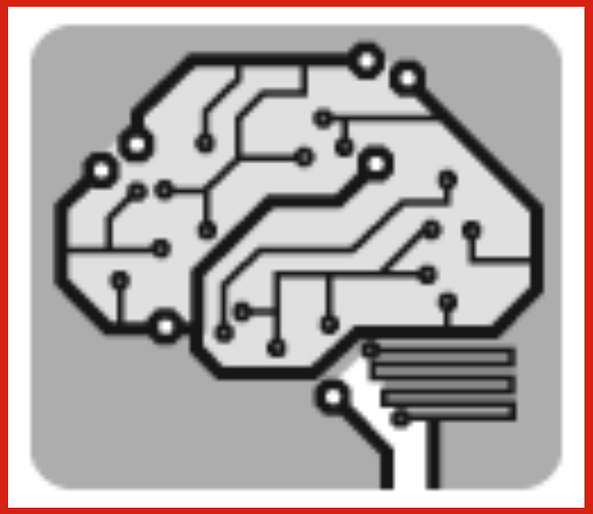
MLCC 2017

Machine Learning Crash Course

Universita' di Genova, Summer, 2017



Instructor: Lorenzo Rosasco

Organizers: Gian Maria Marconi, Fabio Anselmi,
Workshop organizer: Raffaello Camoriano

Dibris

The Center for Brains, Minds & Machines

Massachusetts Institute of Technology

iit @MIT ISTITUTO ITALIANO DI TECNOLOGIA
LABORATORY FOR COMPUTATIONAL AND STATISTICAL LEARNING

ML-GOA

SLIPGURU

STATISTICAL LEARNING & IMAGE PROCESSING
GENOA UNIVERSITY RESEARCH UNIT

Dibris

CompBio

Machine Learning

Computer Vision

Learning Theory

LCSL

Laboratory for Computational & Statistical Learning

iit @MIT

6+3 Faculty

7 PostDoc

~15 PhD+ master

From RegML to MLCC

RegML- Regularization Methods for Machine Learning

(baby 9.520@MIT)

ADVANCED

-2010, 35 attendees
-2011, 50 attendees
-2012, 50 attendees (@BISS)
-2013, 85 attendees
-2014, 95 attendees
-2016, 120 attendees
-2017, 80 attendeed (@OSLO)

MLCC- Machine Learning Crash Course

(baby ISML2@DIBRIS)

INTRO

-2014, 85 attendees
-2015, 120+ attendees
-2017, 120+ attendees

MLCC Objective

ML Desert Island Compilation

An introduction to ***essential*** Machine Learning:

- Concepts
- Algorithms

Course at a Glance

CLASS	DAY	TIME	SUBJECT
1	26/06	9:30 - 11:00	Introduction to Machine Learning
2	26/06	11:30 - 13:00	Local Methods and Model Selection
3	26/06	14:00 - 16:00	Laboratory 1: Local Methods for Classification
4	27/06	9:30 - 11:00	Regularization Networks I: Linear Models
5	27/06	11:30 - 13:00	Regularization Networks II: Kernels
6	27/06	14:00 - 16:00	Laboratory 2: Regularization Networks
Talk	28/06	9:30 - 10:10	*Pietro Leo*, Executive Architect & CTO, IBM Italy
Talk	28/06	10:10 - 10:35	*Enrico Ferrari*, R&D Manager, Rulex Inc.
Talk	28/06	10:35 - 11:00	*Matteo Santoro*, CEO, Camelot Biomedical Systems s.p.a.
	28/06	11:00 - 11:30	Coffee Break
Talk	28/06	11:30 - 12:10	*Giovanni Zappella*, Machine Learning Scientist, Amazon Developement Center Germany
Talk	28/06	12:10 - 12:35	*Luca Nardelli*, Founder and CTO @ Erya, Horus Technology
Talk	28/06	12:35 - 13:00	*Alessandro Verri*, Professor, University of Genoa
	28/06	Afternoon	Free
7	29/06	9:30 - 11:00	Dimensionality Reduction and PCA
8	29/06	11:30 - 13:00	Variable Selection and Sparsity
9	29/06	14:00 - 16:00	Laboratory 3: PCA and Sparsity
10	30/06	9:30 - 11:00	Clustering
11	30/06	11:30 - 13:00	Data Representation: Deep Learning

Day 1: Local Methods and Model Selection

Day 2: Regularization and nonparametrics

MLCC Workshop!

COMPANIES!

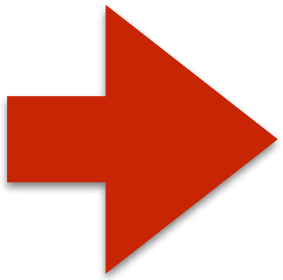
Day 3: Dimensionality Reduction and Sparsity

Day 4: DL & clustering

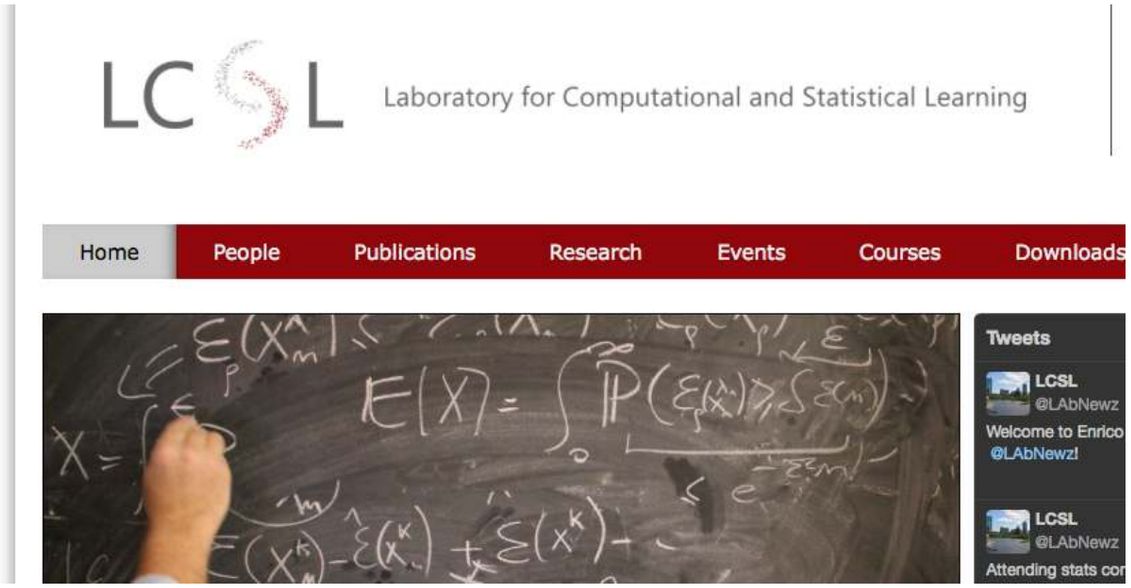
Note: Wed afternoon is vacation!

Prerequisites and References

Prerequisites: The mathematical tools needed for the course are basic probability, calculus and linear algebra.



lcsl.mit.edu



- L. Rosasco. Introductory Machine Learning Notes.

References:

- T. Hastie, R. Tibshirani, and J. Friedman. **The Elements of Statistical Learning: Prediction, Inference and Data Mining.** Second Edition, Springer Verlag, 2009 (available for free from the author's website).

Further readings :

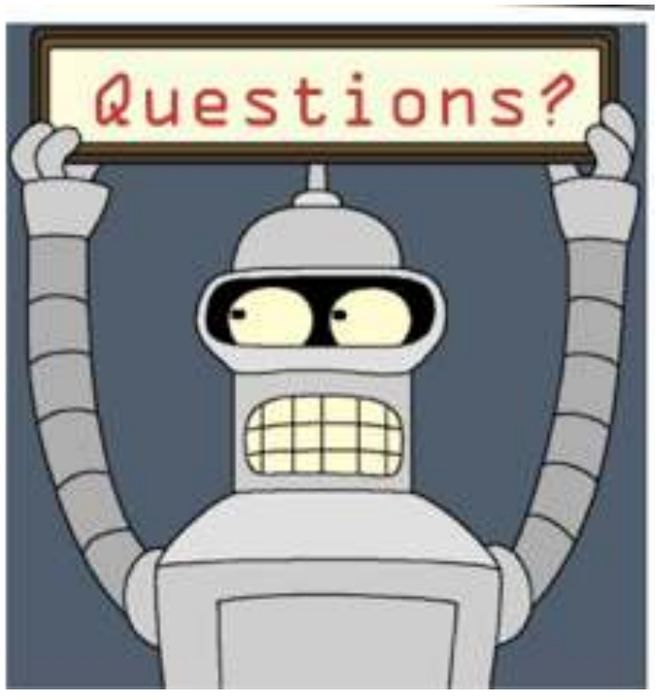
- T. Poggio and S. Smale. **The Mathematics of Learning: Dealing with Data.** Notices of the AMS, 2003
- Pedro Domingos. **A few useful things to know about machine learning.** Communications of the ACM CACM Homepage archive. Volume 55 Issue 10, October 2012 Pages 78-87.
- ....

Useful Links

- MIT 9.520: Statistical Learning Theory and Applications, Fall 2013 (http://www.mit.edu/~9.520/).
- Stanford CS229 Machine Learning Autumn 2013 (http://cs229.stanford.edu). See also the *Coursera* version (https://www.coursera.org/course/ml).

This Course Has a Rule

ASK

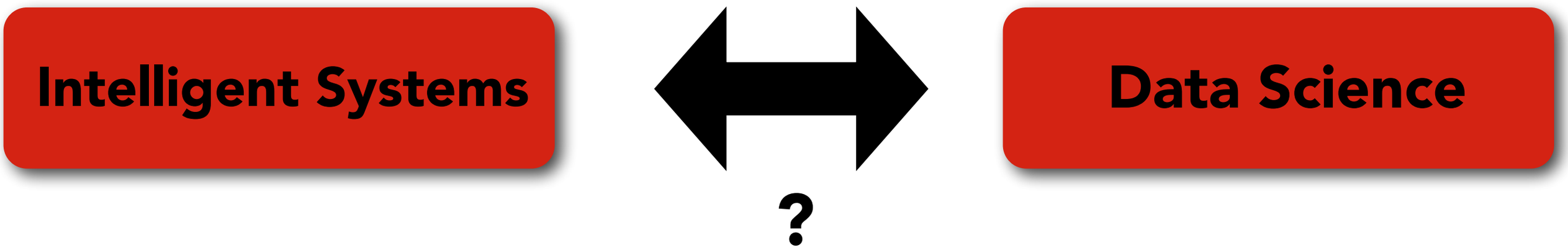


+attendance!

Today

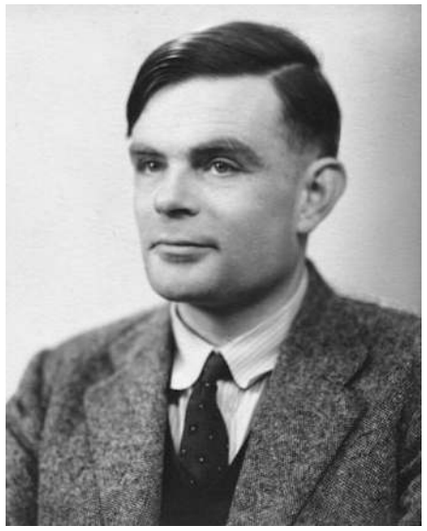
- A quick tour of machine learning
- Basic statistical learning theory
- Local algorithms
- Model selection

What is (Machine) Learning?



(Artificial) Intelligence: A Working Definition

Turing test



Alan Turing 1912-1954

Ingredients for AI

- natural language processing
- knowledge representation
- automated reasoning
- machine learning
- computer vision
- robotics to manipulate

A Glimpse to the Past

1943

Arturo Rosenblueth, Norbert Wiener and Julian Bigelow coin the term "cybernetics". Wiener's popular book by that name published in 1948.

…..

1948

John von Neumann (quoted by E.T. Jaynes) in response to a comment at a lecture that it was impossible for a machine to think: "You insist that there is something a machine cannot do. If you will tell me *precisely* what it is that a machine cannot do, then I can always make a machine which will do just that!". Von Neumann was presumably alluding to the Church-Turing thesis which states that any effective procedure can be simulated by a (generalized) computer.

...

1950

Alan Turing proposes the Turing Test as a measure of machine intelligence.

1950

Claude Shannon published a detailed analysis of chess playing as search.

1955

The first Dartmouth College summer AI conference is organized by John McCarthy, Marvin Minsky, Nathan Rochester of IBM and Claude Shannon.

.....

Late 1990s

Web crawlers and other AI-based information extraction programs become essential in widespread use of the World Wide Web.

1997

The Deep Blue chess machine (IBM) beats the world chess champion, Garry Kasparov.

….

Type to enter text

10/15 years ago

How are we doing now?

Pedestrians Detection at Human Level Performance

Speech Recognition

How do we do this???

Big Data revolution

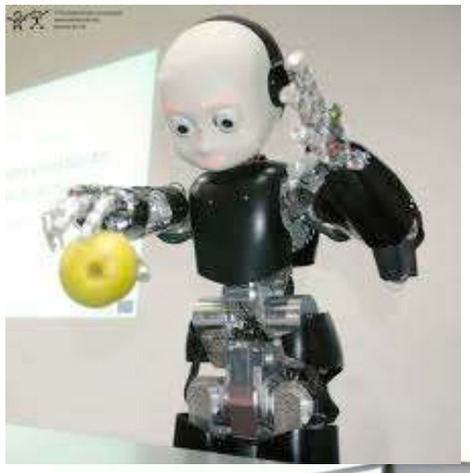
DATA

Computers

“It takes these very simple-minded instructions—‘Go fetch a number, add it to this number, put the result there, perceive if it’s greater than this other number’—but executes them at a rate of, let’s say, 1,000,000 per second. At 1,000,000 per second, the results appear to be magic.” [**Playboy, Feb. 1, 1985**]

Type to enter text

+Machine Learning

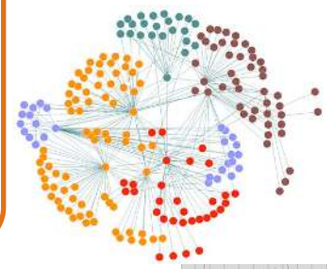


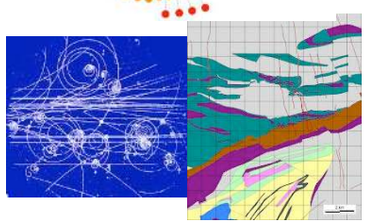
Machine Learning

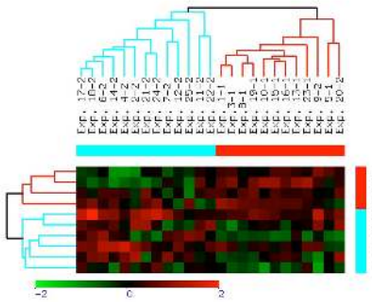
systems

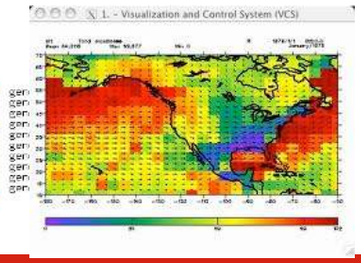
learn from data

rather than being programmed









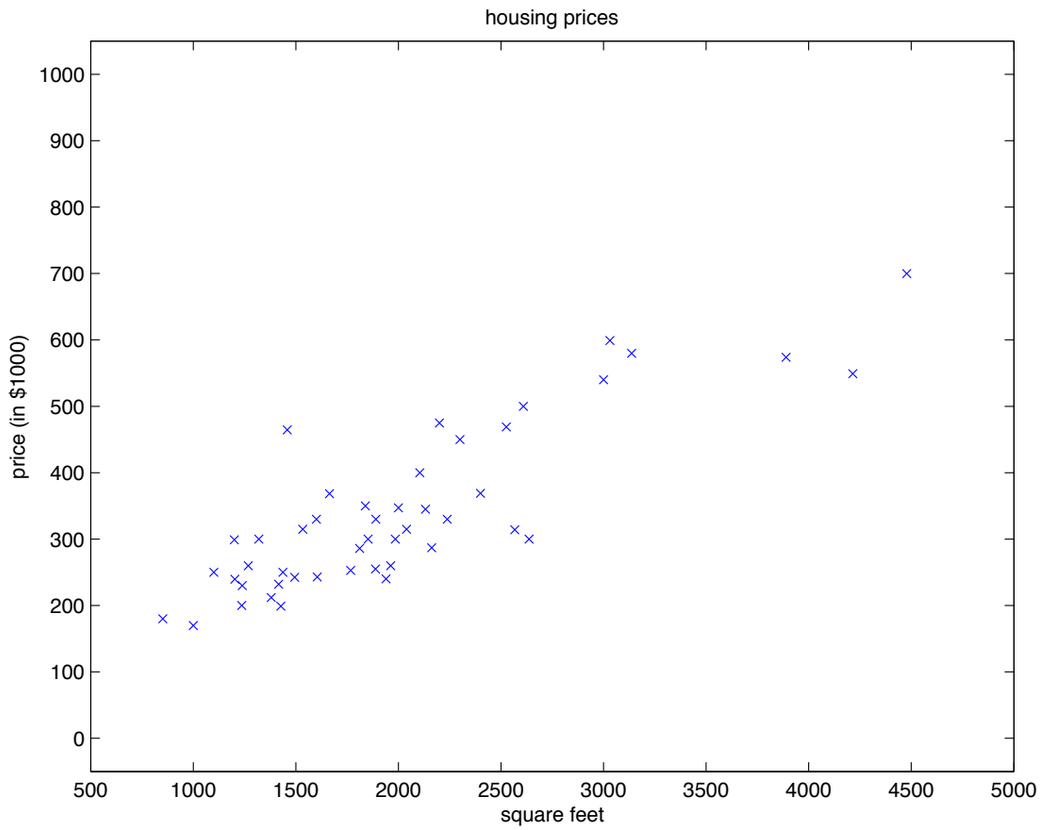
Regression

example taken from Coursera

DATA

Living area (feet^2)	Price (1000$s)
2104	400
1600	330
2400	369
1416	232
3000	540
⋮	⋮

$$(x_1, y_1), \ldots, (x_n, y_n)$$



Living area (feet^2)	#bedrooms	Price (1000$s)
2104	3	400
1600	3	330
2400	3	369
1416	2	232
3000	4	540
⋮	⋮	⋮

Text Classification

» 2007-05-22 18:33 - News in English

Soccer: Milan weighs attack options

Gilardino tipped to lead the line, Inzaghi late card

Sport (football)

» 2007-05-22 17:29 - News in English

Wine wards off senile dementia

A glass a day stops mild impairment worsening, Italians say

Health

» 2007-05-22 16:27 - News in English

Treasury may sell all of Alitalia

Formal bids for carrier due by July 2

Politics (economy)

» 2007-05-22 15:58 - News in English

Afghanistan: Rome cold on Bush call

'We respcnd to parliament, no one else, ' FM says

Politics (foreign)

» 2007-05-22 14:11 - News in English

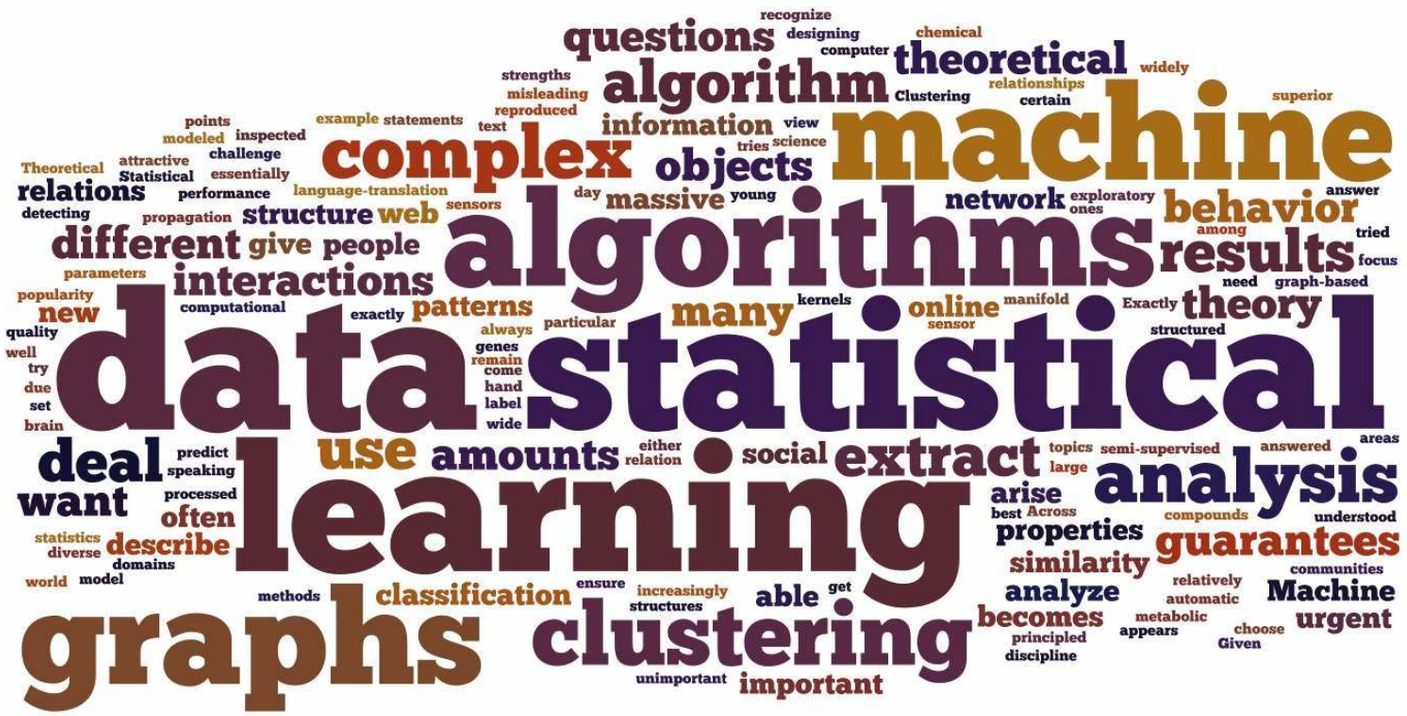
Electronic nose sniffs asthma

Device developed by Italian researcher in Netherlands

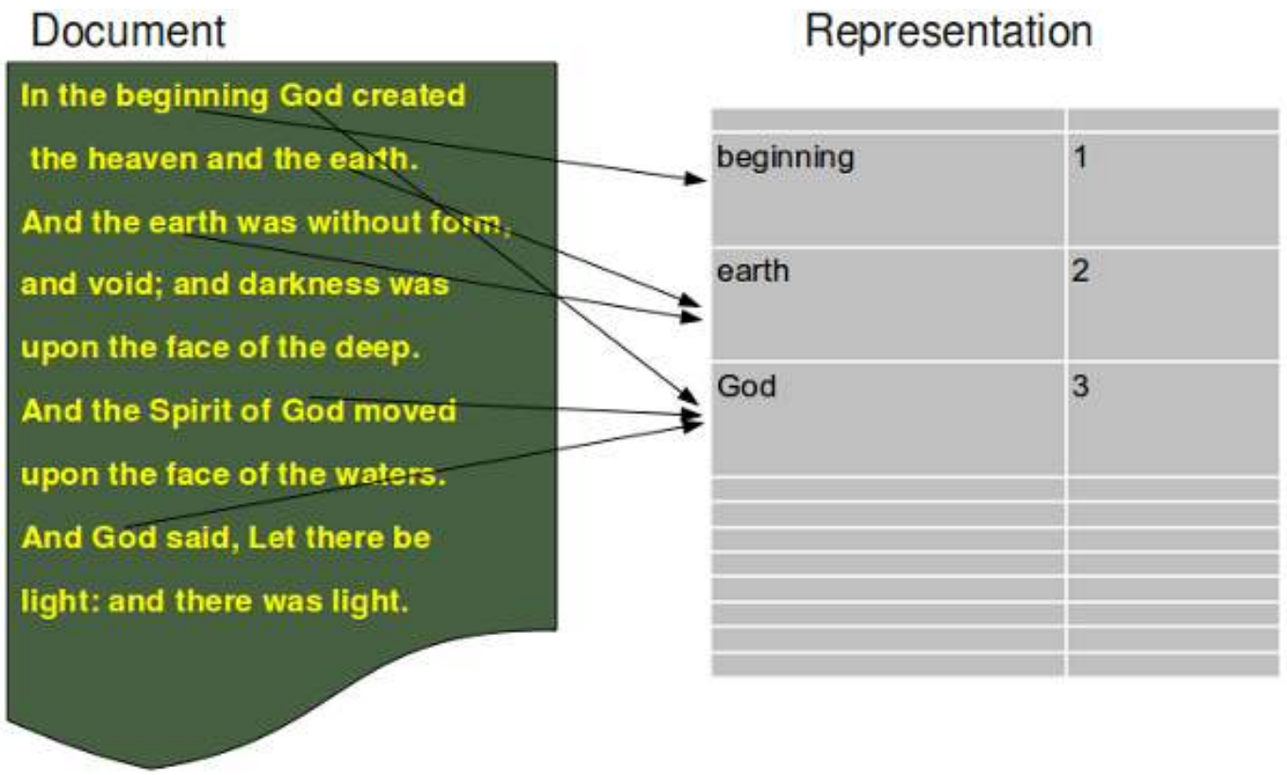
Technology

Subject	Date	Time	Body	Spam?
I has the viagra for you	03/12/1992	12:23 pm	Hi! I noticed that you are a software engineer so here's the pleasure you were looking for…	Yes
Important business	05/29/1995	01:24 pm	Give me your account number and you'll be rich. I'm totally serial	Yes
Business Plan	05/23/1996	07:19 pm	As per our conversation, here's the business plan for our new venture Warm regards…	No
Job Opportunity	02/29/1998	08:19 am	Hi !I am trying to fill a position for a PHP …	Yes
[A few thousand rows ommitted]				
Call mom	05/23/2000	02:14 pm	Call mom. She's been trying to reach you for a few days now	No

Text Classification: Bag of Words

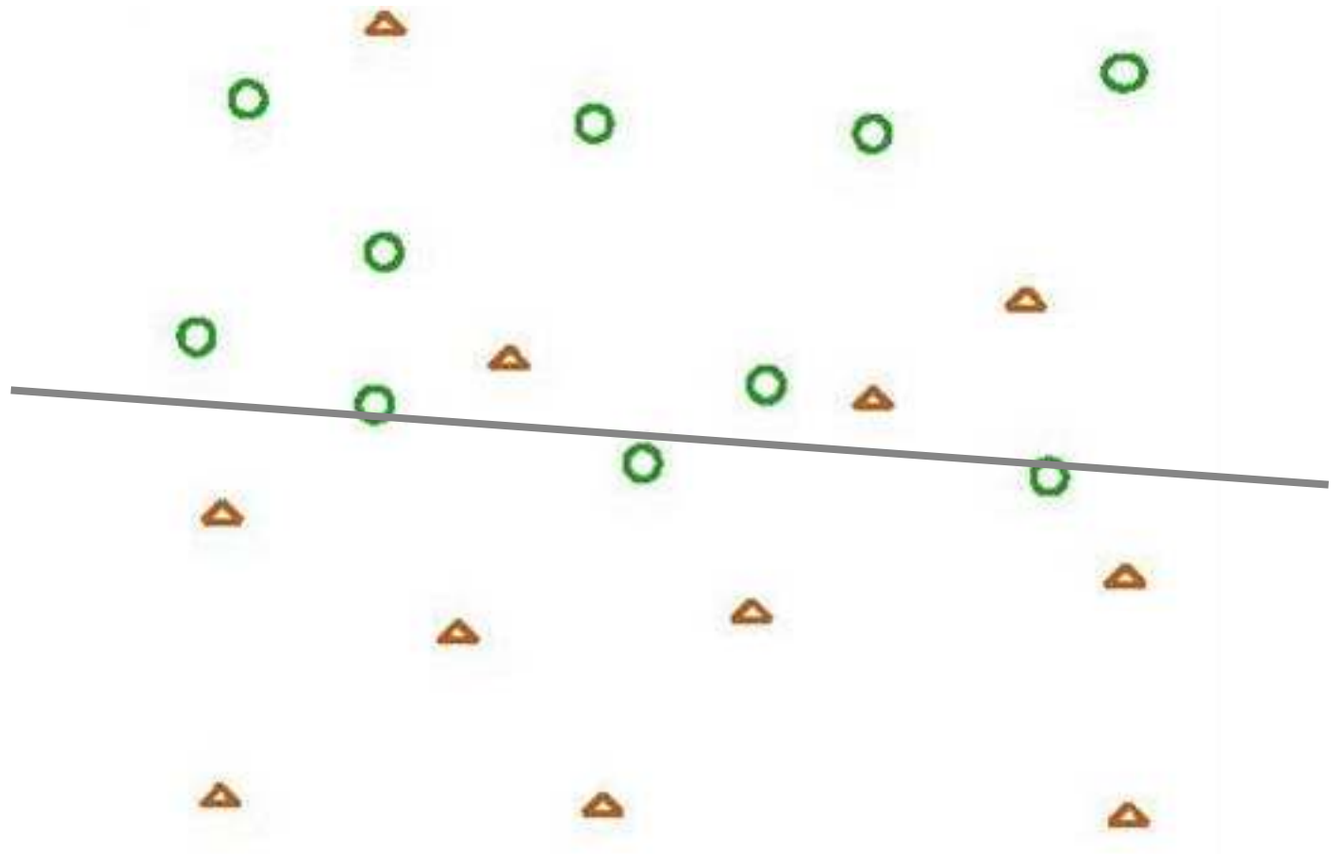


$$Y_n = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}$$



$$X_n = \begin{pmatrix} x_1^1 & \dots & \dots & \dots & x_1^p \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ x_n^1 & \dots & \dots & \dots & x_n^p \end{pmatrix}$$

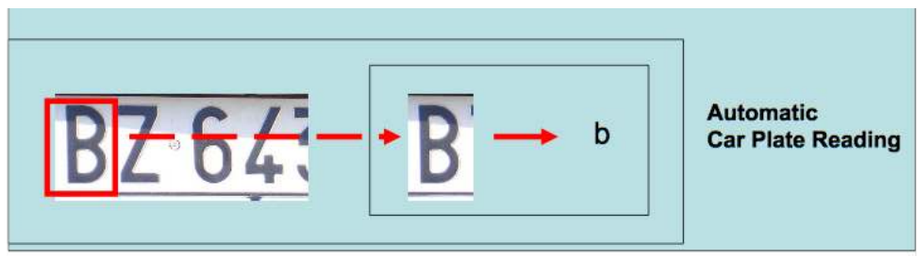
Basic Setting: Classification



$$(x_1, y_1), \ldots, (x_n, y_n)$$

$$X_n = \begin{pmatrix} x_1^1 & \ldots & \ldots & \ldots & x_1^p \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ x_n^1 & \ldots & \ldots & \ldots & x_n^p \end{pmatrix} \qquad Y_n = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}$$

Image Classification



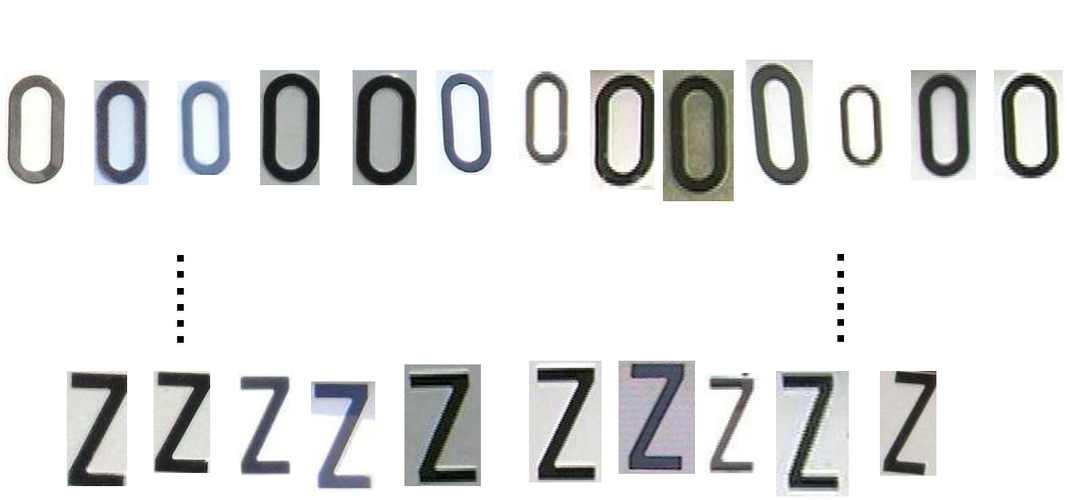
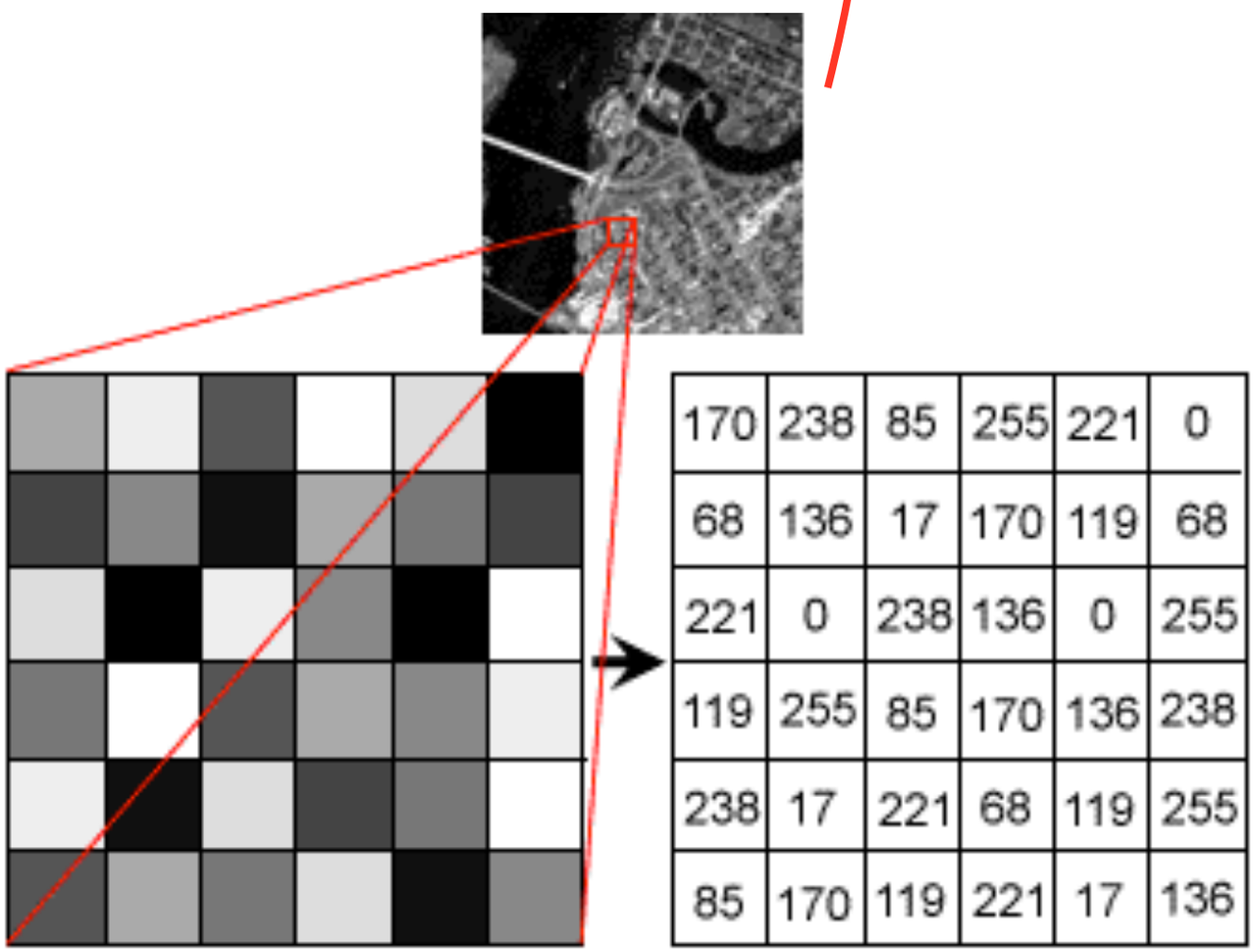


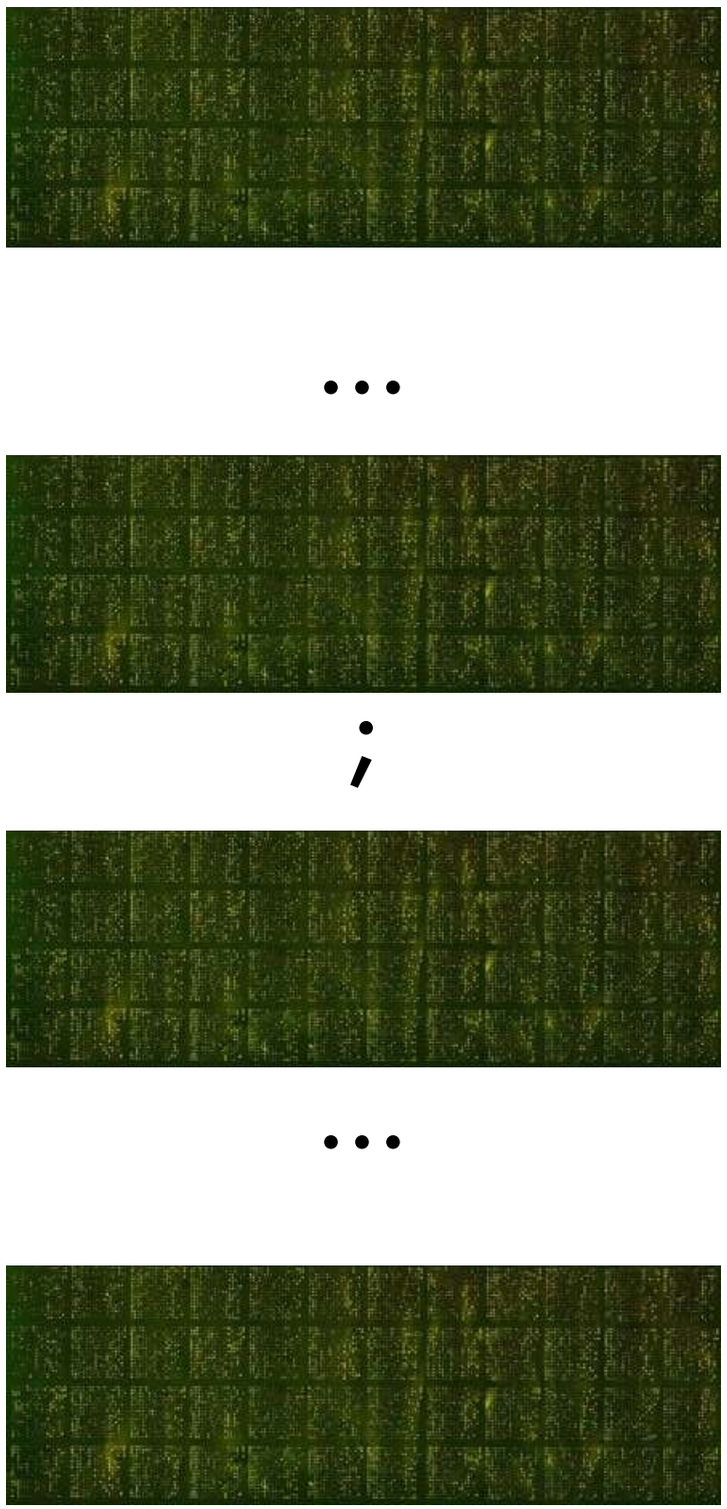
Image Classification

$$X_n = \begin{pmatrix} x_1^1 & \dots & \dots & \dots & x_1^p \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ x_n^1 & \dots & \dots & \dots & x_n^p \end{pmatrix}$$



170	238	85	255	221	0
68	136	17	170	119	68
221	0	238	136	0	255
119	255	85	170	136	238
238	17	221	68	119	255
85	170	119	221	17	136

Biology

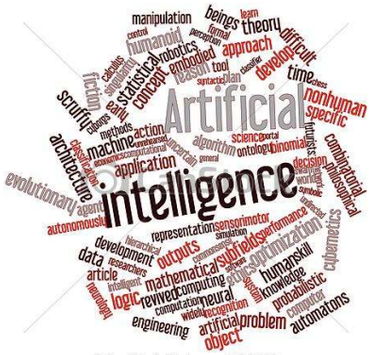


n patients p gene expression measurements

$$X_n = \begin{pmatrix} x_1^1 & \dots & \dots & \dots & x_1^p \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ x_n^1 & \dots & \dots & \dots & x_n^p \end{pmatrix}; Y_n = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}$$

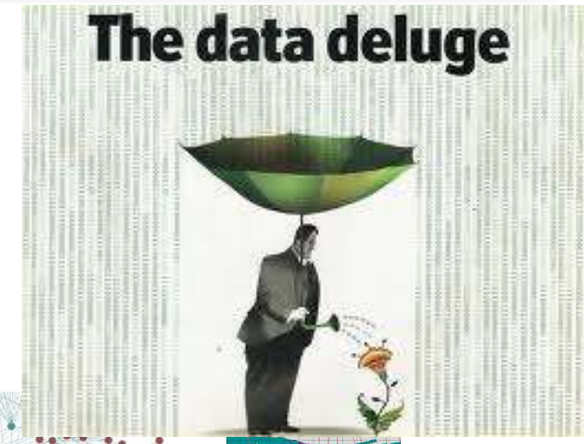
Machine Learning

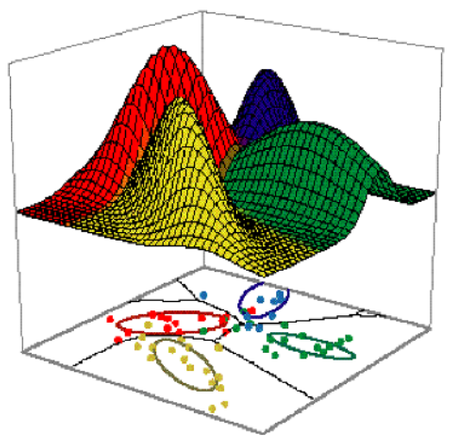
Intelligent Systems

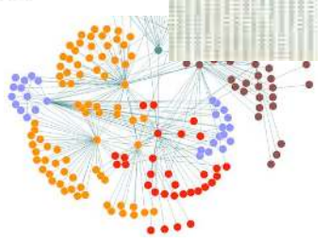


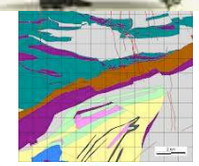
© Can Stock Photo - csp11714066

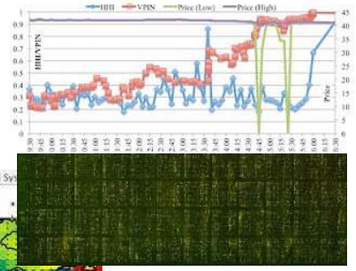
Data Science

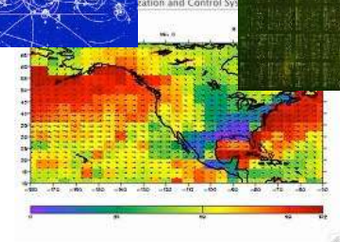












Today

- A quick tour of machine learning ✓
- Basic statistical learning theory

- Local algorithms
- Model selection