1449, Page 1449, Page 1450, Page 1451, Page 1452, Page 1453, Page 1454, Page 1455, Page 1456, Page 1457, Page 1458, Page 1459, Page 1459, Page 1460, Page 1461, Page 1462, Page 1463, Page 1464, Page 1465, Page 1466, Page 1467, Page 1468, Page 1469, Page 1469, Page 1470, Page 1471, Page 1472, Page 1473, Page 1474, Page 1475, Page 1476, Page 1477, Page 1478, Page 1479, Page 1479, Page 1480, Page 1481, Page 1482, Page 1483, Page 1484, Page 1485, Page 1486, Page 1487, Page 1488, Page 1489, Page 1489, Page 1490, Page 1491, Page 1492, Page 1493, Page 1494, Page 1495, Page 1496, Page 1497, Page 1498, Page 1499, Page 1499, Page 1500, Page 1501, Page 1502, Page 1503, Page 1504, Page 1505, Page 1506, Page 1507, Page 1508, Page 1509, Page 1509, Page 1510, Page 1511, Page 1512, Page 1513, Page 1514, Page 1515, Page 1516, Page 1517, Page 1518, Page 1519, Page 1519, Page 1520, Page 1521, Page 1522, Page 1523, Page 1524, Page 1525, Page 1526, Page 1527, Page 1528, Page 1529, Page 1529, Page 1530, Page 1531, Page 1532, Page 1533, Page 1534, Page 1535, Page 1536, Page 1537, Page 1538, Page 1539, Page 1539, Page 1540, Page 1541, Page 1542, Page 1543, Page 1544, Page 154



THE UNIVERSITY OF  
SYDNEY

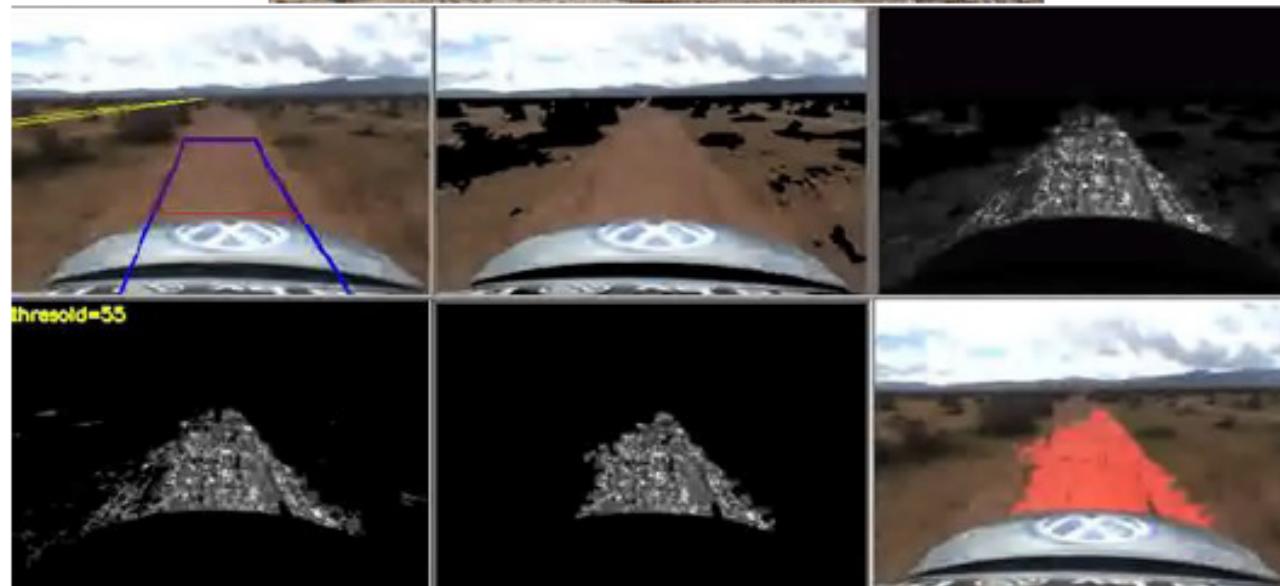
# Financial prediction





THE UNIVERSITY OF  
SYDNEY

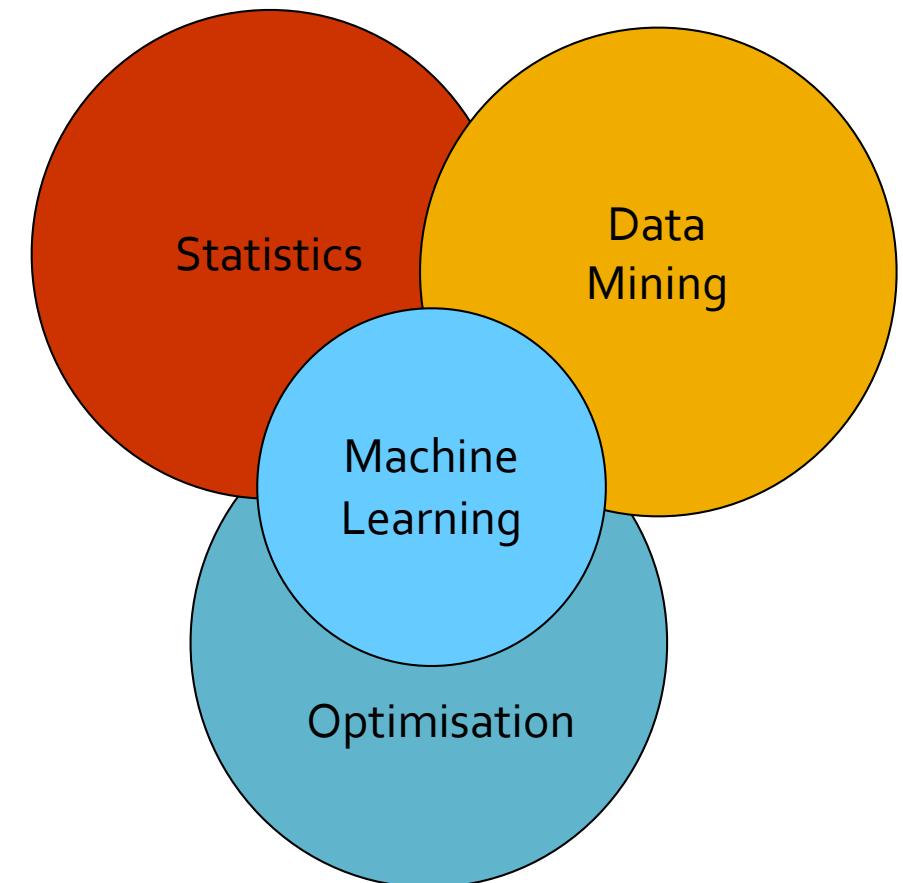
# Robotics





# This Course: COMP5318

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





# Machine Learning Problems

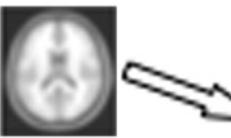
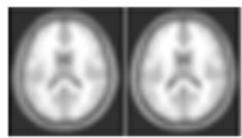
- Prediction
  - Classification and Regression
- Clustering, segmentation and summarisation
  - Find patterns in the data
- Outlier/anomaly detection
  - Find unusual patterns



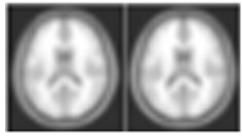
THE UNIVERSITY OF  
SYDNEY

# Elements of Machine Learning

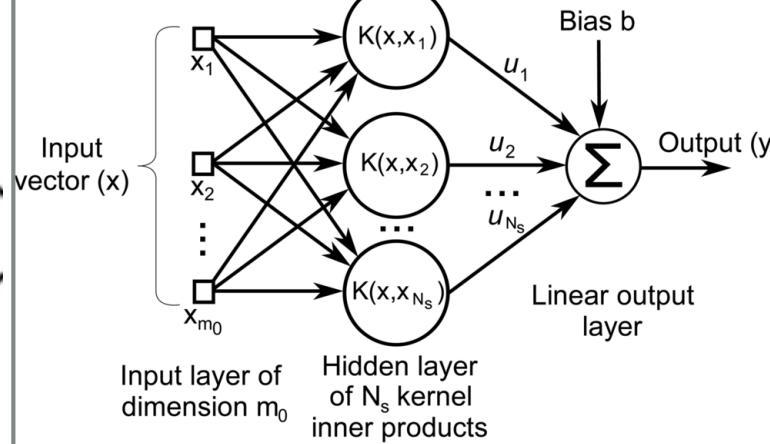
Group 1



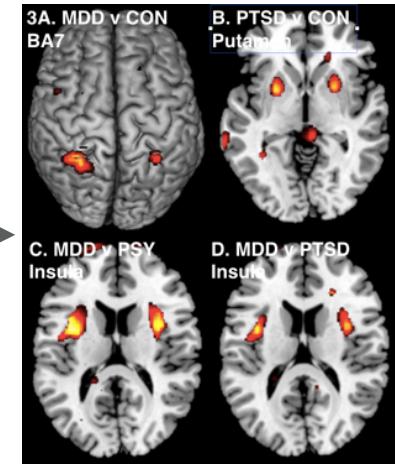
Group 2



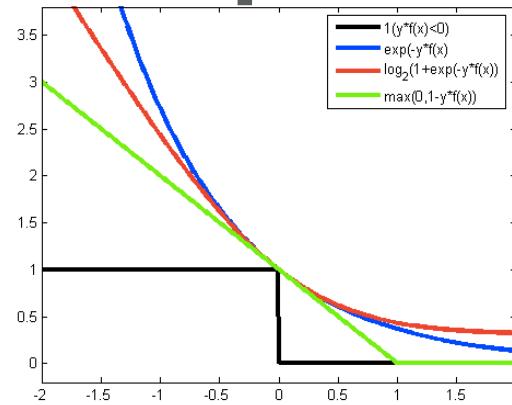
Mathematical Model



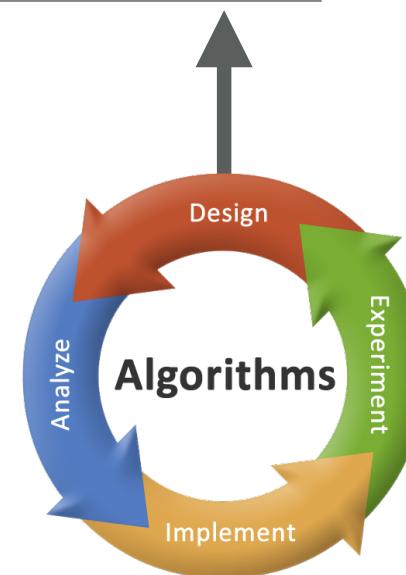
Predictions/Patterns



Data



Objective function





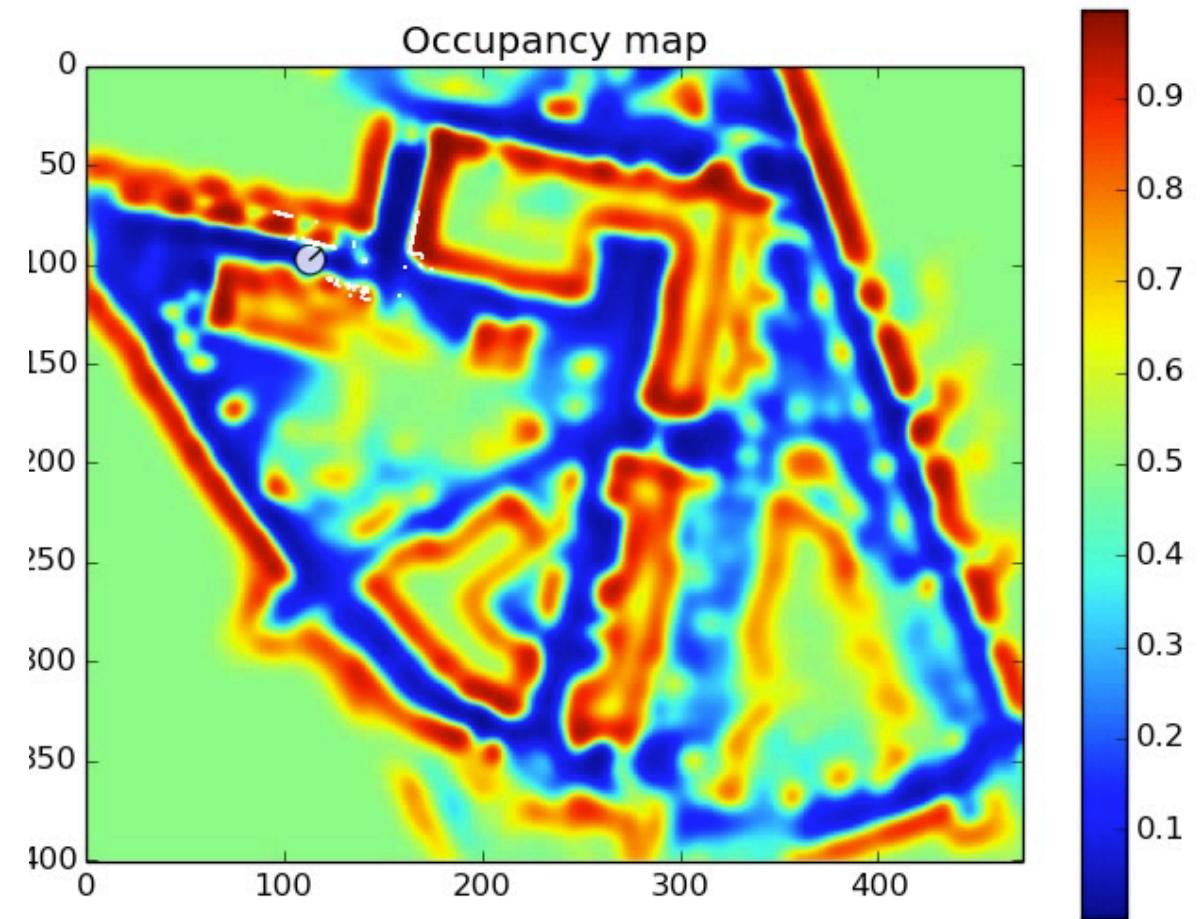
THE UNIVERSITY OF  
SYDNEY

# Classification

Aerial View



Occupancy map

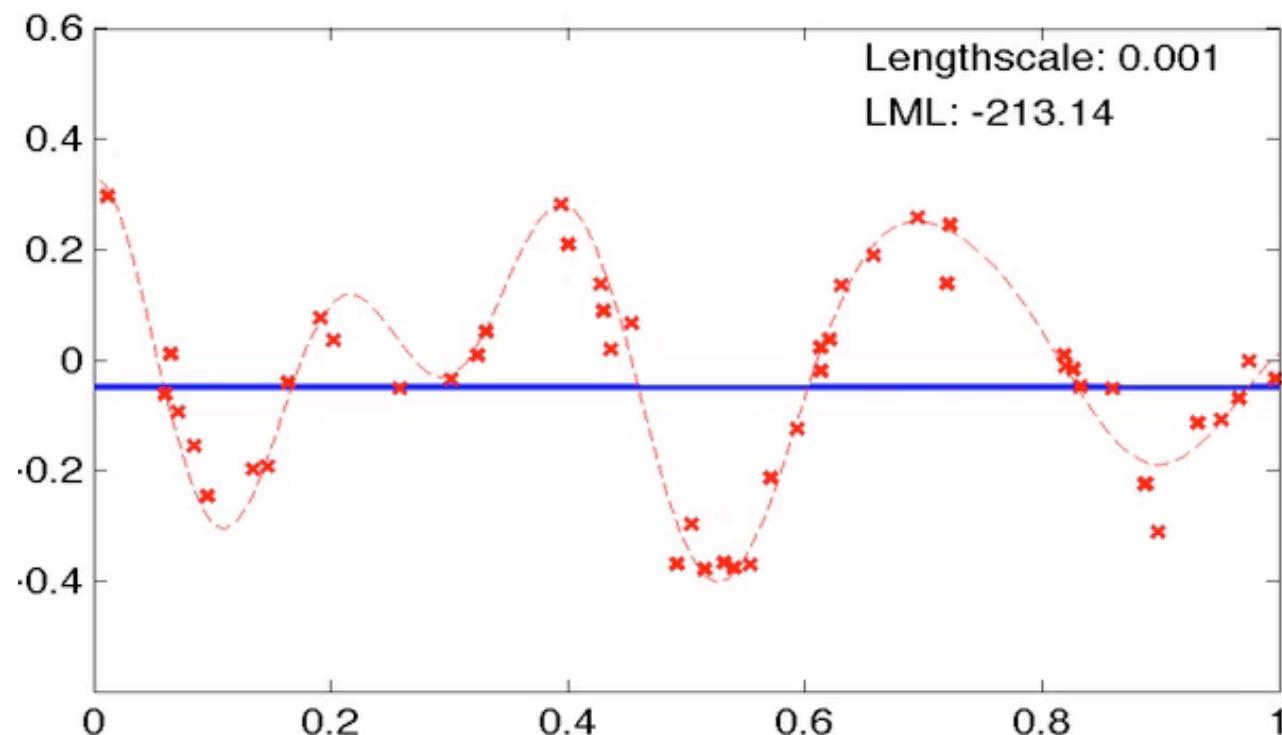




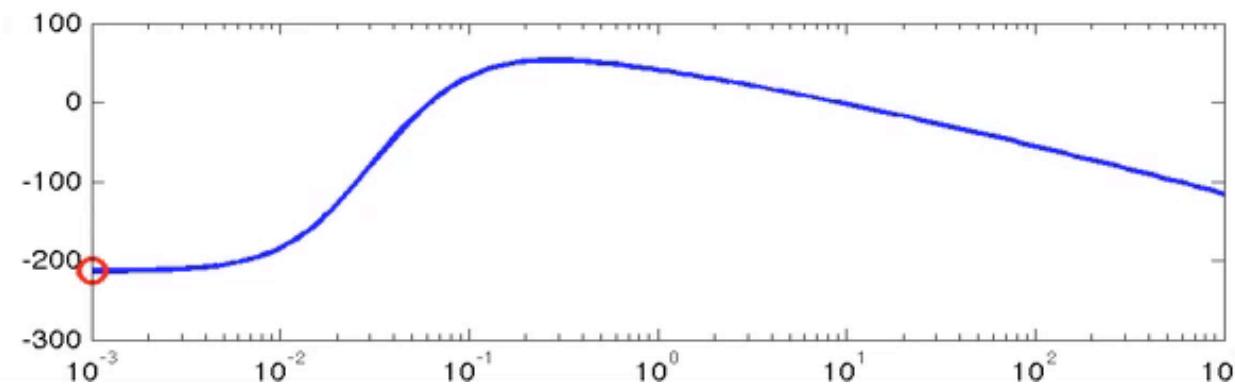
THE UNIVERSITY OF  
SYDNEY

# Regression

Problem



Objective





THE UNIVERSITY OF  
SYDNEY



# Neural Networks

Iterations  
000,000

Learning rate  
0.03

Activation  
ReLU

Regularization  
L2

Regularization rate  
0.001

Problem type  
Regression

## DATA

Which dataset do you want to use?



Ratio of training to test data: 40%

Noise: 20

Batch size: 10

**REGENERATE**

## FEATURES

Which properties do you want to feed in?

- $x_1$
- $x_2$
- $x_1^2$
- $x_2^2$
- $x_1 x_2$
- $\sin(x_1)$
- $\sin(x_2)$

## 3 HIDDEN LAYERS

+

-

+

-

+

-

4 neurons

4 neurons

2 neurons

$x_1$

$x_1$

$x_1$

$x_2$

$x_2$

$x_2$

$x_1^2$

$x_1^2$

$x_1^2$

$x_2^2$

$x_2^2$

$x_2^2$

$x_1 x_2$

$x_1 x_2$

$x_1 x_2$

$\sin(x_1)$

$\sin(x_1)$

$\sin(x_1)$

$\sin(x_2)$

$\sin(x_2)$

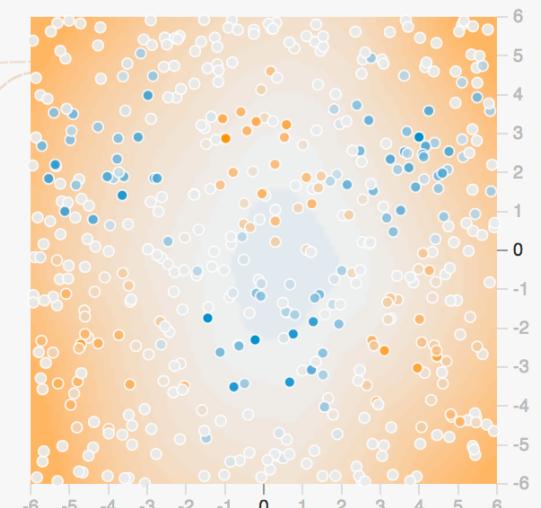
$\sin(x_2)$

This is the output from one neuron. Hover to see it larger.

The outputs are mixed with varying weights, shown by the thickness of the lines.

## OUTPUT

Test loss 0.121  
Training loss 0.134



Colors shows data, neuron and weight values.



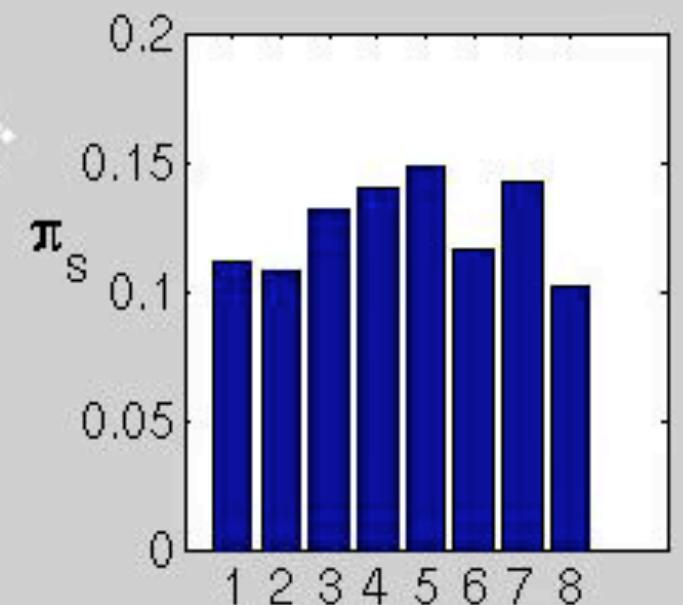
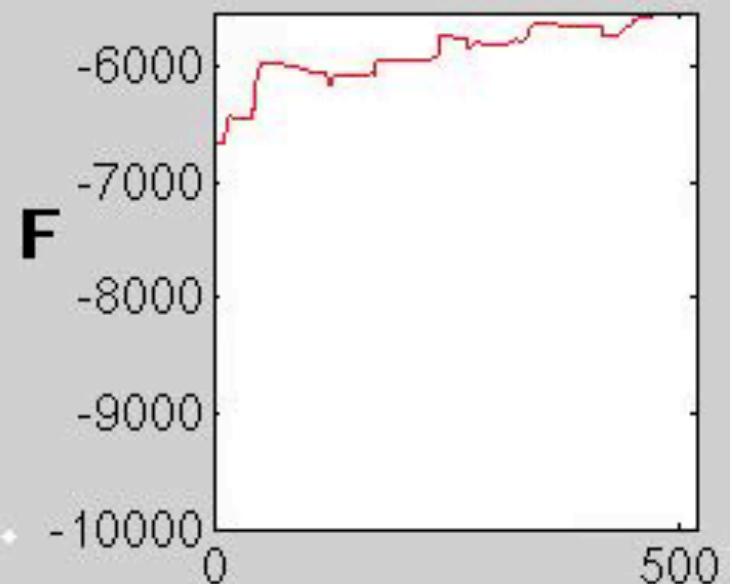
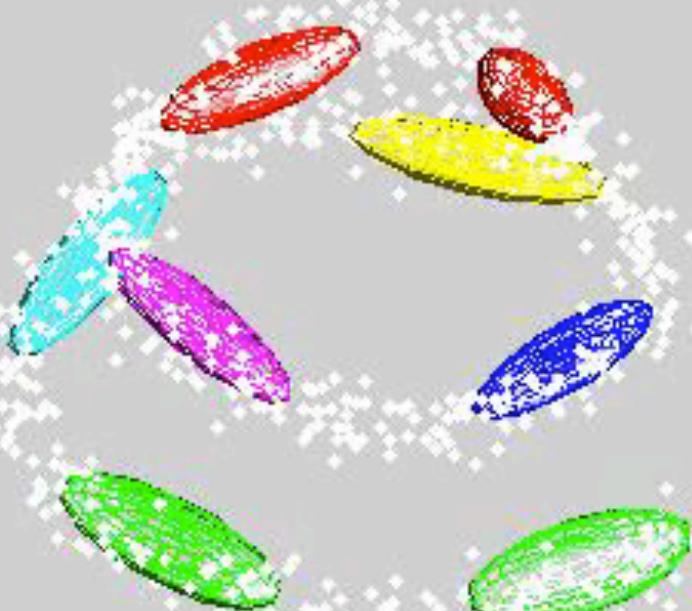
Show test data

Discretize output

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



# Clustering





# Common representation

IMAGE/  
VIDEO

TEXT/  
COMMENT

TIME  
SERIES

SYSTEM  
LOGS

NETWORK

TABULAR/  
RATING

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



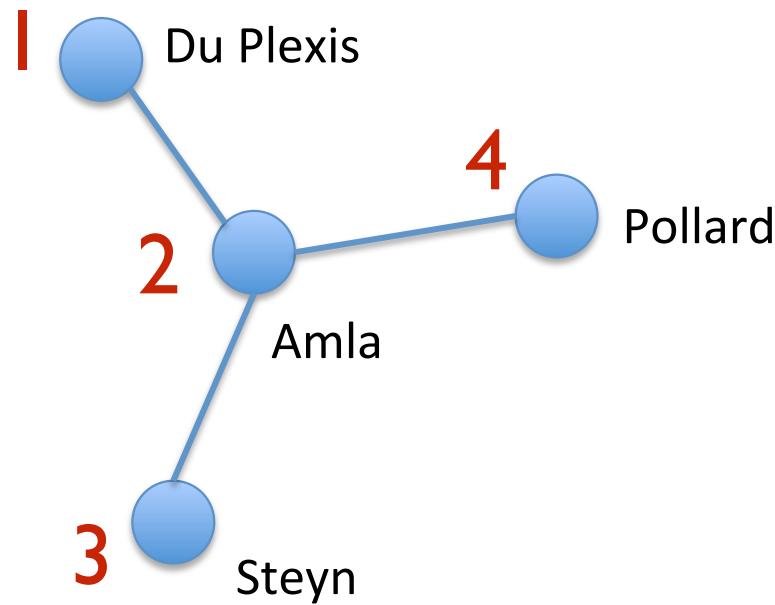
# Text to matrix

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

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



# Network data



Nodes	Nodes			
	0	1	0	0
	1	0	1	1
Nodes	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
---	----	---	---	----	----	----	----	----



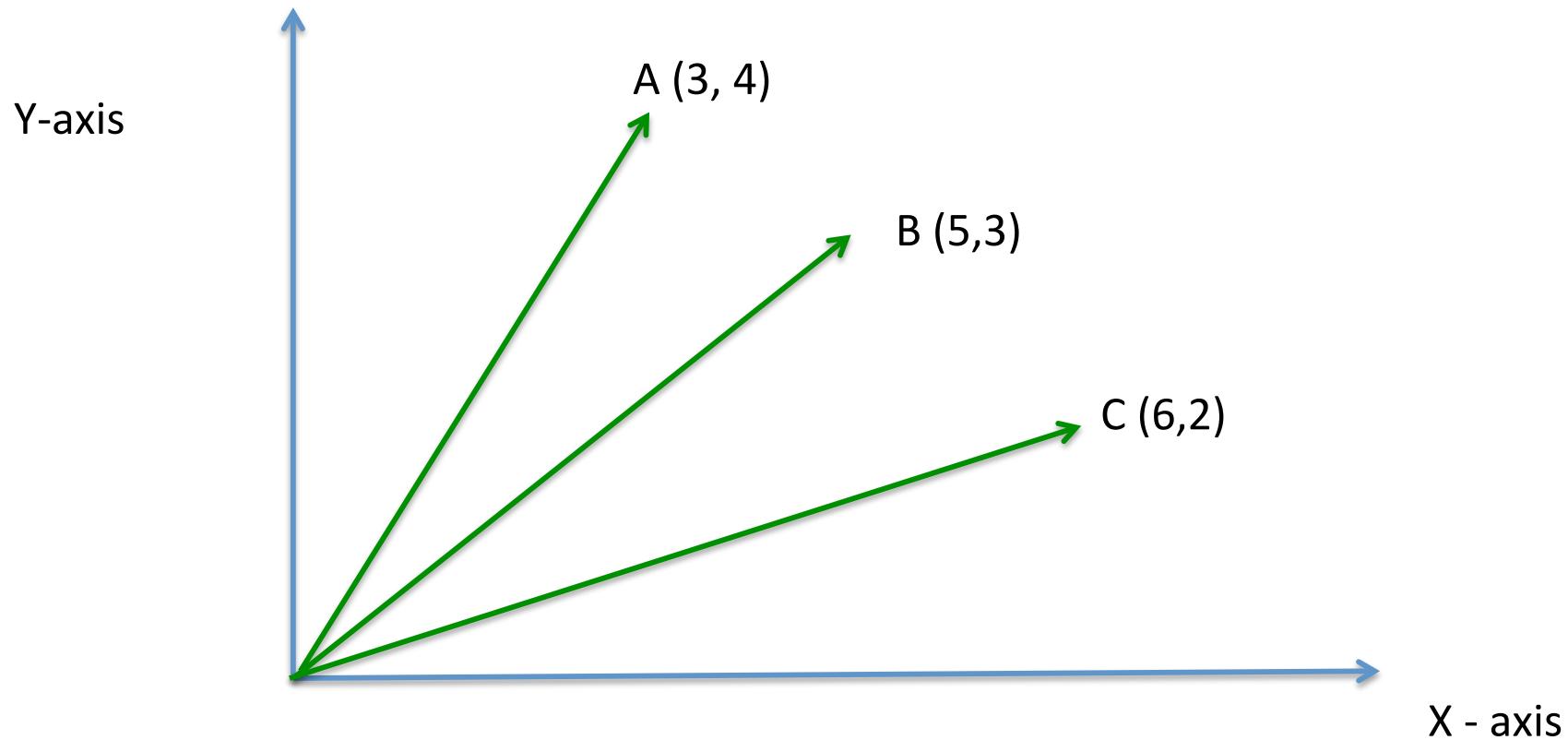
# Similarity Computation

- 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

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

$$\mathbf{x} \cdot \mathbf{y} = (x_1 y_1 + x_2 y_2 + \dots + x_n y_n)$$

- Norm (length) of a vector

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



# Similarity Computation

- The similarity between two vectors  $x$  and  $y$  is given by
  - $\text{sim}(x,y) = \mathbf{x} \cdot \mathbf{y} / (\|\mathbf{x}\| \|\mathbf{y}\|)$



# 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
  - $\|x\| = (3^2 + 1^2 + 2^2 + 4^2)^{0.5} = 5.477$
  - $\|y\| = 3.162$
- $\text{sim}(x, y) = x \cdot y / (\|x\| \|y\|) = 0.8660$

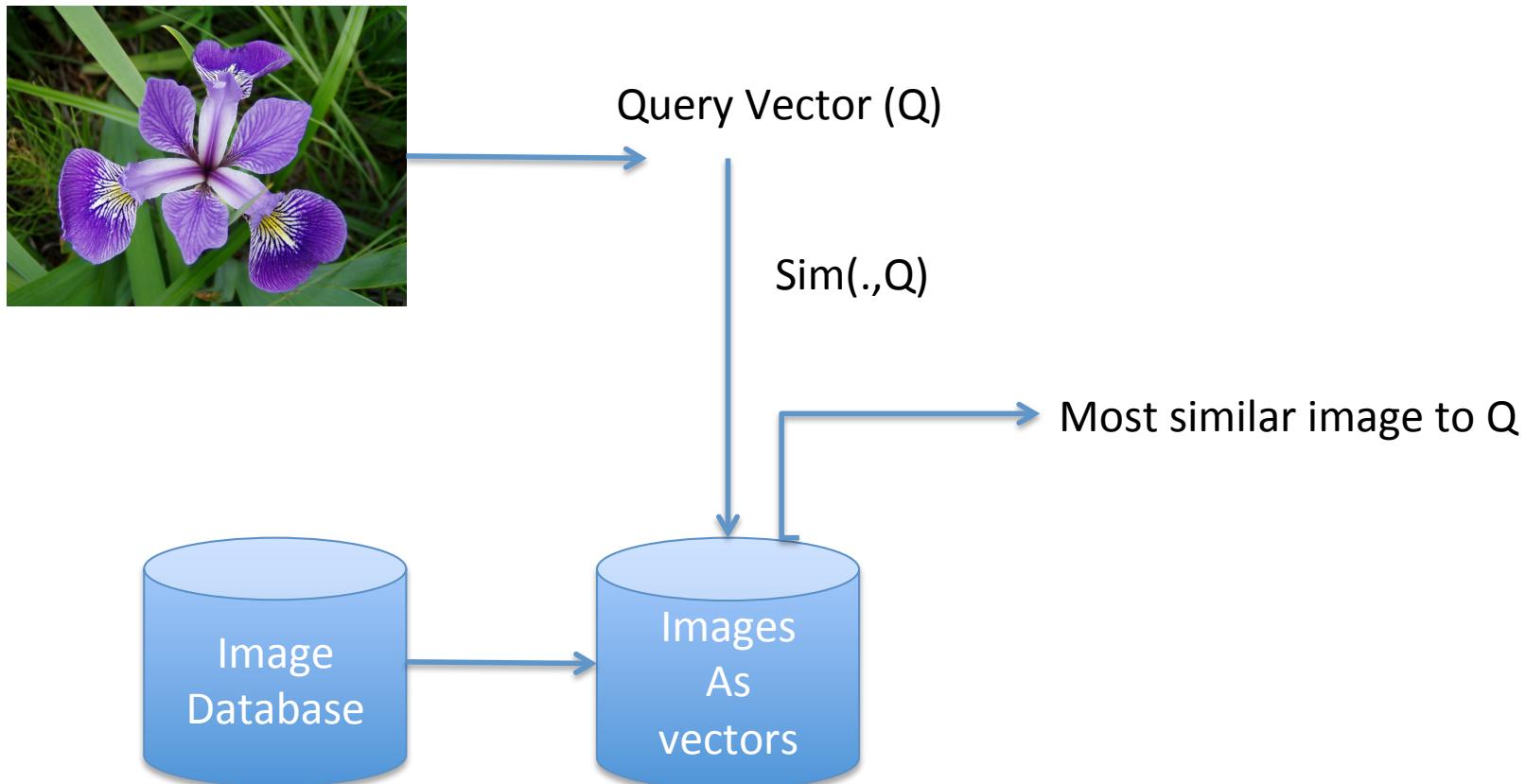


# Properties

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



# Image search engine





# Object Recognition

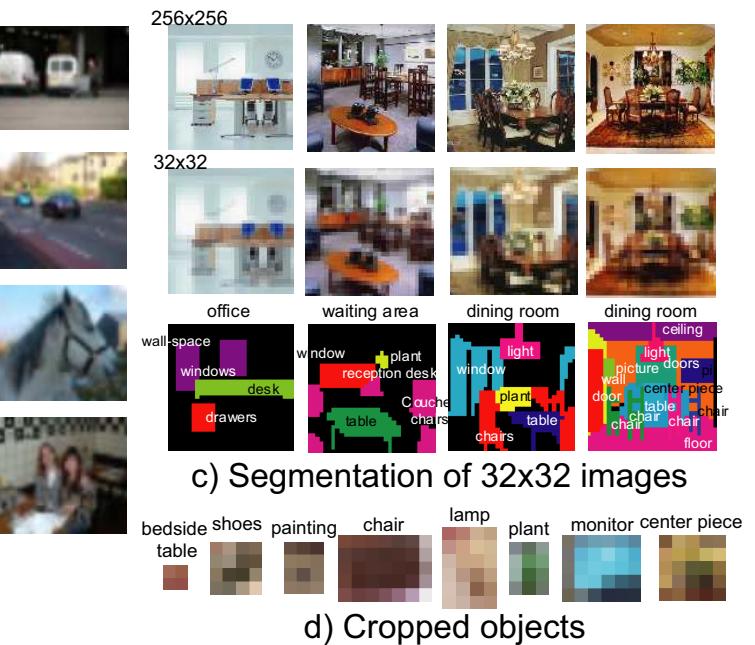
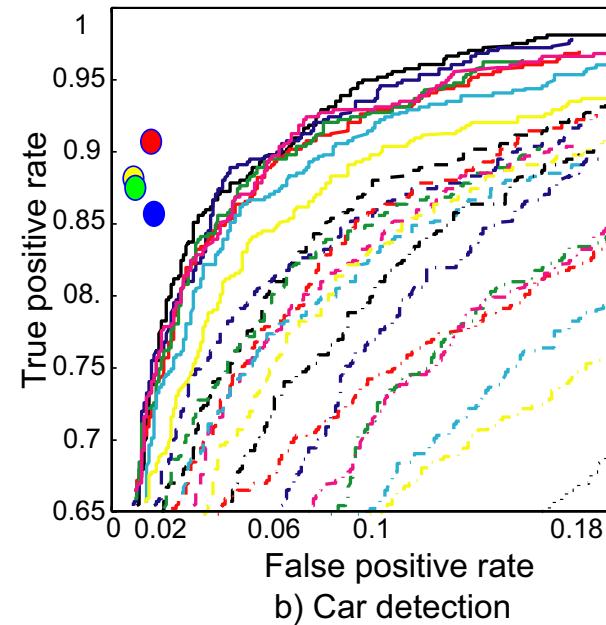
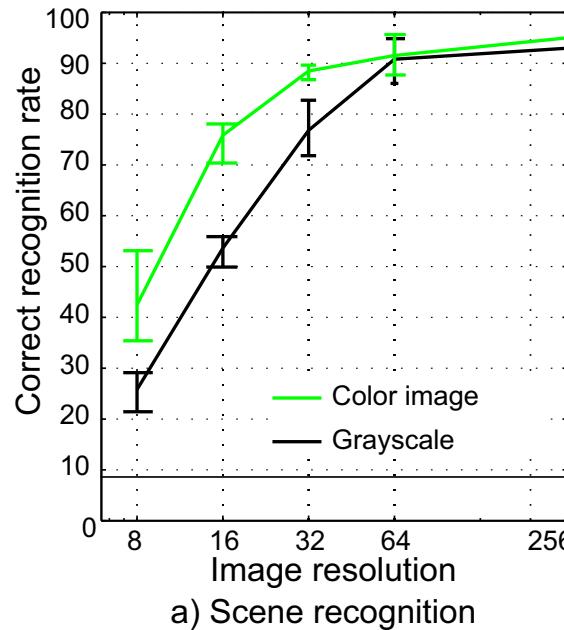


Fig. 1. a) Human performance on scene recognition as a function of resolution. The green and black curves show the performance on color and gray-scale images respectively. For color  $32 \times 32$  images the performance only drops by 7% relative to full resolution, despite having 1/64th of the pixels. 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).



THE UNIVERSITY OF  
SYDNEY

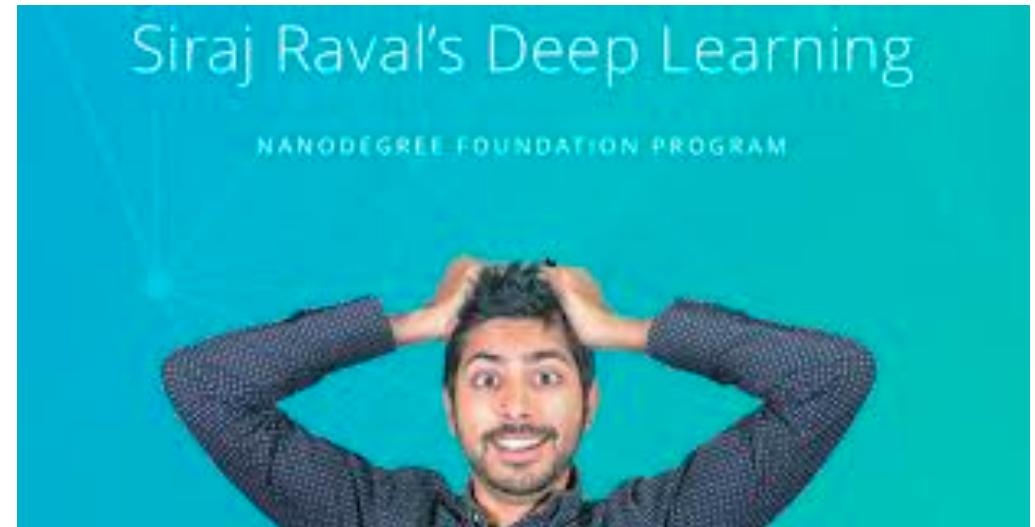
# Data Mining and Turing Test

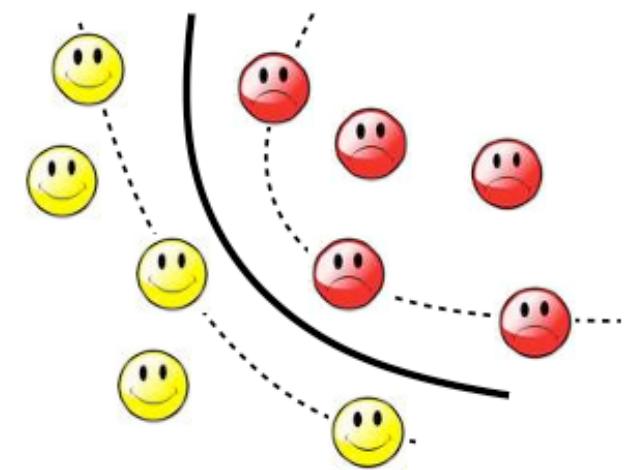
Source: BBC Radio 4 - Analysis Podcast



THE UNIVERSITY OF  
SYDNEY

# YouTube channels





# My research