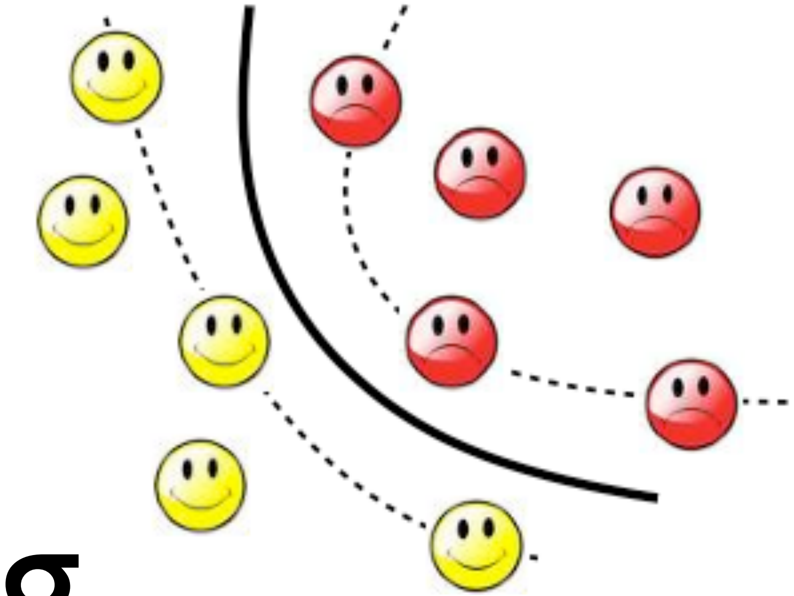
THE UNIVERSITY OF SYDNEY



Machine Learning and Data Mining

(COMP 5318)

School of Computer Science

Nguyen Hoang Tran

Team

- Unit Coordinator and Lecturer:
  - Dr Nguyen Hoang Tran, room 428, J12
- Teaching Assistants: Canh Dinh, Zhiyi Wang
- Tutors

Format of the lectures

- 10-15 min review from previous week
- 1h-1h30min of new content
- 5-10 min of examples

Assumed knowledge

- Linear algebra, calculus
- Basics of probability theory
- Programming skills (Python)

Labs: Python

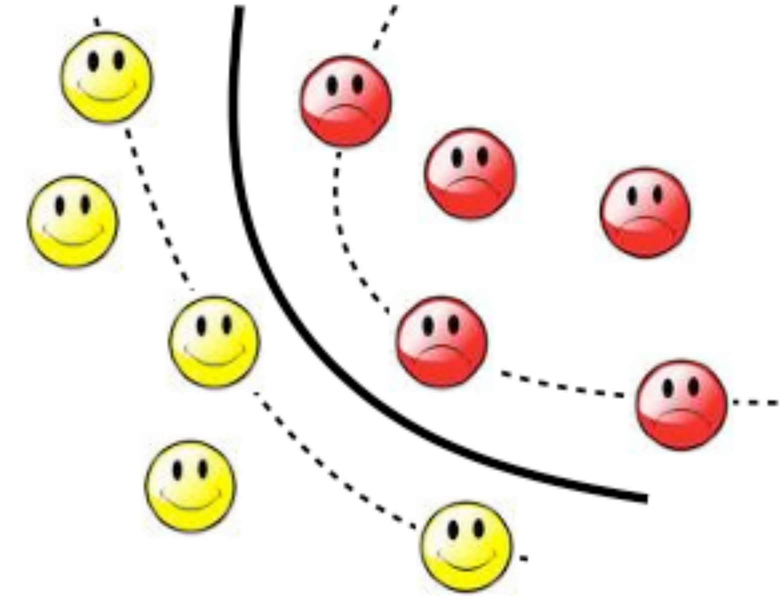
- Python is a high-level programming language designed to enforce good coding practices.
- Interactive and very natural to use.
- Extremely versatile and excellent for prototyping.
- Great libraries for machine learning eg. scikit-learn, TensorFlow, Keras, Pytorch

www.python.org

Tutorial 1

- Check on canvas
- Introduction to Python
- Bring questions to your tutor next week

THE UNIVERSITY OF SYDNEY



Let’s talk about Machine Learning

What is Machine Learning?

Informally: Making predictions from data

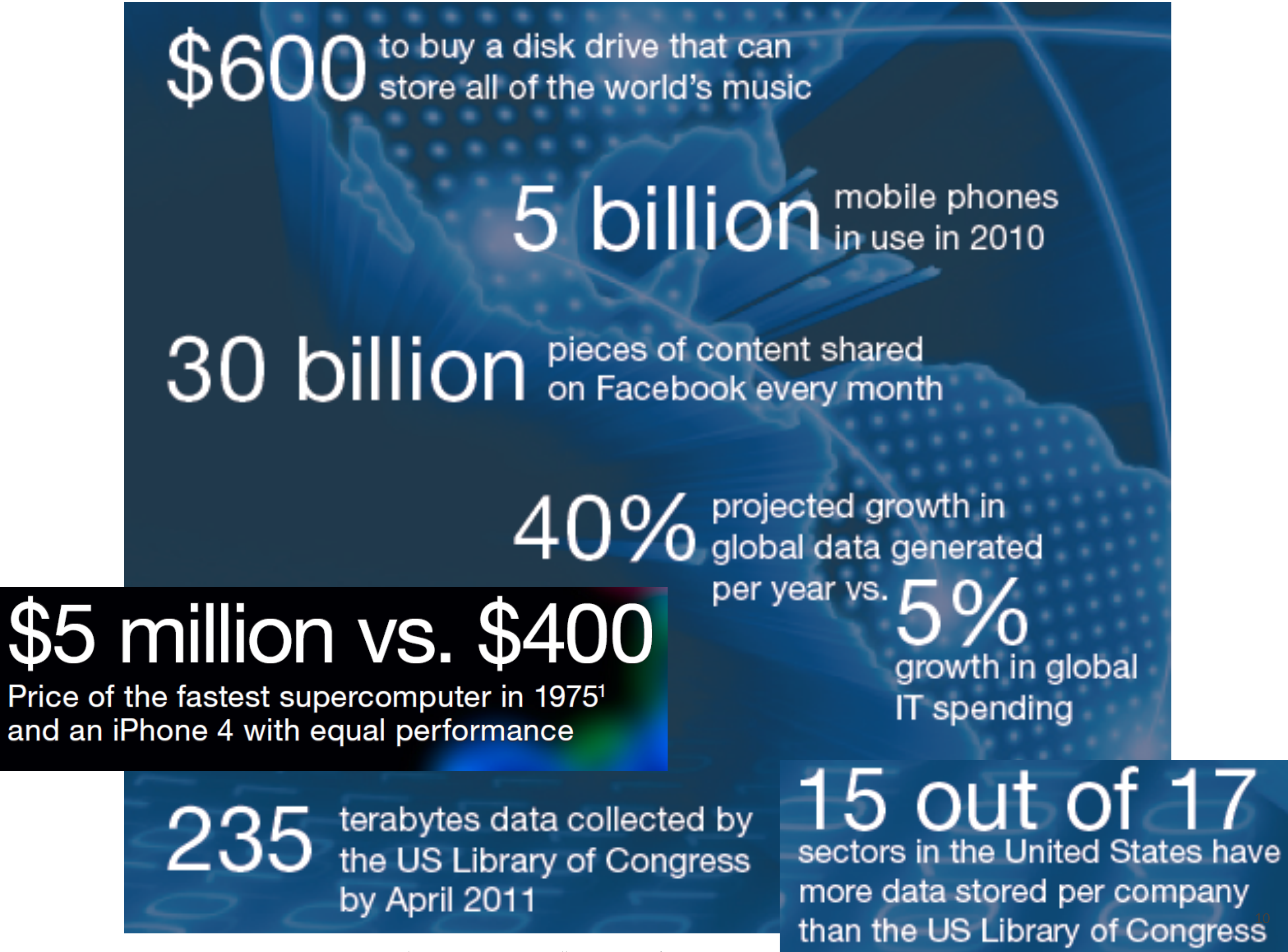
Formally: The construction of a statistical model that is an underlying distribution from which the data is drawn from, or using which we can classify the data into different categories..

ML / DM

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

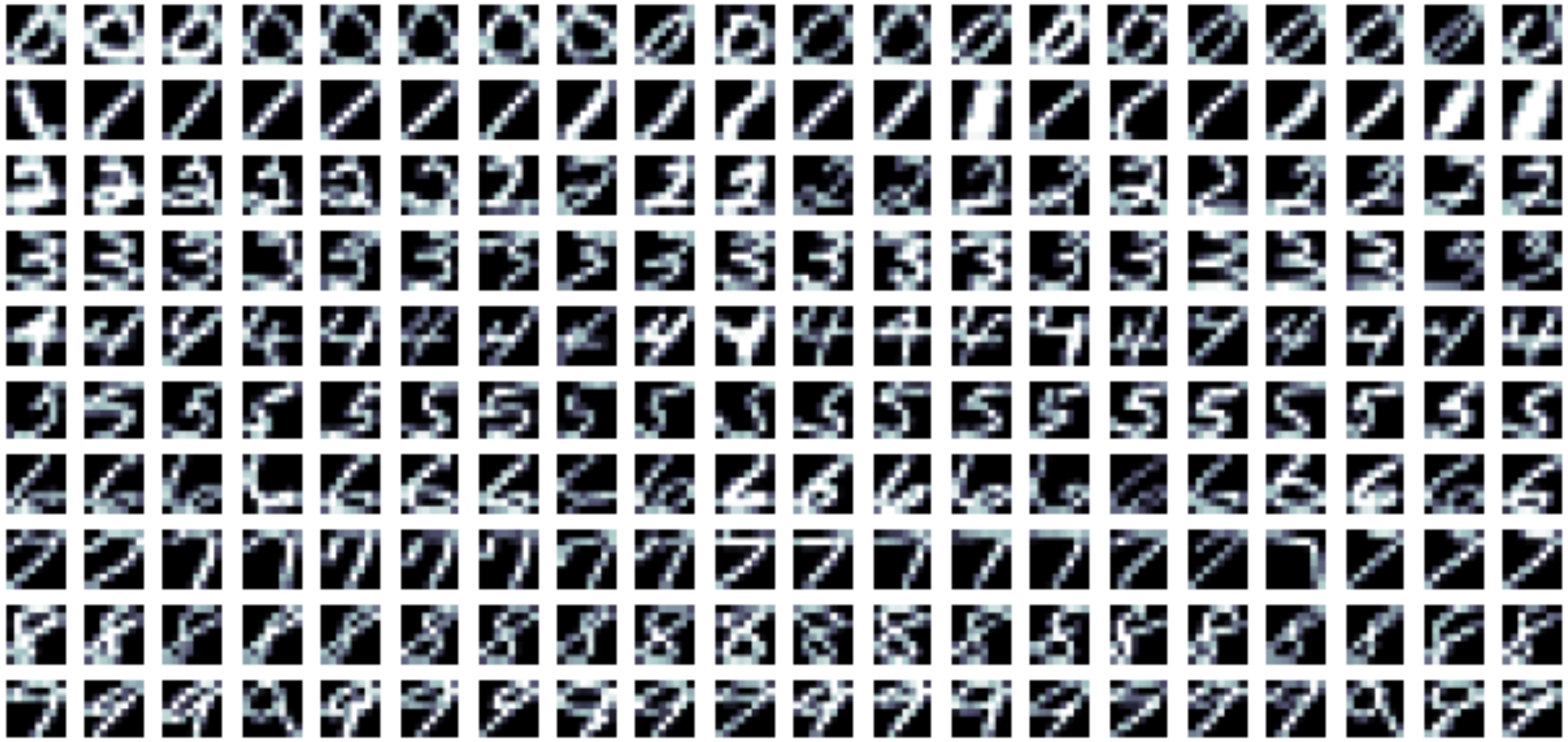
Data Mining ≈ Big Data ≈ Statistics ≈ Machine Learning ≈ Data Science

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


$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[1] 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

Speech recognition

Object and handwriting recognition



(NORB image from Yann LeCun)

Information retrieval


Google Search: Unsupervised Learning

Web Images Groups News Froogle more »

Unsupervised Learning Search Advanced Search Preferences

Web Results 1 - 10 of about 150,000 for Unsupervised Learning. (0.27 seconds)

Mixture modelling, Clustering, Intrinsic classification ...
Mixture Modelling page. Welcome to David Dowe's clustering, mixture modelling and unsupervised learning page. Mixture modelling (or ...
www.csse.monash.edu.au/~dld/mixture.modelling.page.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. Endorsed by SIGNLL ...
www.ai.sri.com/~kehler/unsup-acl-99.html - 5k - Cached - Similar pages

Unsupervised learning and Clustering
cgm.cs.mcgill.ca/~soss/cs644/projects/wijhe/ - 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/~mccallum/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 - 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/ - 15k - Cached - Similar pages
[More results from 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 ...
acl.ldc.upenn.edu/J/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 - 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/acl-wkshp.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.lans.ece.utexas.edu/ulg/ - 14k - Cached - Similar pages

Result Page: 1 2 3 4 5 6 7 8 9 10 Next

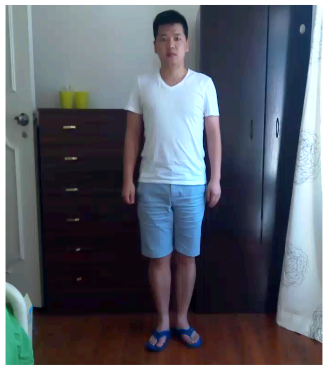
1 of 2 06/10/04 15:44

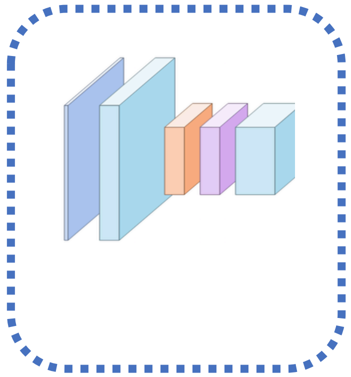
Web Pages

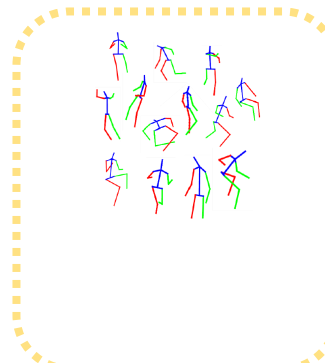
Retrieval
Categorisation
Clustering
Relations between pages

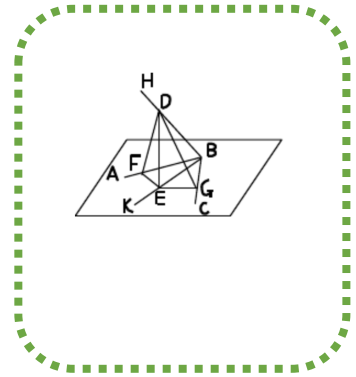
Financial prediction

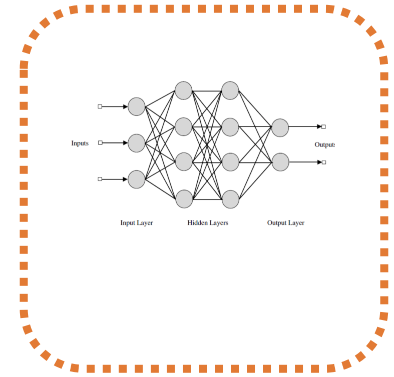
Robotics: pose estimation



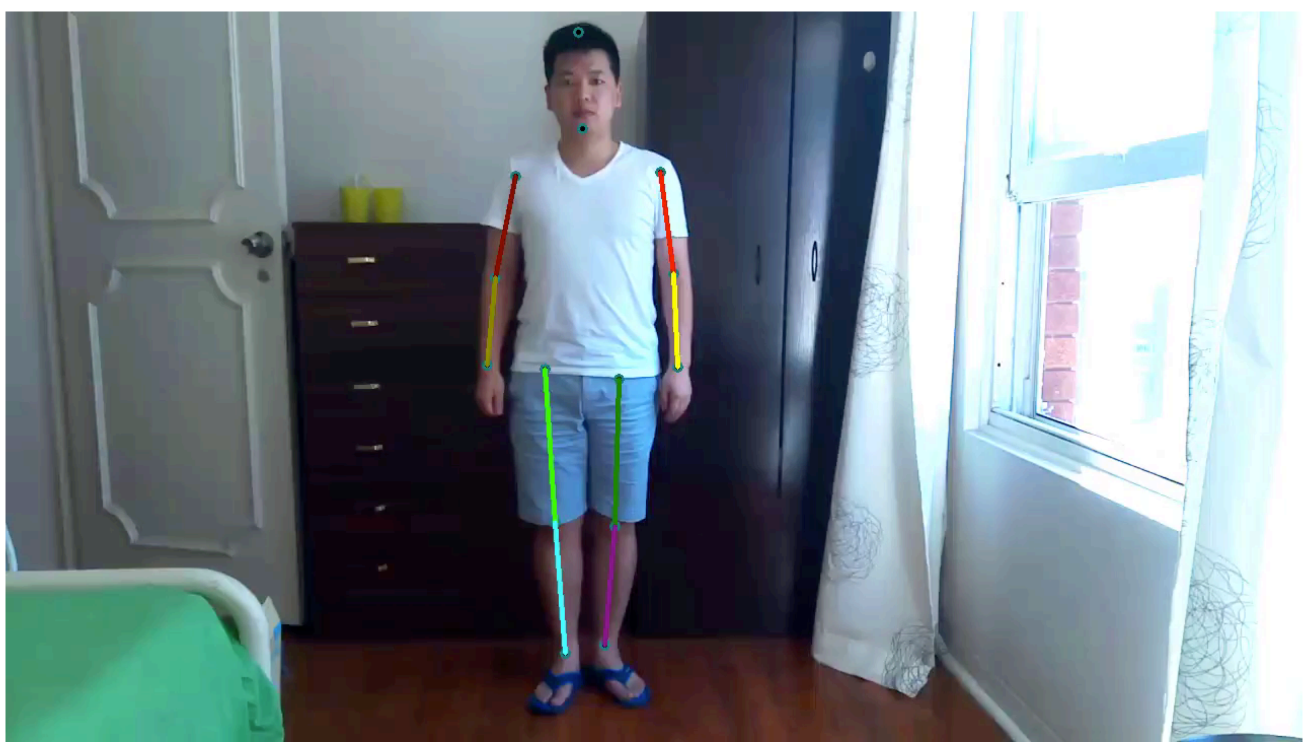


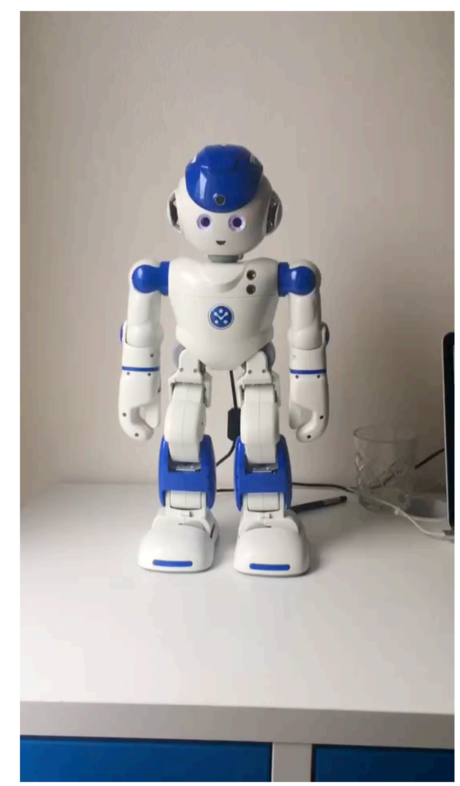






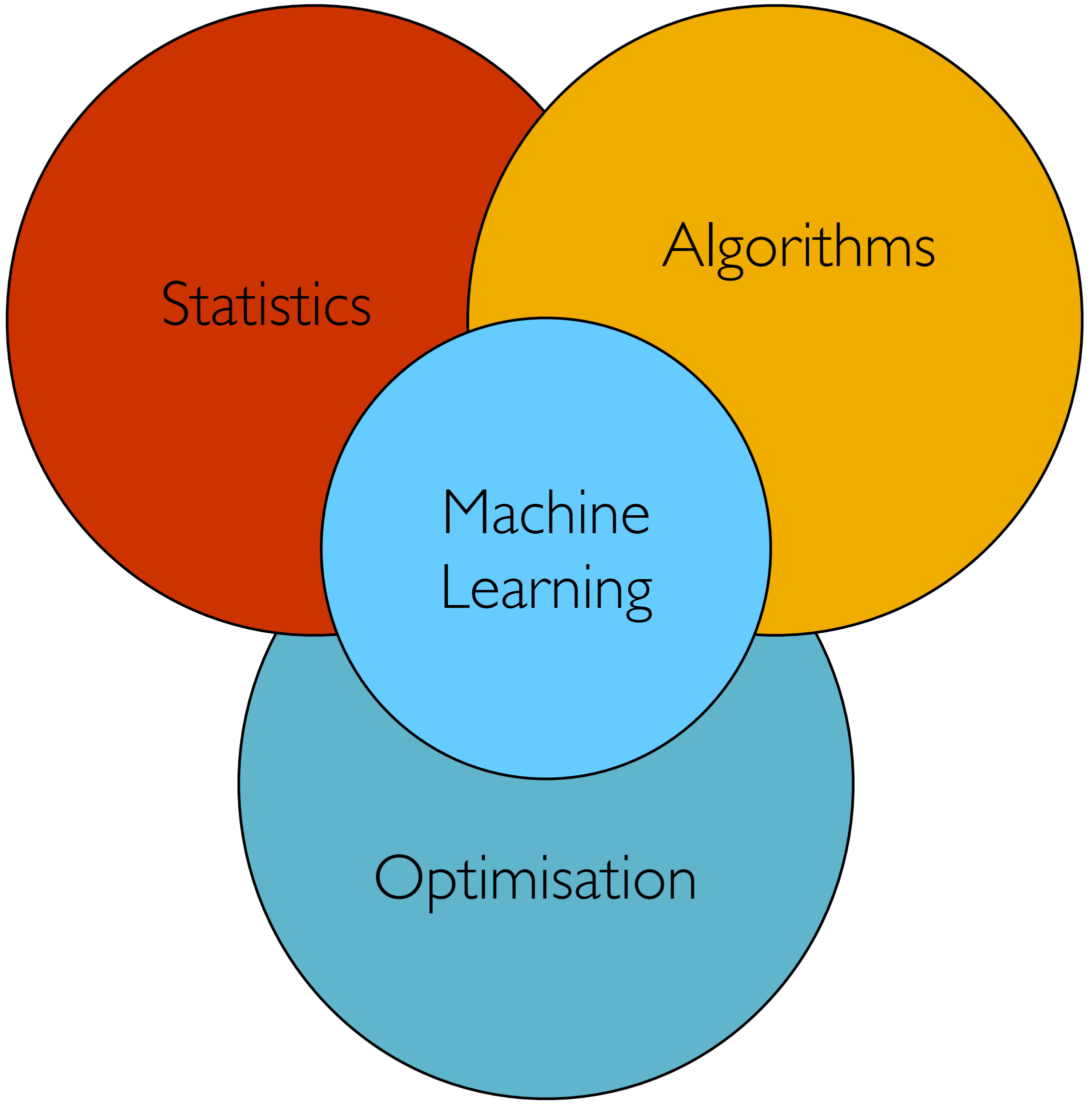
2D Pose Estimation → 3D Pose Recovery → Pose Transformation → Pose Stabilizer





This Course: COMP 5318

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



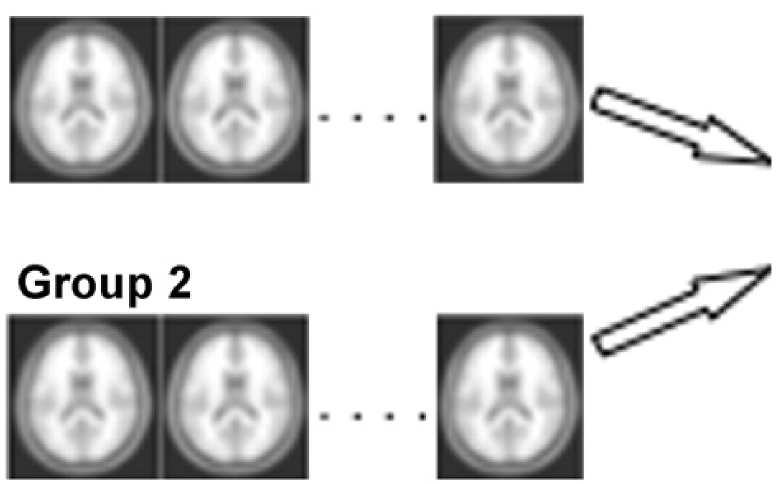
Machine Learning Problems

THE UNIVERSITY OF SYDNEY

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

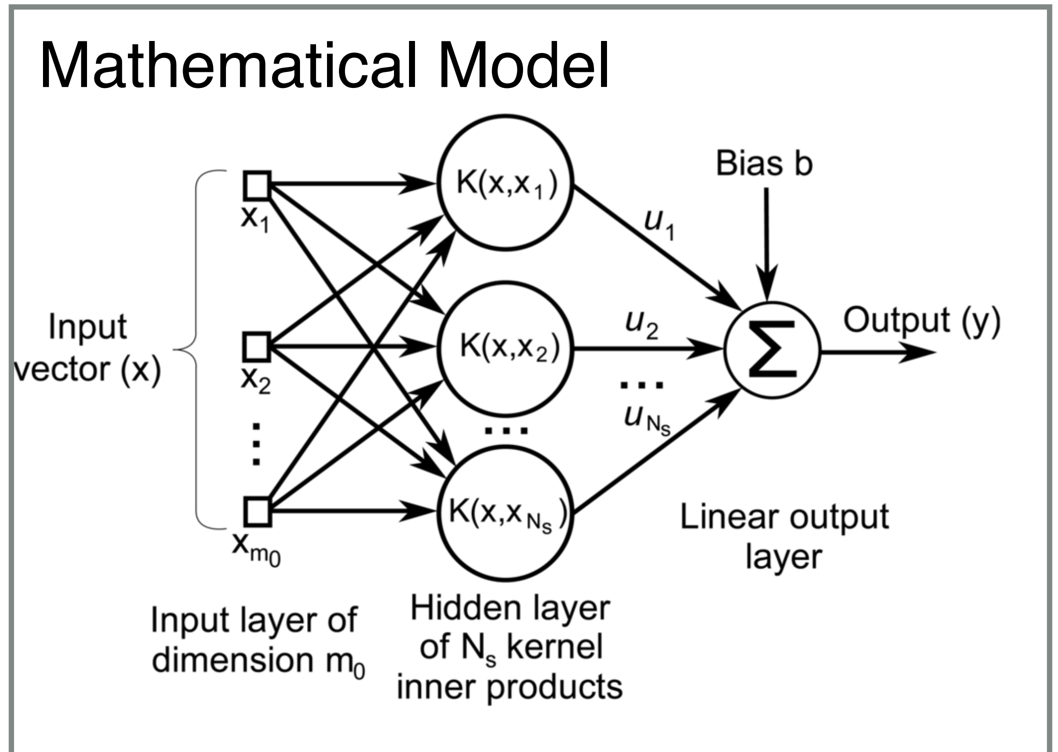
Elements of Machine Learning

THE UNIVERSITY OF SYDNEY

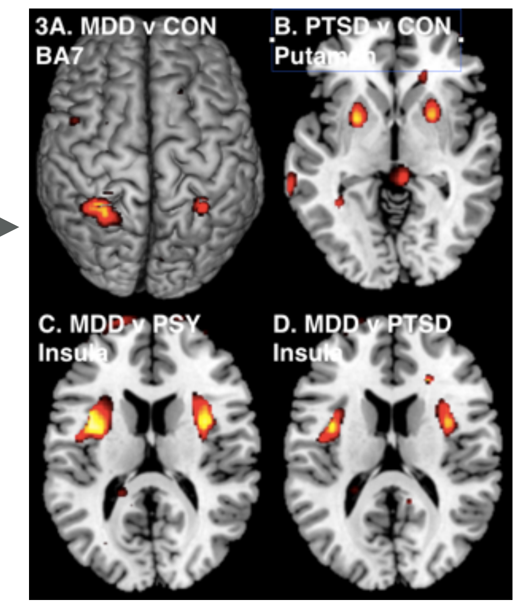


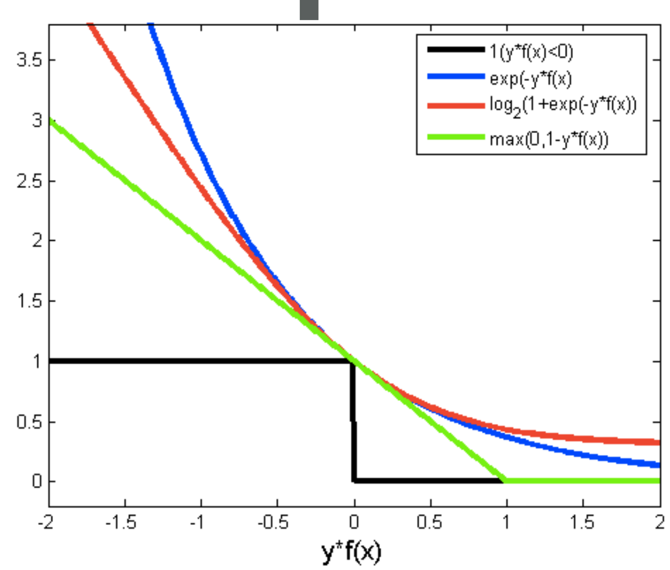
Data

Mathematical Model

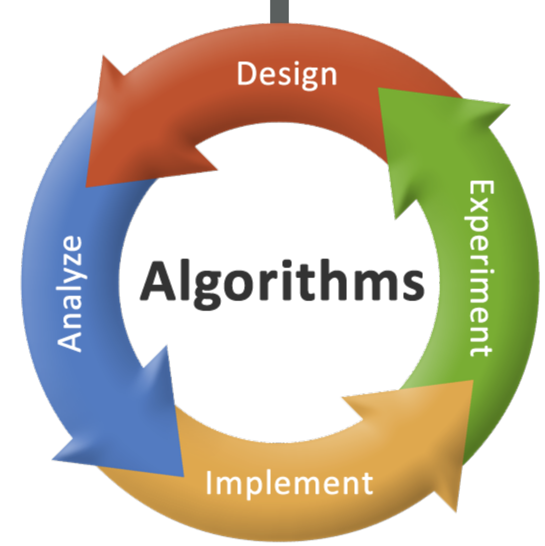


Predictions/Patterns





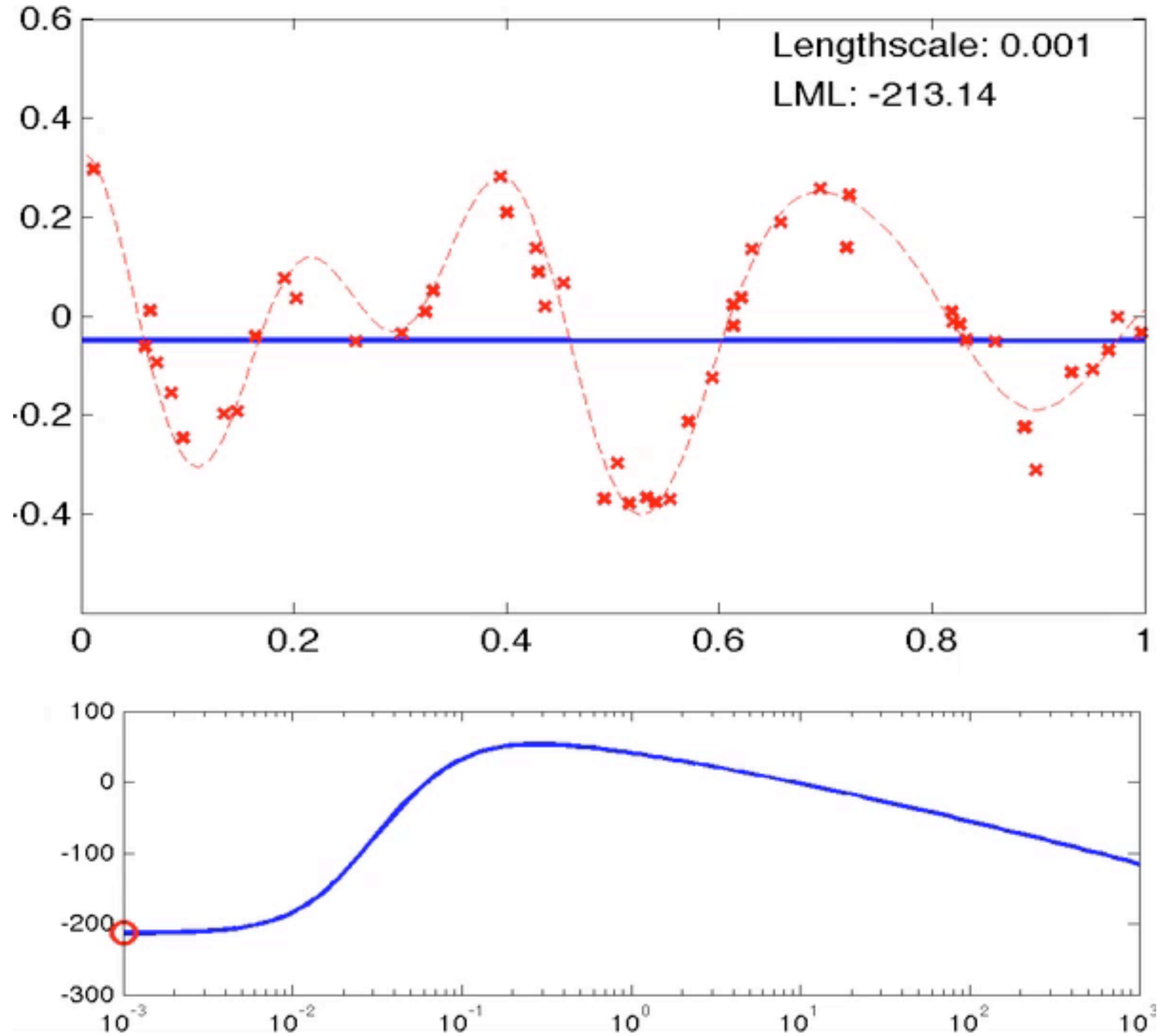
Objective function



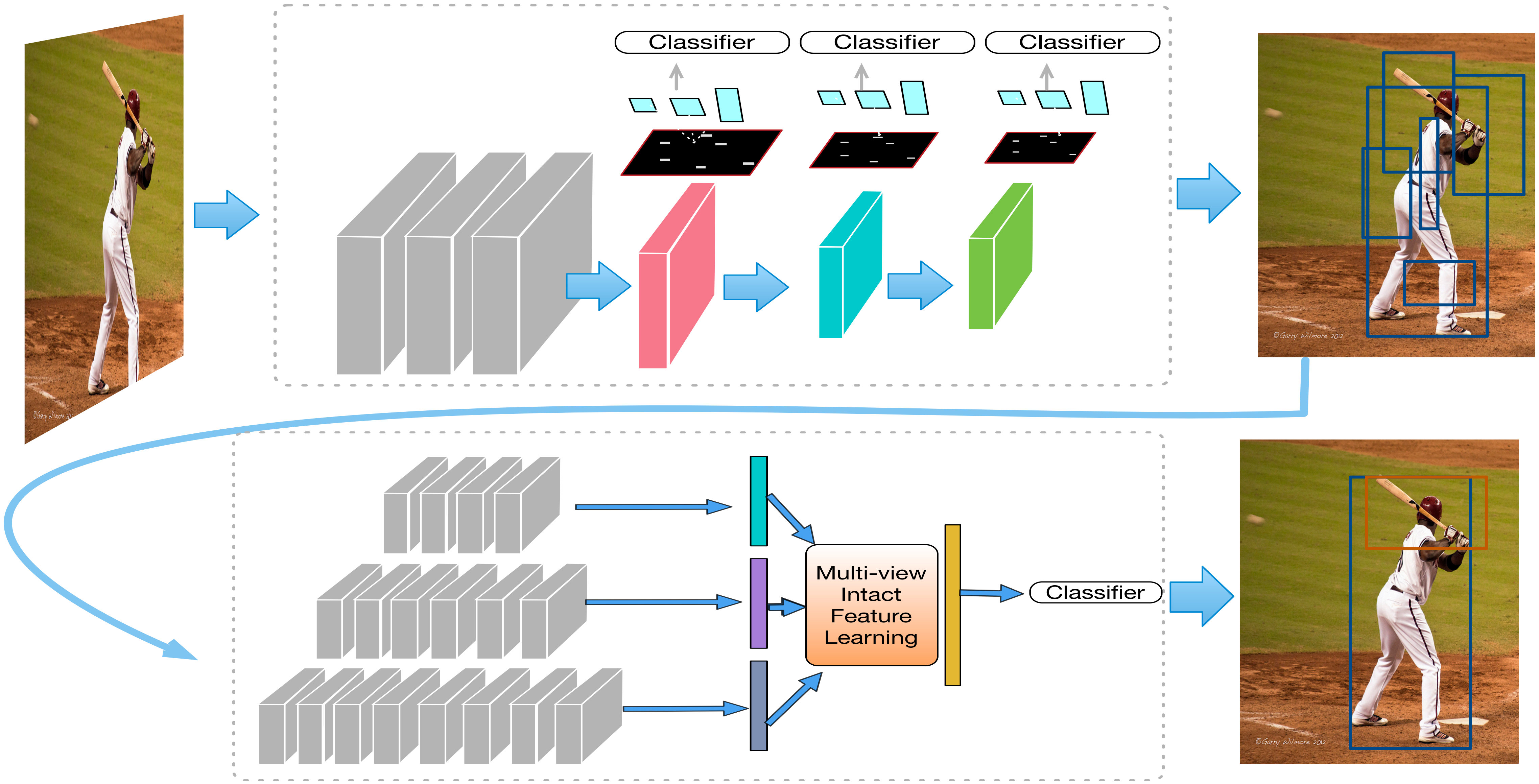
Regression

Problem

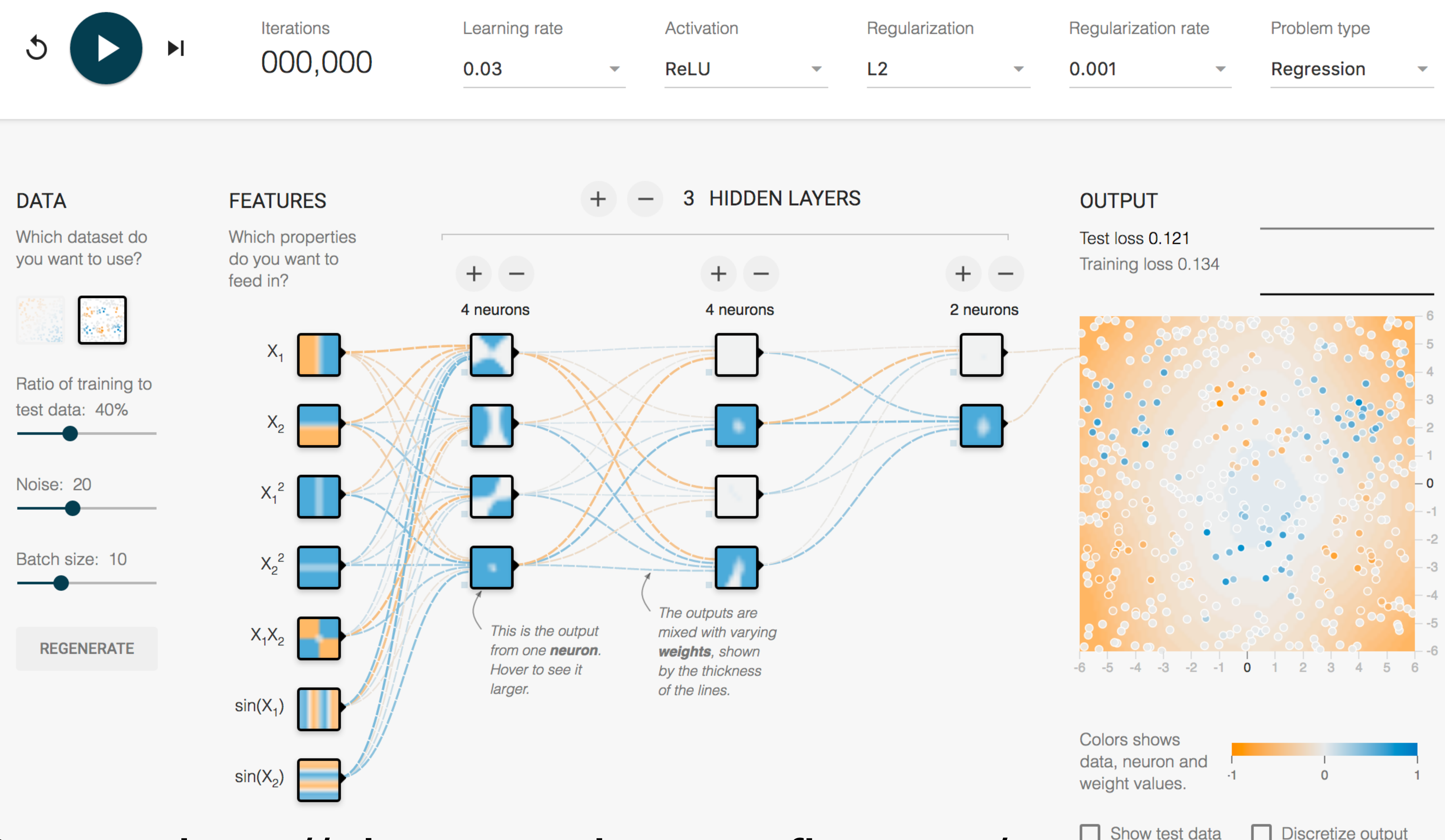
Objective



Classification for object detection

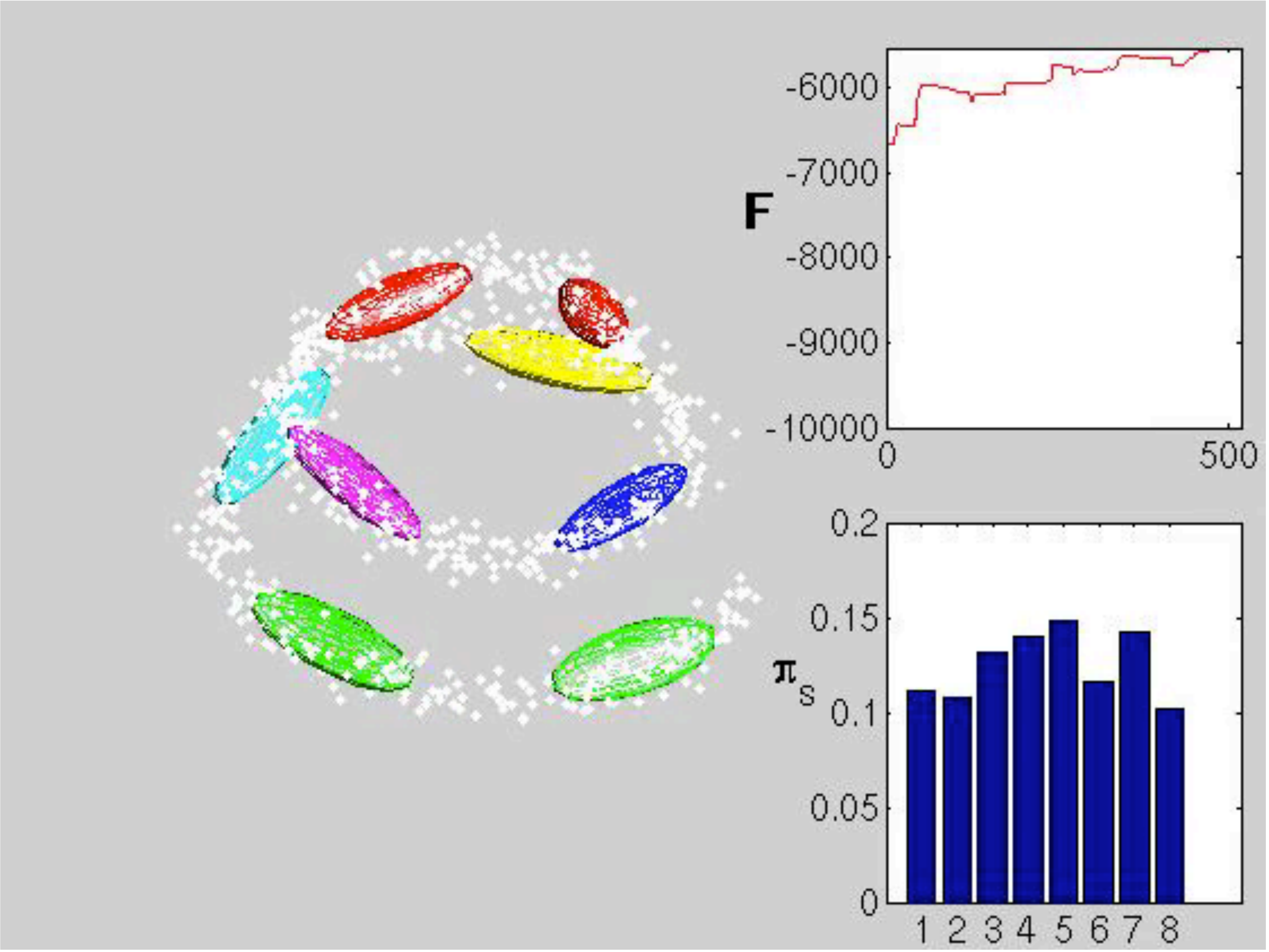


Neural Networks



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

Clustering



Common representation

IMAGE/ VIDEO

TEXT/ COMMENT

TIME SERIES

SYSTEM LOGS

NETWORK

TABULAR/ RATING

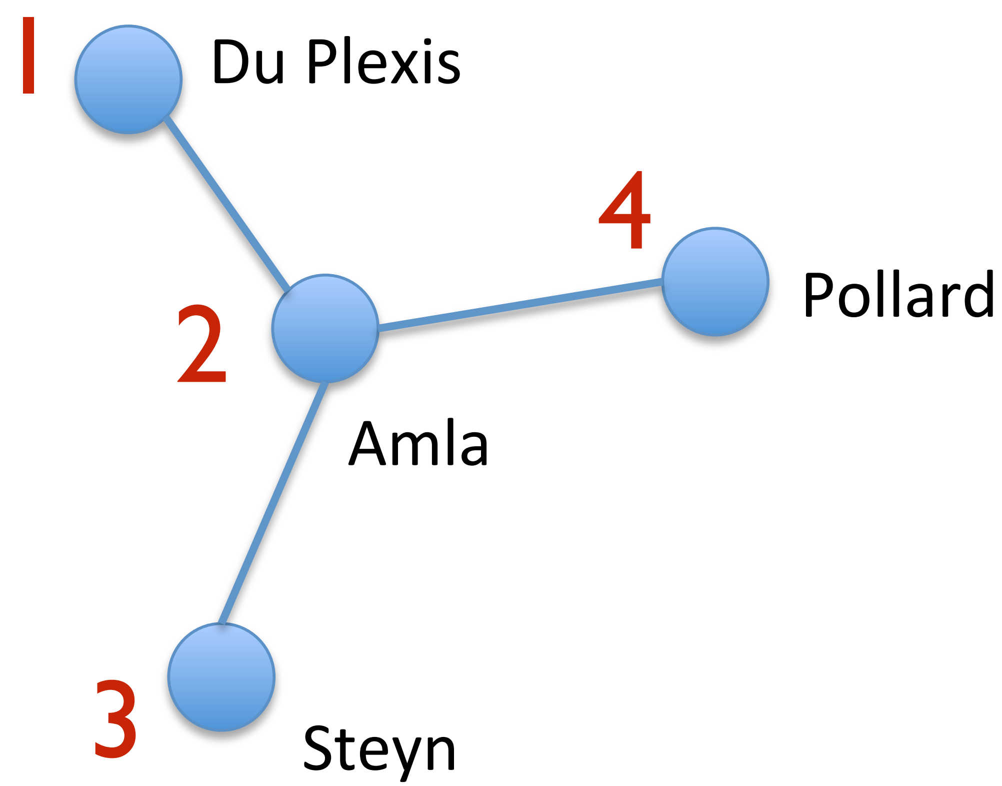
Is there a common way to represent data of different modalities ?

Text to matrix

- Document- Word Matrix
- Document 1: "AACCBBAAA"
- Document 2: "CCAABBDD"

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

Network data



$$
\begin{array}{cc}
 & \text{Nodes} \\
\text{Nodes} & \begin{bmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}
\end{array}
$$

Image data

www.sydney.visitorsbureau.com.au

700 x 500

4	45	6
6	12	33
22	17	44

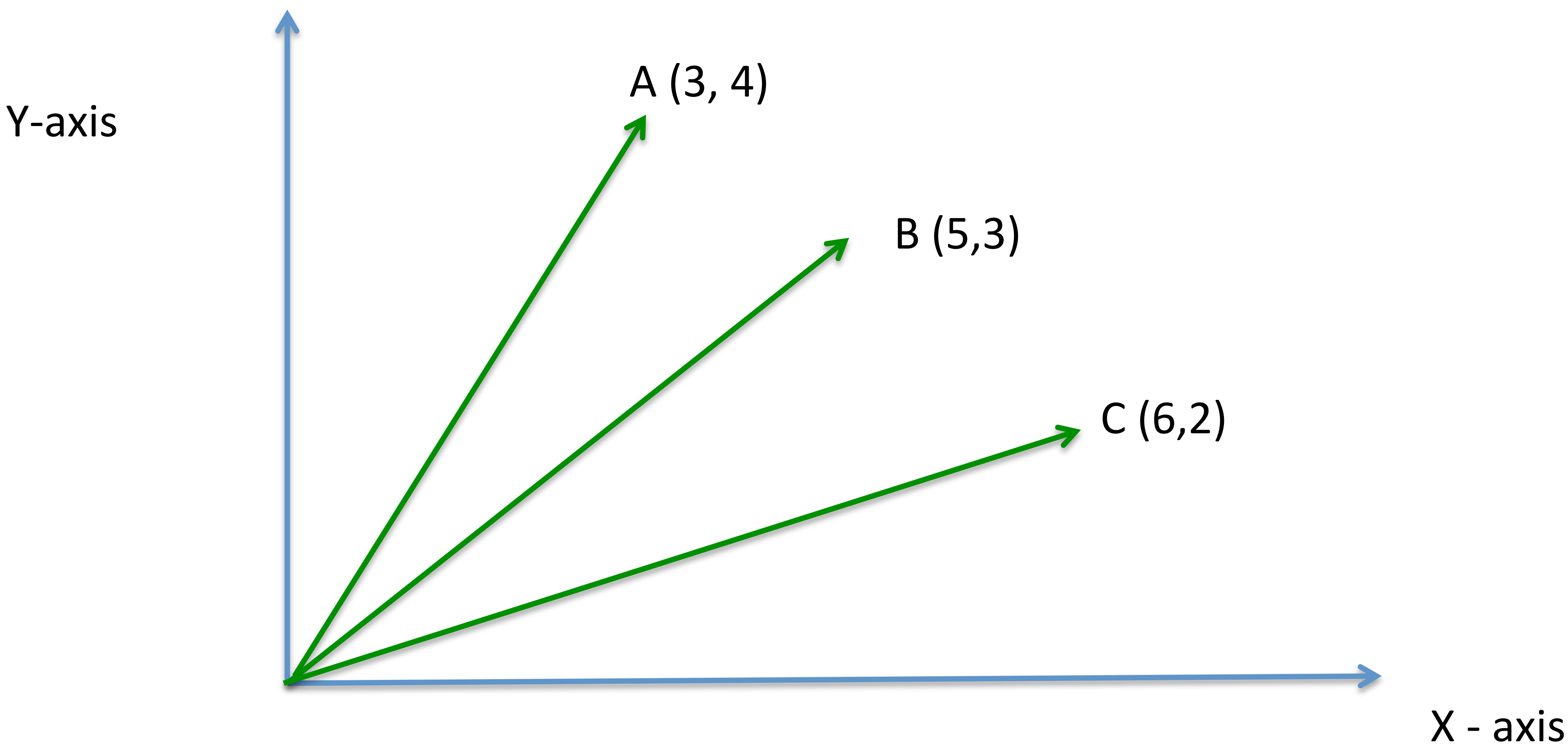
4	45	6	6	12	33	22	17	44

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

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

$$x.y = (x_1 y_1 + x_2 y_2 + ... + x_n y_n);$$

- Norm (length) of a vector

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

Similarity Computation

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

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

Example

- Let $x =< 3, 1, 2, 4 >$, $y =< 1, 2, 1, 2 >$
- Step 1: Compute the dot-product

$$x.y = 3.1 + 1.2 + 2.1 + 4.2 = 15$$

- Step 2: Compute length of x vector

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

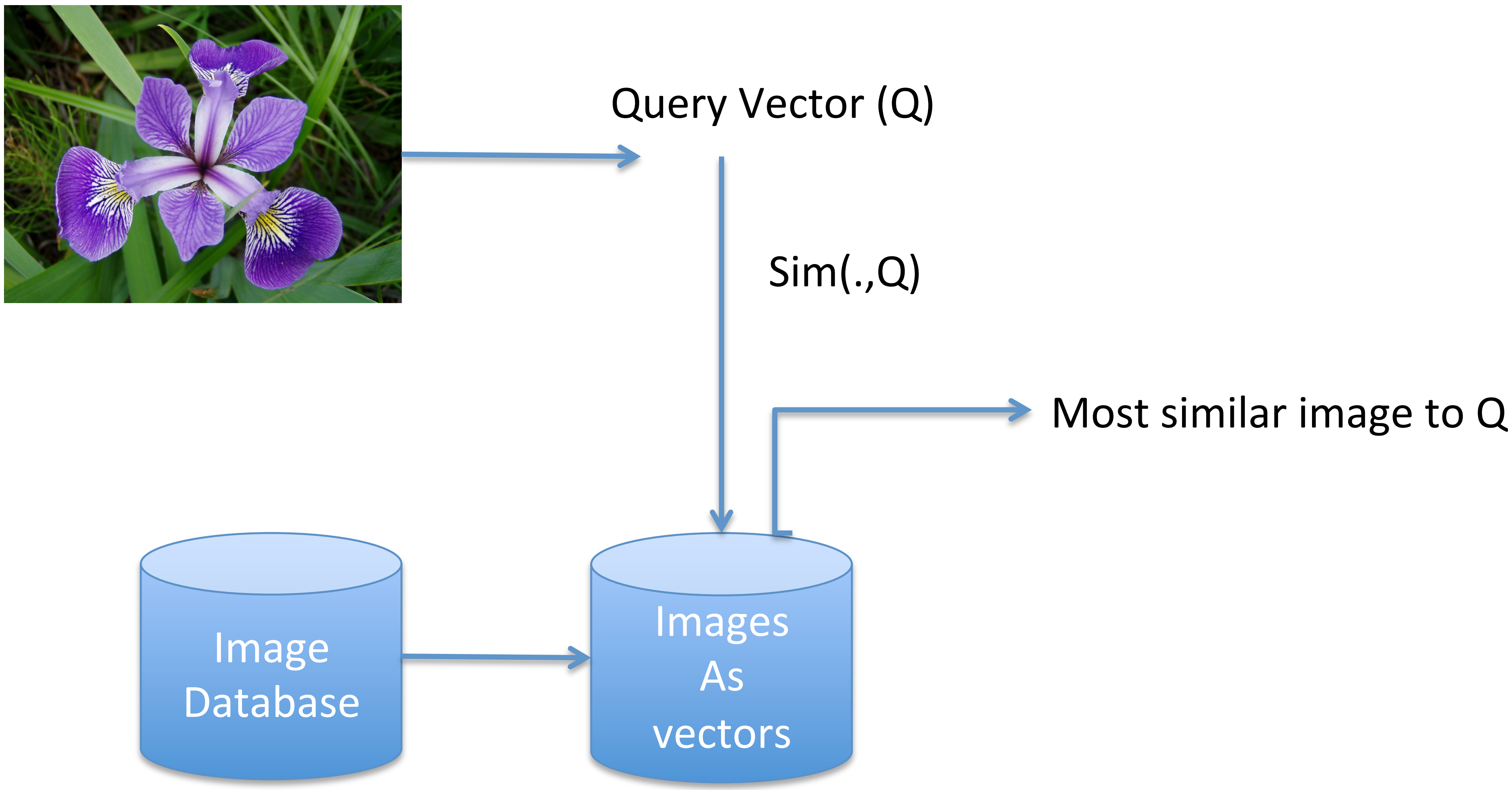
$$\|y\| = 3.162$$

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

Properties

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

Image search engine



Object recognition

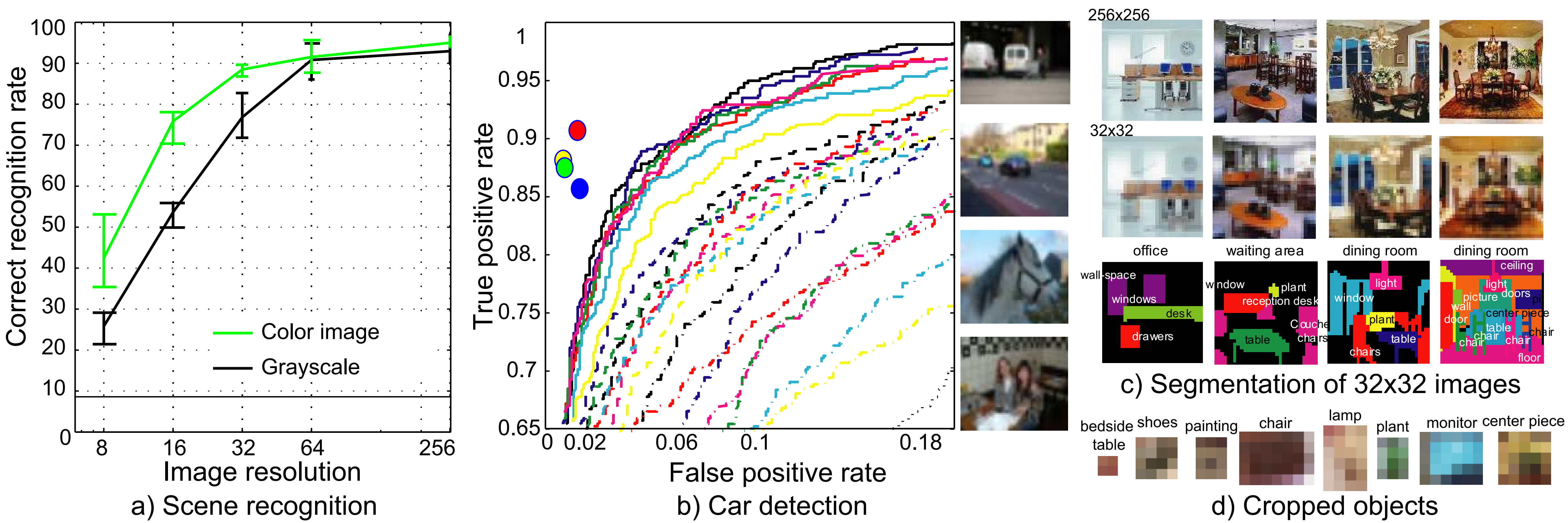
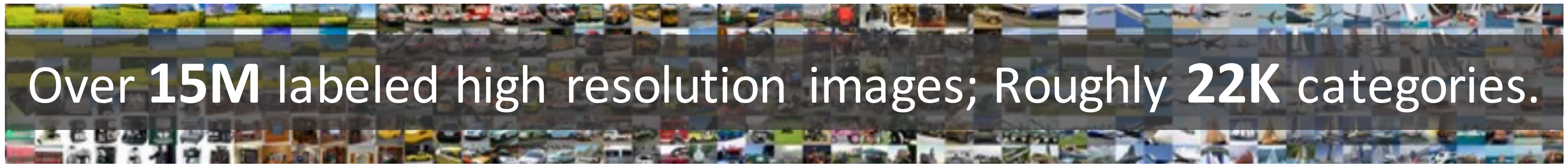
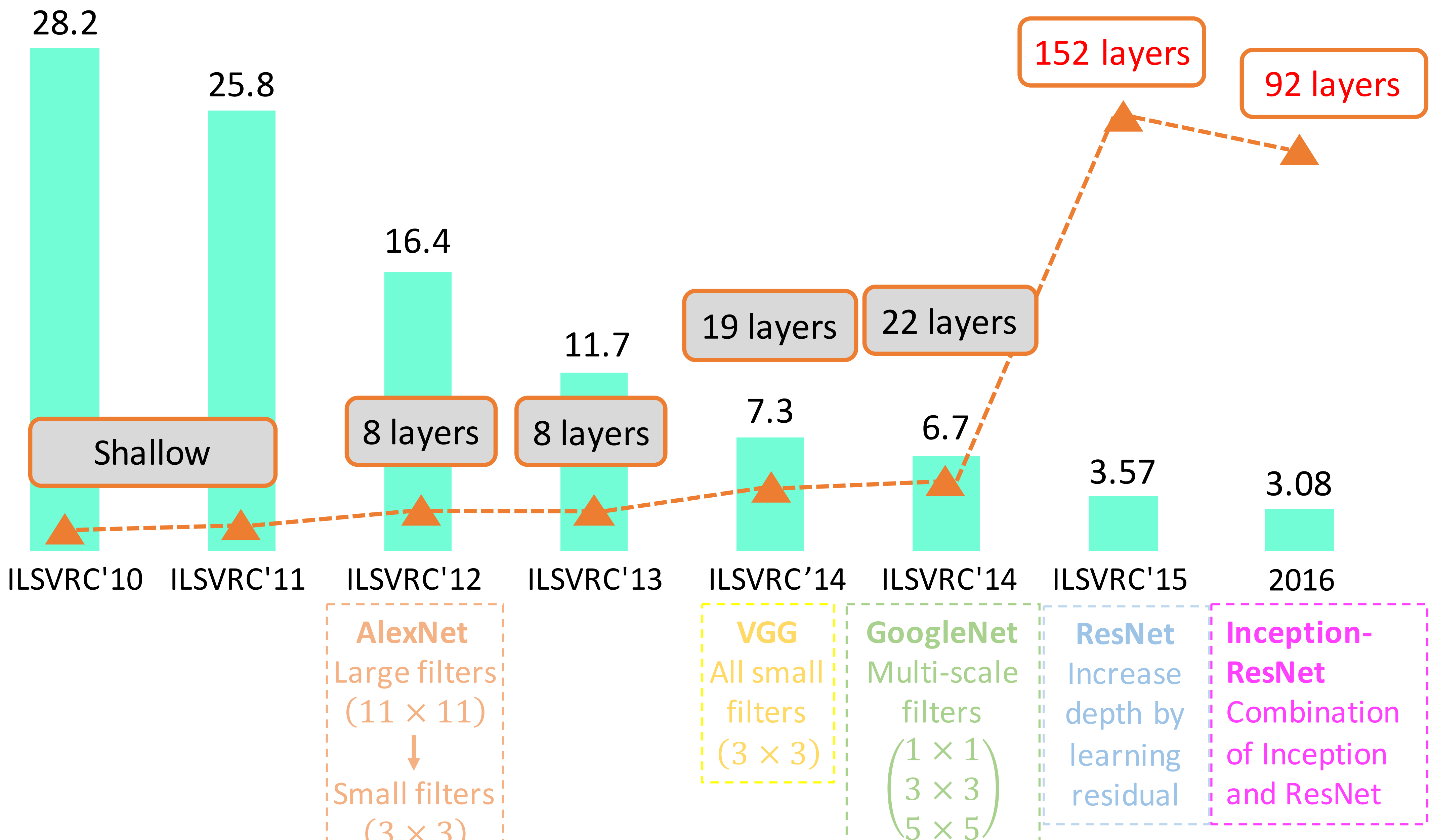


Fig. 1. a) Human performance on scene recognition as a function of resolution. The green and black curves show the performance on color and gray-scale images respectively. For color 32 × 32 images the performance only drops by 7% relative to full resolution, despite having 1/64th of the pixels. b) Car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

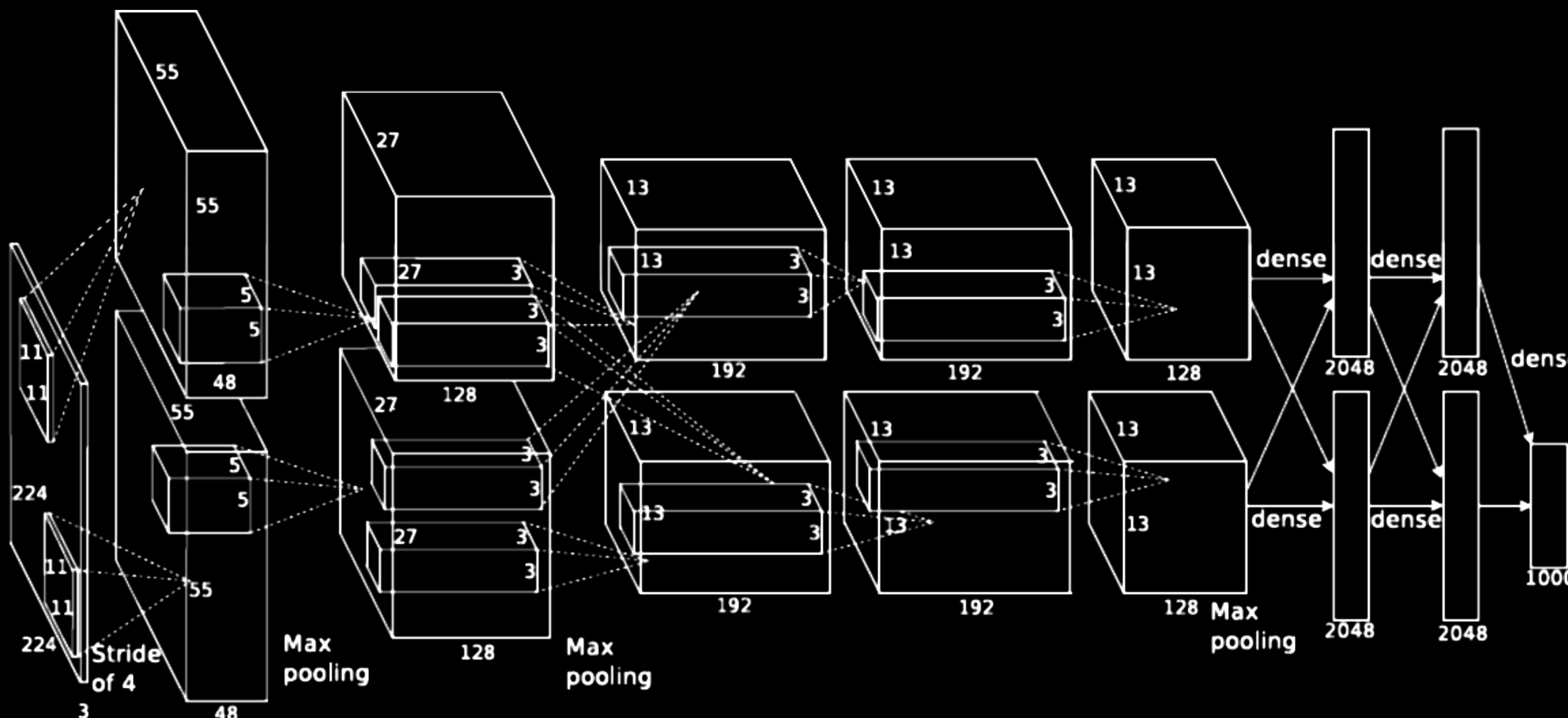
Torralba et al. 80 million tiny images: a large dataset for non-parametric object and scene recognition, PAMI 2008

ImageNet:


Over **15M** labeled high resolution images; Roughly **22K** categories.

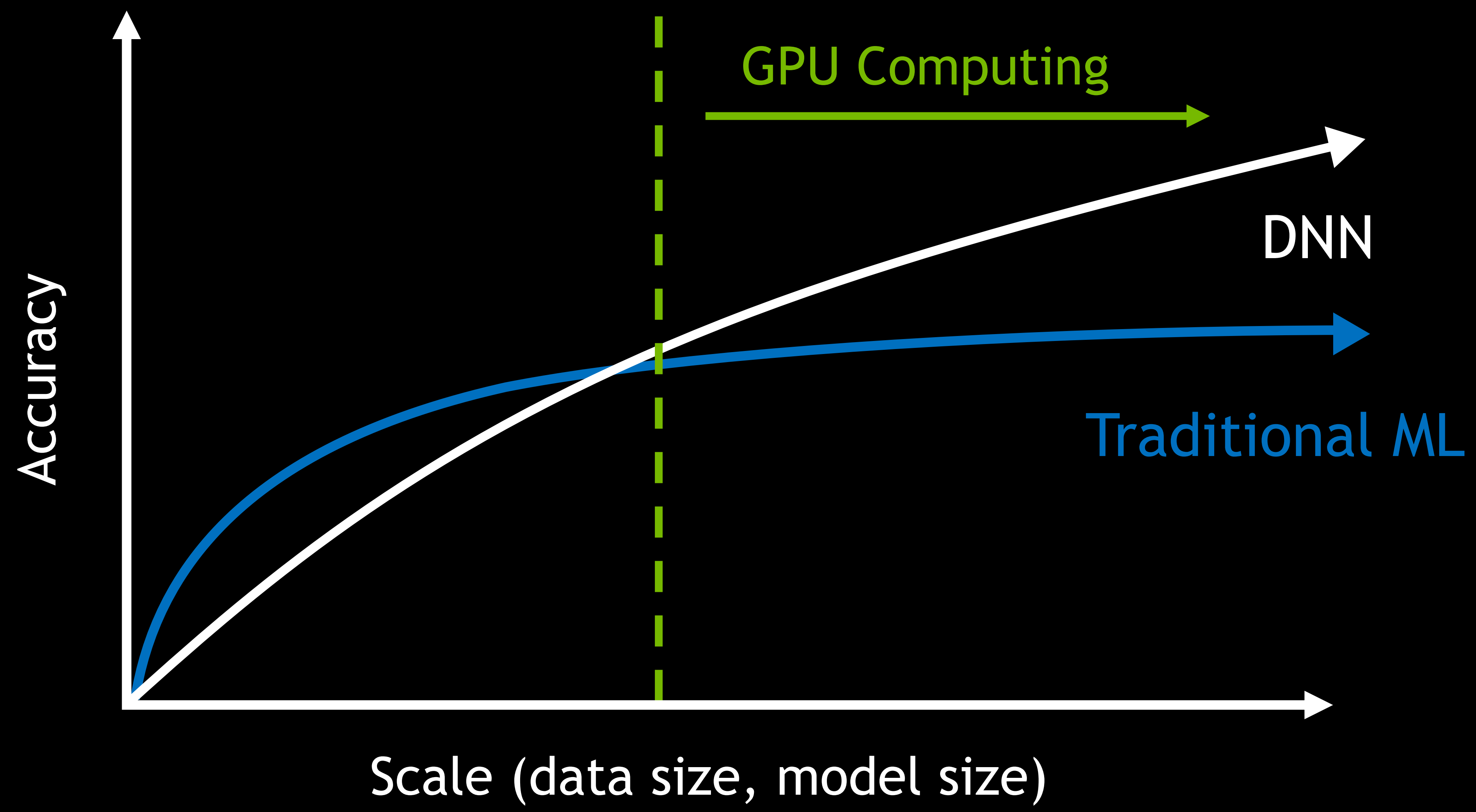

28.2
25.8
16.4
11.7
7.3
6.7
3.57
3.08
Shallow
8 layers
8 layers
19 layers
22 layers
152 layers
92 layers
ILSVRC'10
ILSVRC'11
ILSVRC'12
ILSVRC'13
ILSVRC'14
ILSVRC'14
ILSVRC'15
2016
AlexNet
Large filters
(11 × 11)
Small filters
(3 × 3)
VGG
All small filters
(3 × 3)
GoogleNet
Multi-scale filters
1 × 1
3 × 3
5 × 5
ResNet
Increase depth by learning residual
Inception-ResNet
Combination of Inception and ResNet

Deep Convolutional Neural Networks

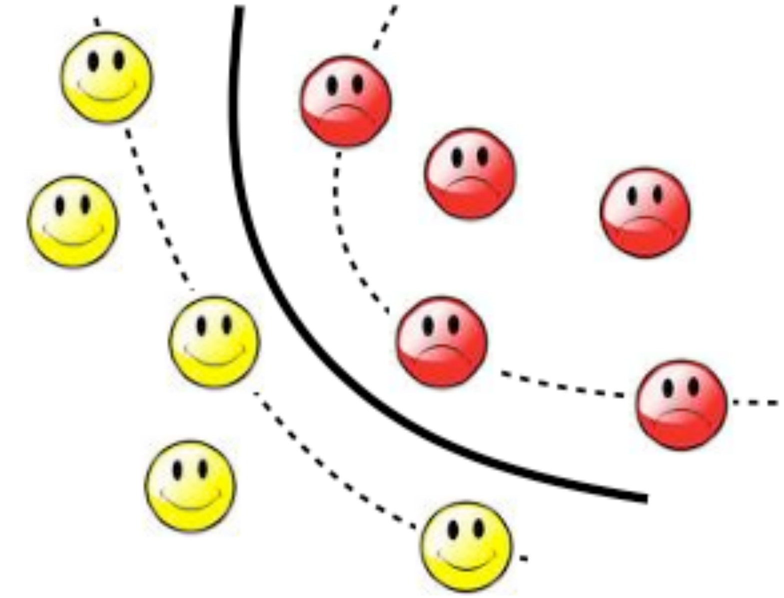


AlexNet 2012

The rise of Deep Learning



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



Thanks!