DIPARTIMENTO
DI INGEGNERIA
DELL'INFORMAZIONE

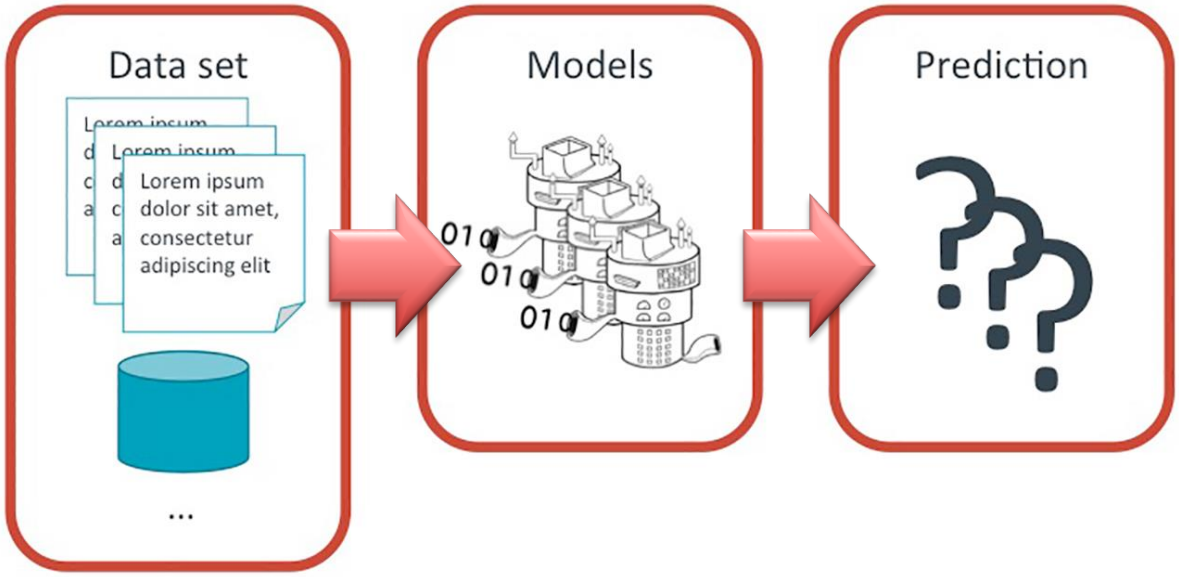
UNIVERSITÀ
DEGLI STUDI
DI PADOVA

Machine Learning

Course Introduction 2023-24

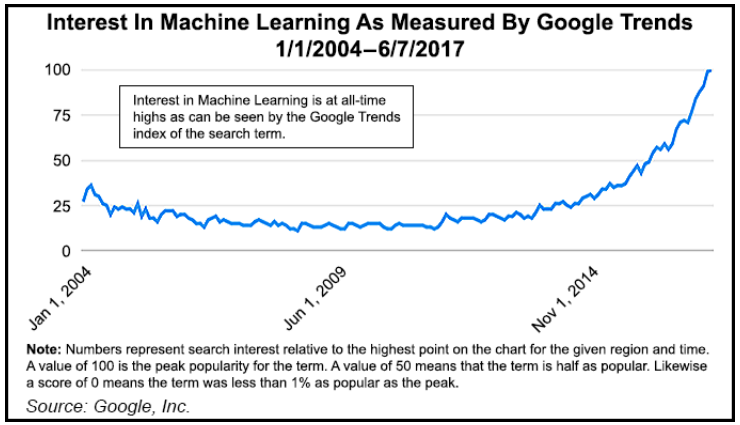
ICT for Internet and Multimedia and Physics of Data
Channel B – **ID last digit from 5 to 9** - P. Zanuttigh

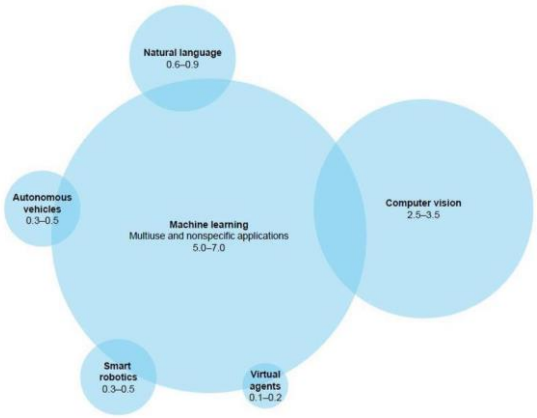
What is Machine Learning ?

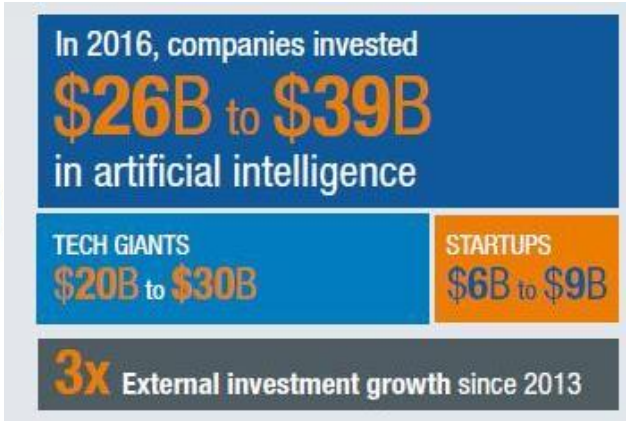


Machine learning (ML) is a set of methods that give computer systems the ability to "*learn*" from (*training*) data to make predictions about novel data samples, *without being explicitly programmed*

Why a Machine Learning Exam ?

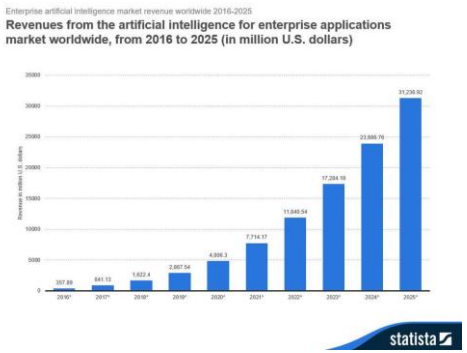


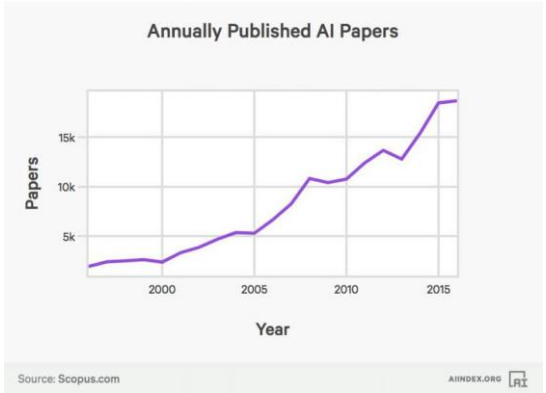


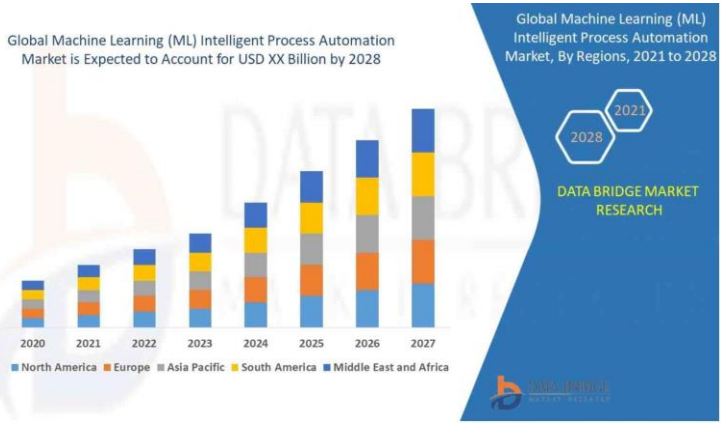


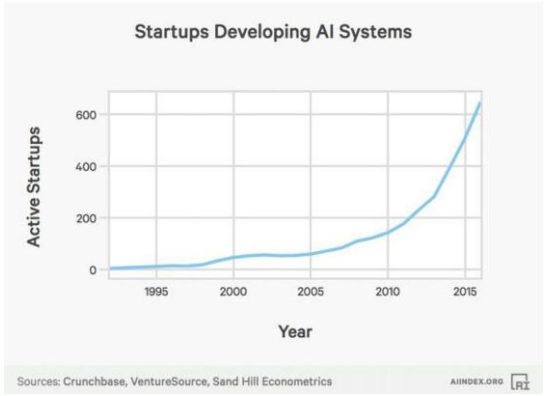
- Many applications of ML with outstanding results
- Huge investments from all around the world
- Availability of large amount of data and computational resources (e.g., GPUs)
- The global machine learning market was valued at **USD 2.40 Billion in 2019** and is projected to **reach USD 47.29 Billion by 2027**, growing at a rate of 44.9% from 2020 to 2027. *Technological advancement is the major driving factor for the global machine learning market*
- Annual global AI software revenue will grow from **$10.1B** in 2018 to **$126.0B** by 2025, at a rate of 43.41% *(from Tractica)*

A Fast Growing Field



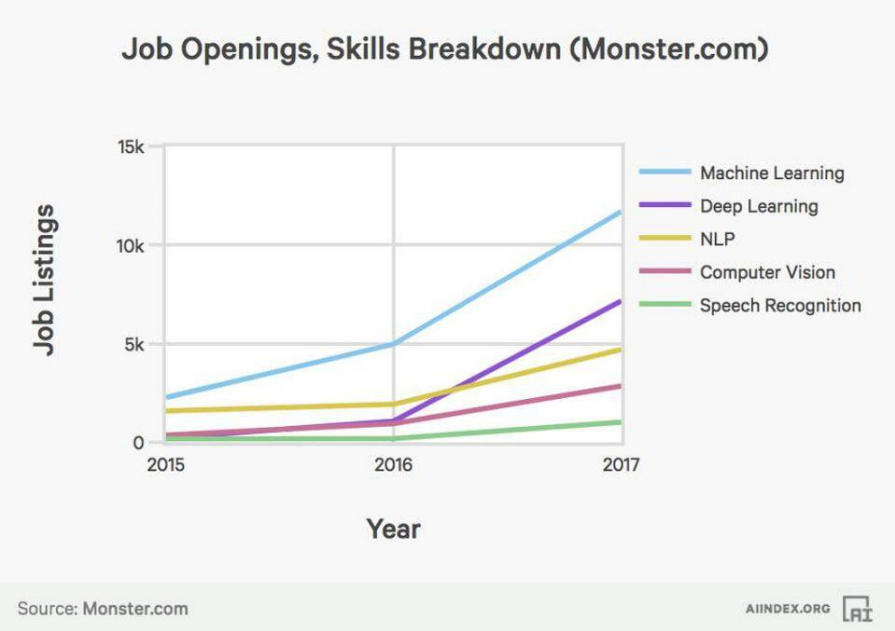


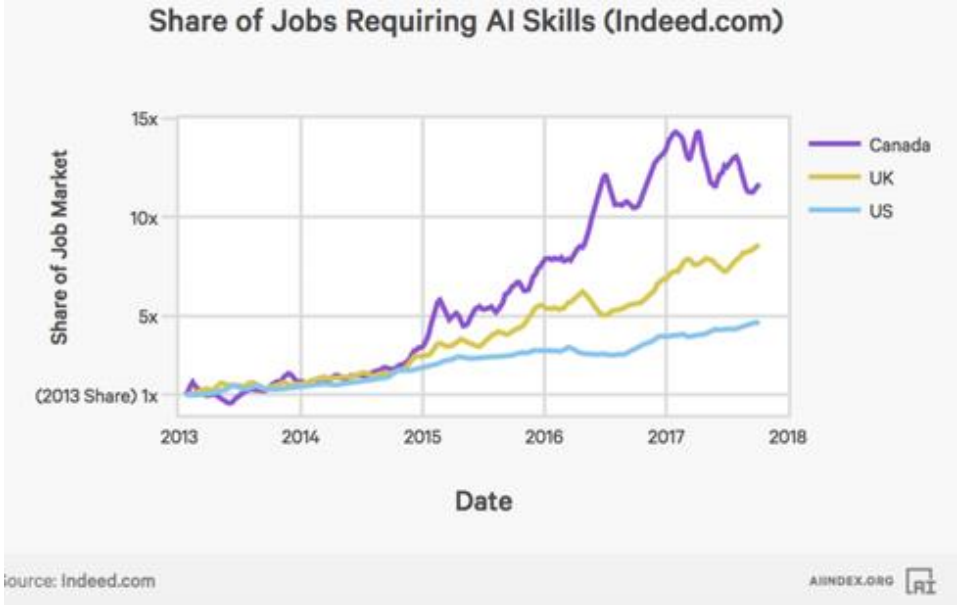




- The research on ML techniques is growing very fast both in the academia and in R&D departments of companies
- Both big companies and many fast growing start-ups involved in ML
- Larger and larger market for ML applications

Many Jobs in the Field



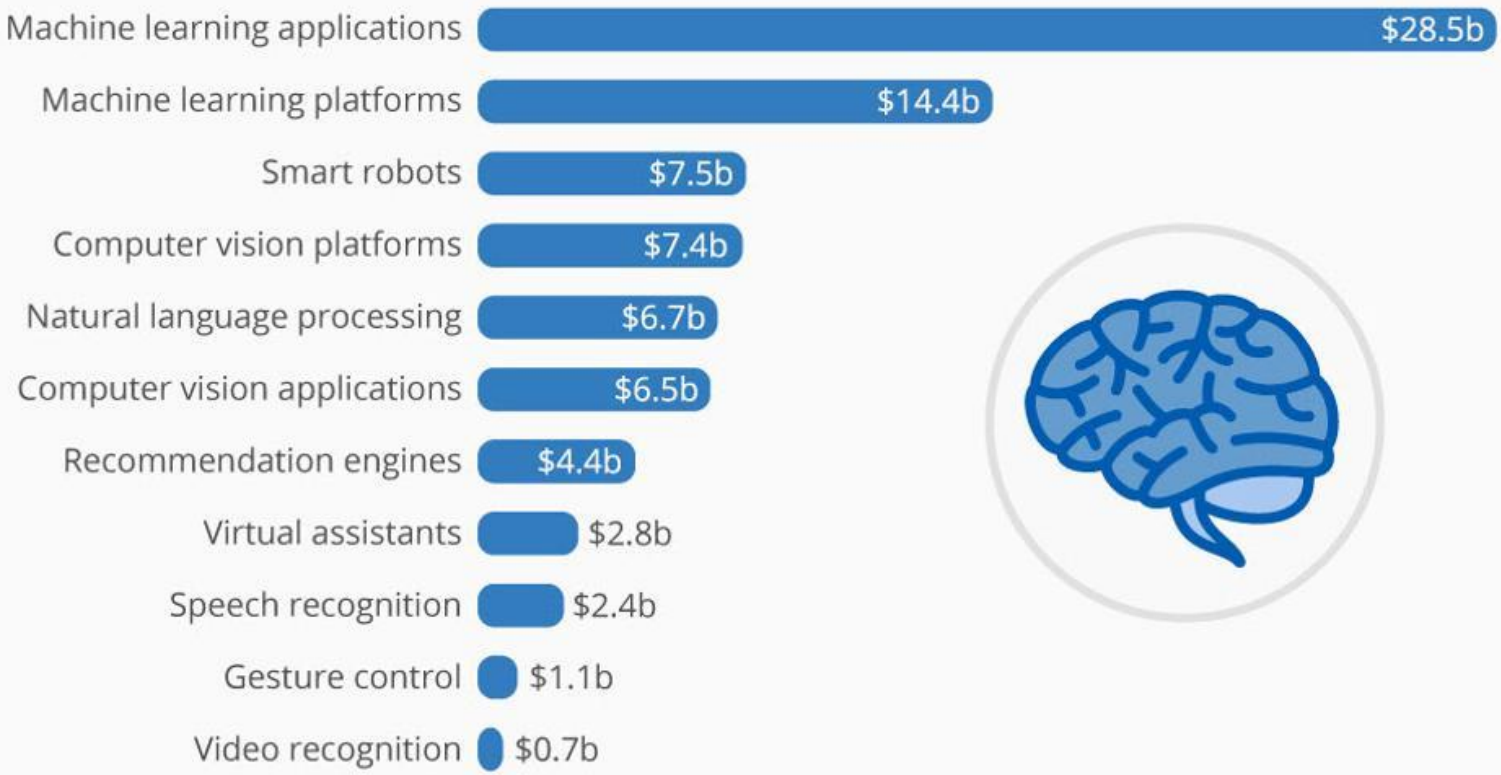


- Many different job opportunities both in Italy and abroad
- Companies struggle in finding ML experts
- According to a recent study machine learning engineering is the best paid job in the United States with an average salary of $146,085

Investments in ML

Machine Learning Tops AI Dollars

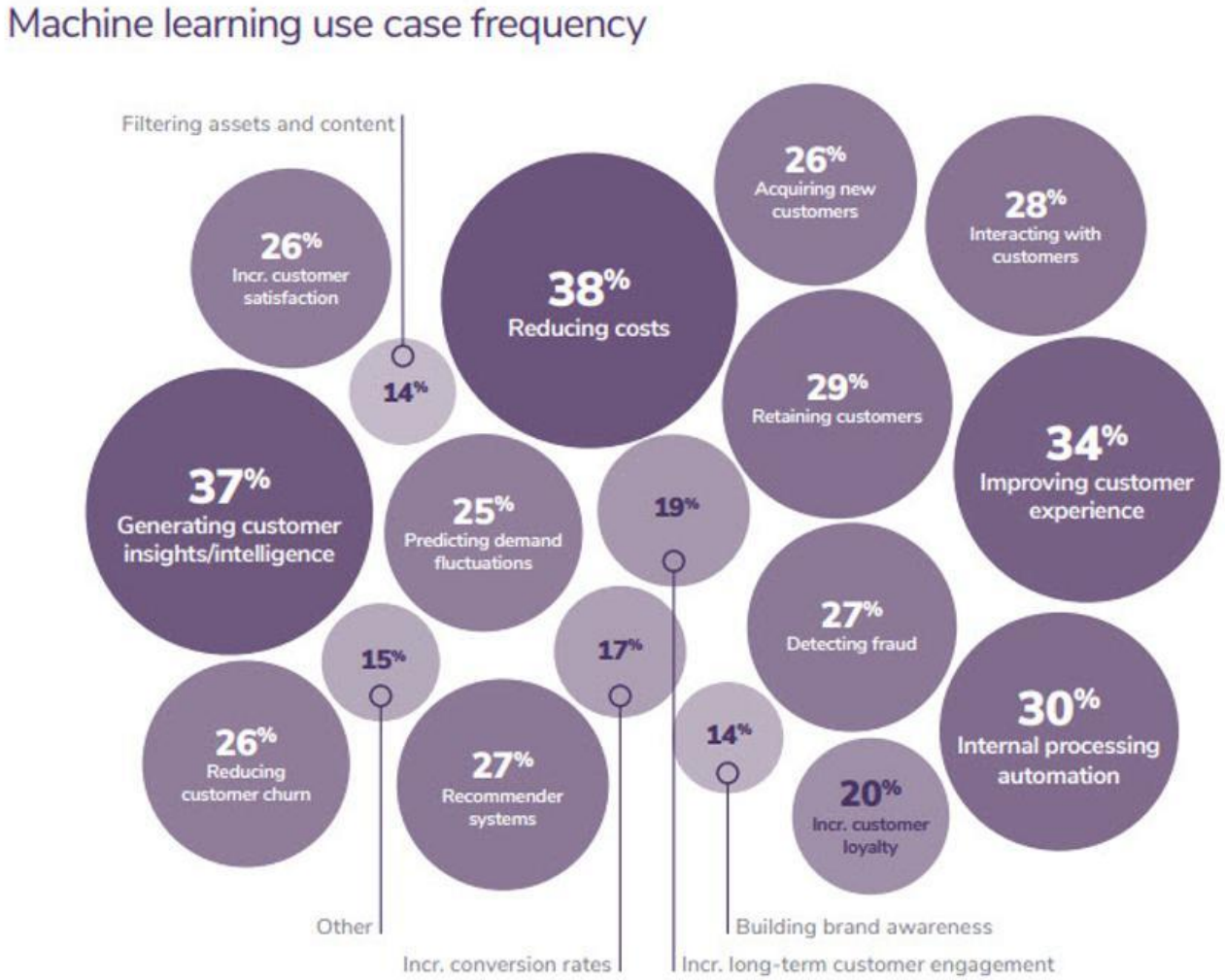
AI funding worldwide cumulative through March 2019 (in billion U.S. dollars), by category



Sources: Venture Scanner, Statista estimates

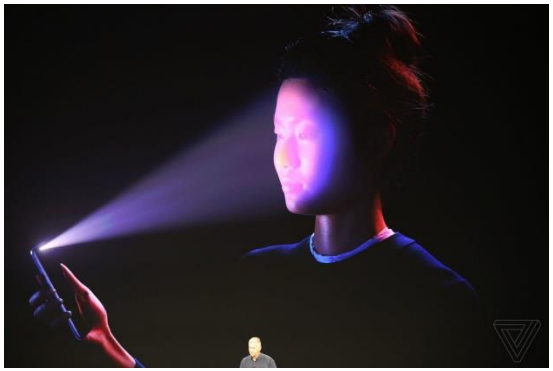
statista

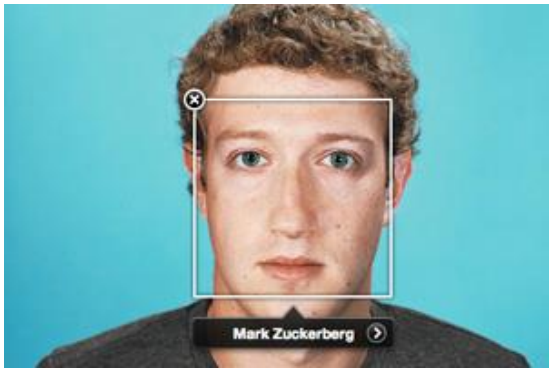
Where is ML used ?

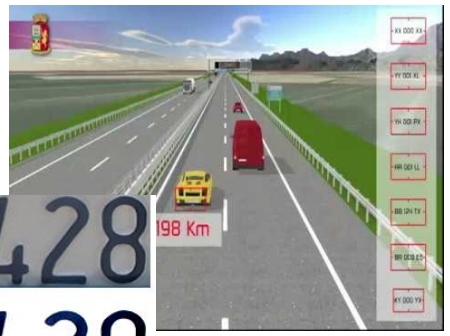


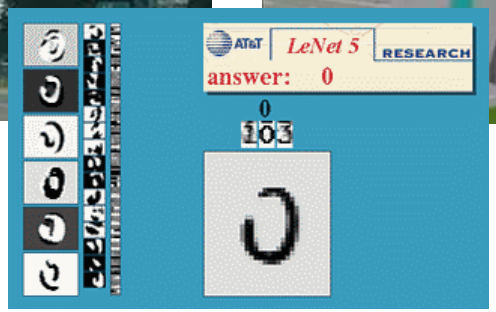
DIPARTIMENTO
DI INGEGNERIA
DELL'INFORMAZIONE

Many Applications: Image Understading

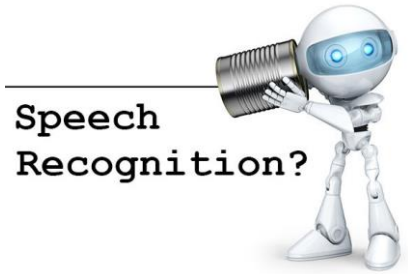


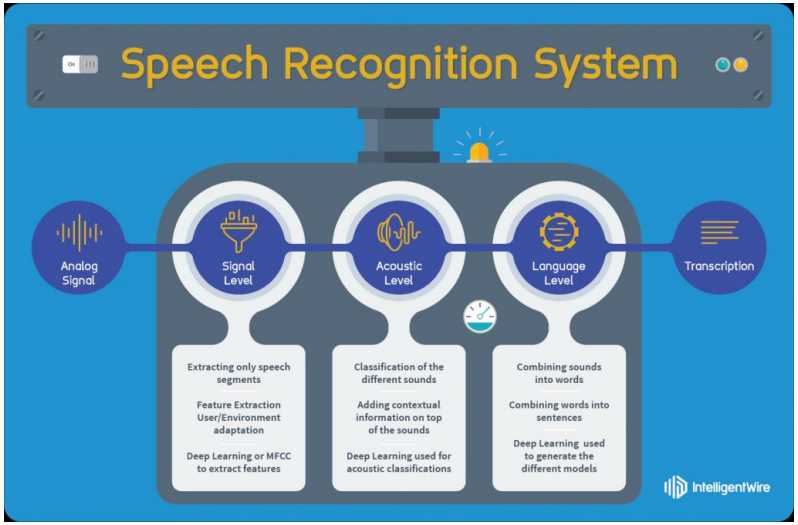


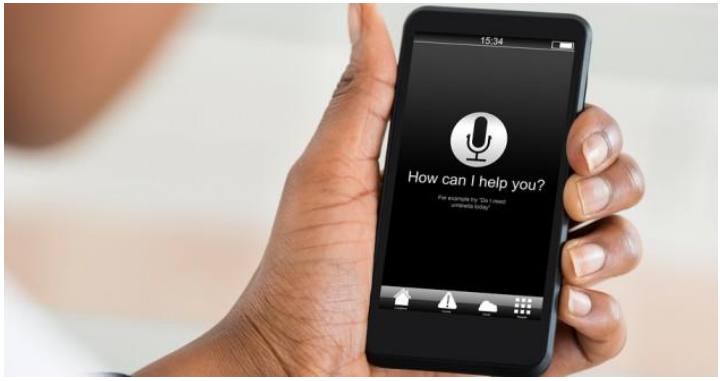




Many Applications: Speech Recognition







Google Home

amazon alexa

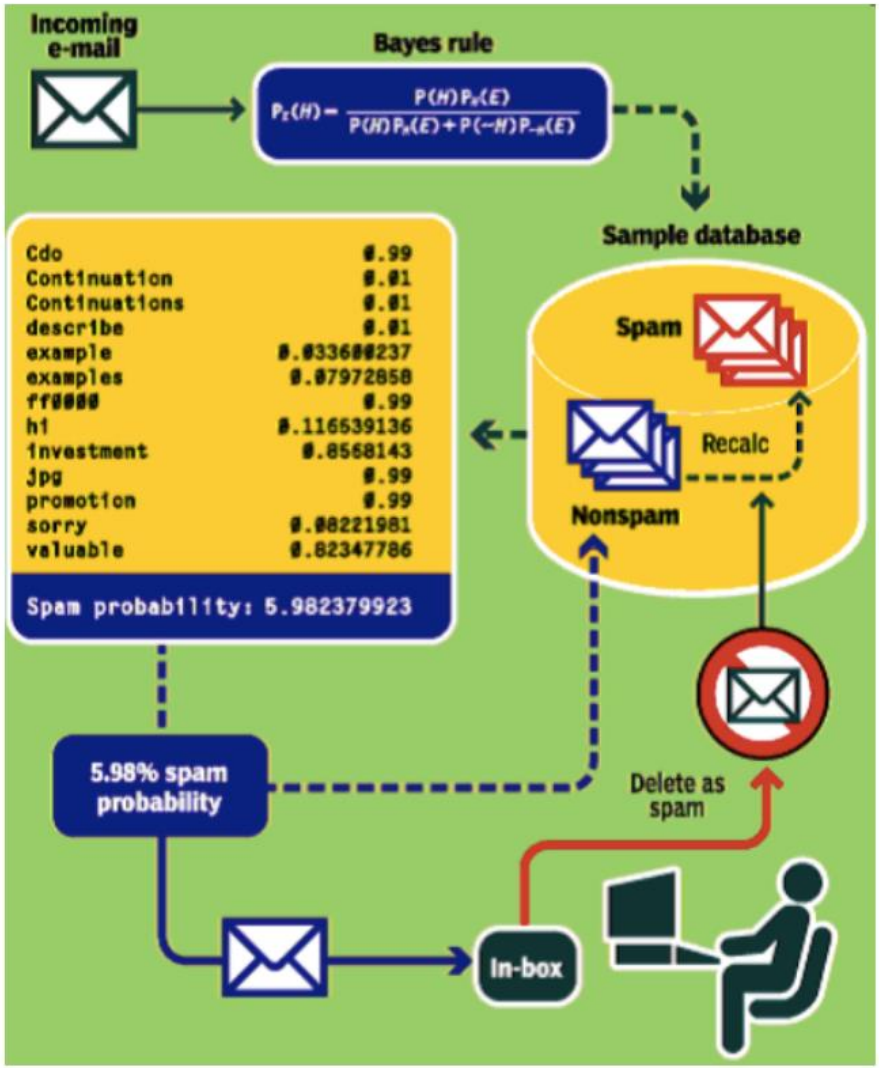
Many Applications: Spam Filtering

Question: Is this e-mail *useful* (ham) or *spam*?

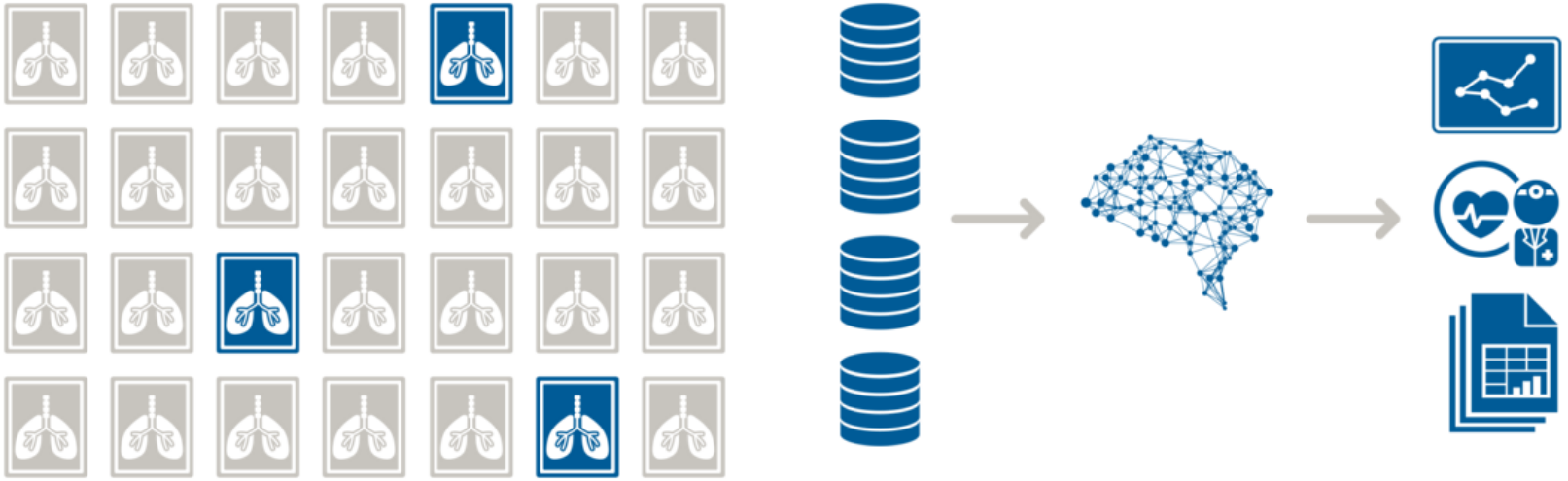
Challenge: There is no simple universal rule to define spam

(Noisy) data: messages previously marked as spam by user

Challenge: Spammers evolve to counter filter innovations

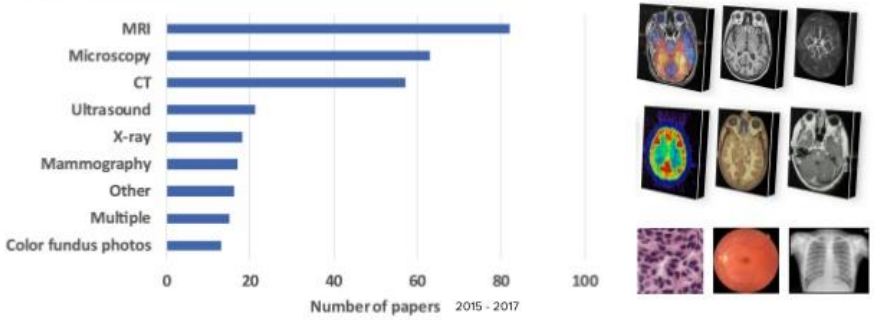


Many Applications: Medical Imaging



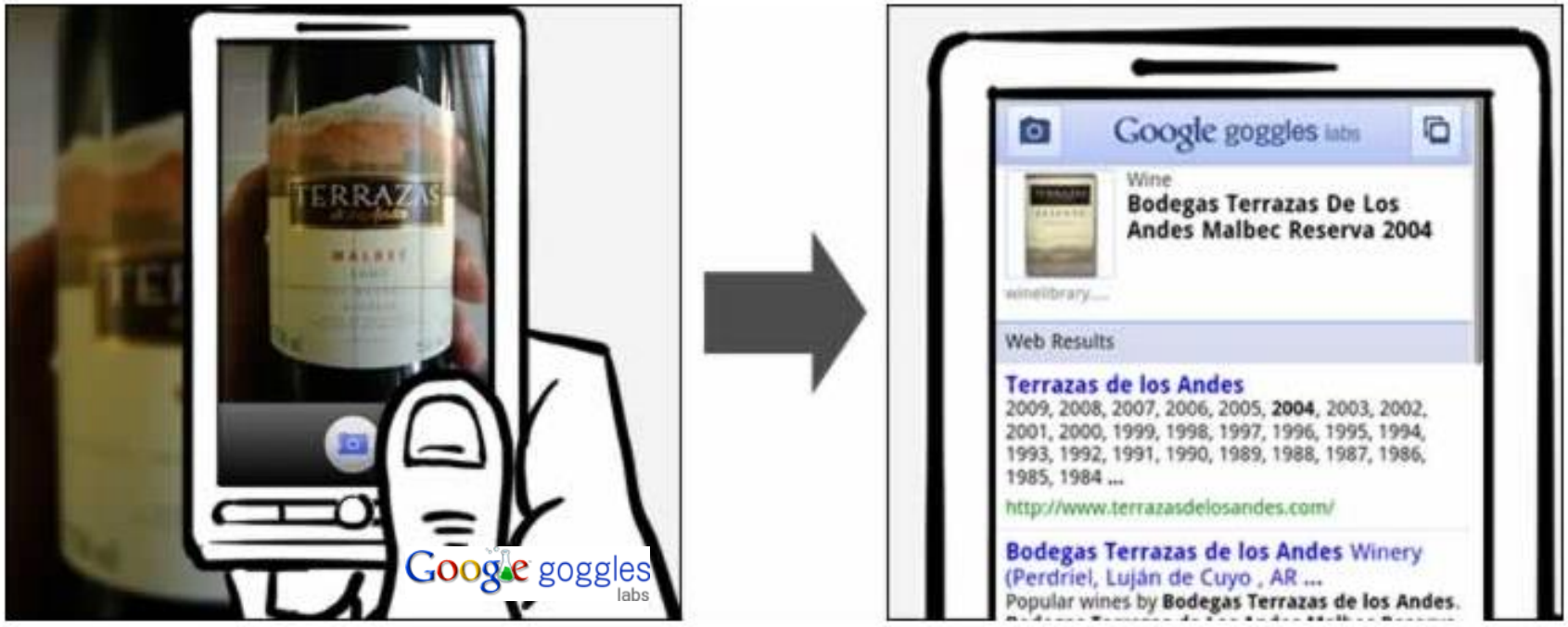
Why deep learning for medical imaging?

Image modalities



Litjens et al, A Survey on Deep Learning in Medical Image Analysis, June 4th, 2017

Many Applications: Search Engines

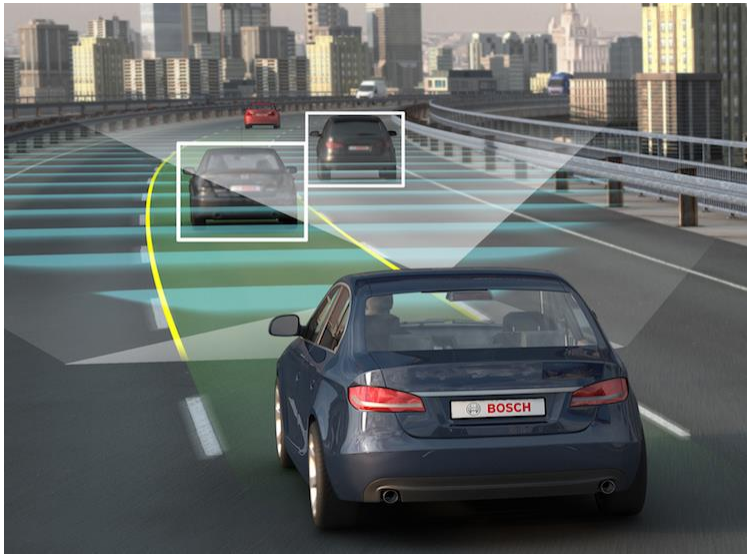


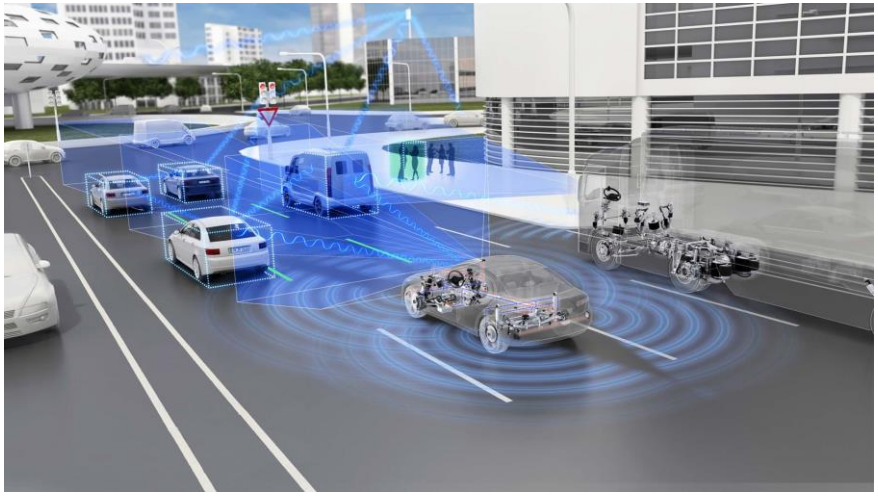
Many Applications: Recommender Systems

- Recommendations are driven by machine learning algorithms
- Over **80%** of Netflix users select films recommended to them by the company's machine learning algorithms

DIPARTIMENTO
DI INGEGNERIA
DELL'INFORMAZIONE

Many Applications: Autonomous Driving





Source: A. Shashua, S. Seitz

DIPARTIMENTO
DI INGEGNERIA
DELL'INFORMAZIONE

Applications: Generating Text, Images and Many Other Things!

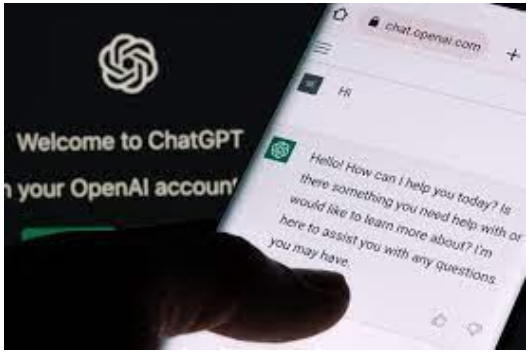


Figure 3. Example results by our proposed StackGAN, GAWWN [20], and GAN-INT-CLS [22] conditioned on text descriptions from CUB test set. GAWWN and GAN-INT-CLS generate 16 images for each text description, respectively. We select the best one for each of them to compare with our StackGAN.

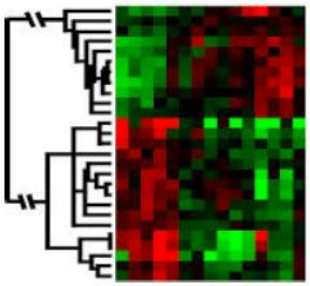
Did You Know ?!

DeepMind: AI company *(developed software which "learn" to play videogames)* acquired by Google in 2014 (now Google DeepMind) for 500 M$!!!

AlphaGo, developed by GoogleDeepMind in October 2015 beats European Champion of Go

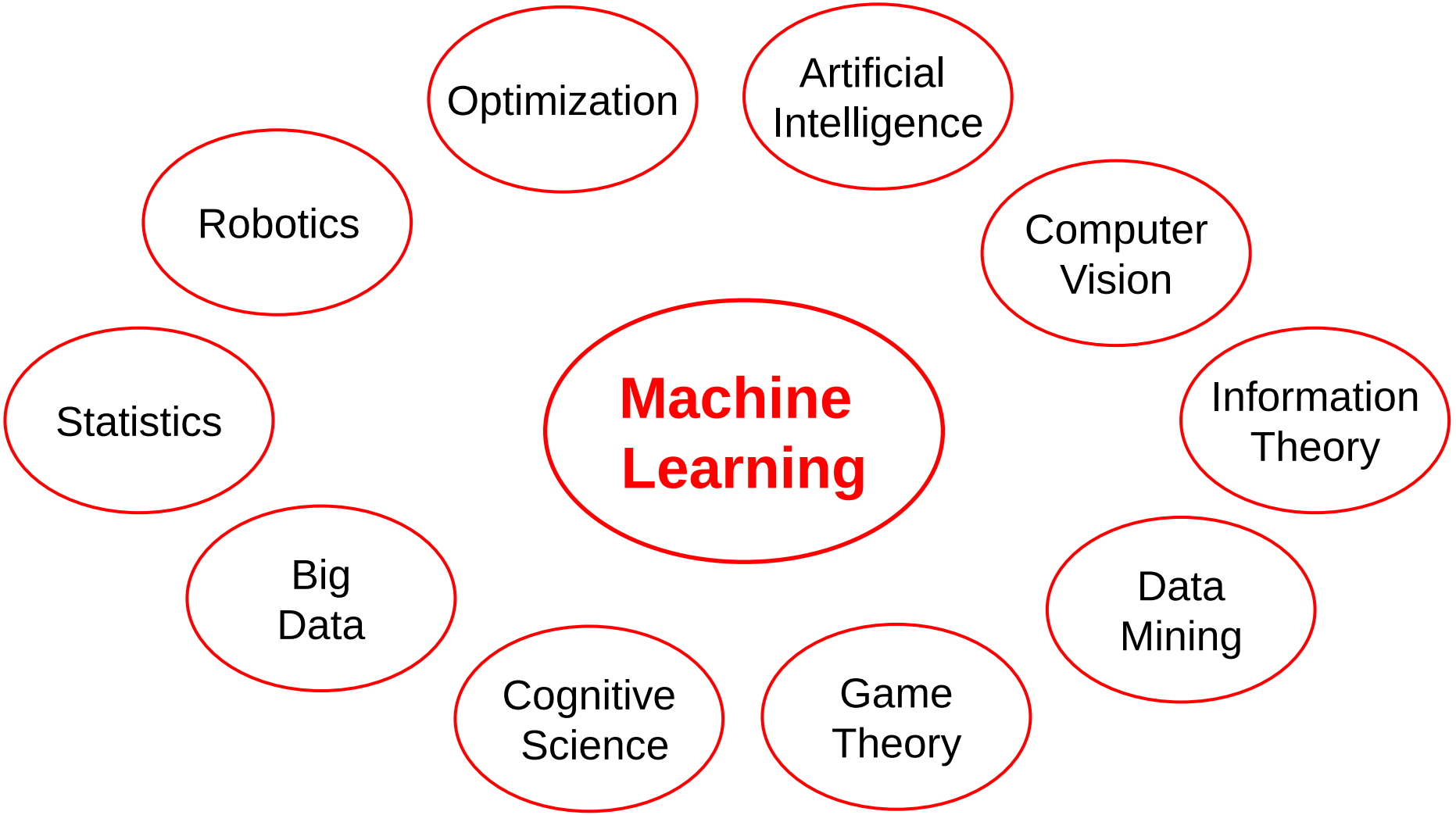
When is ML useful ?



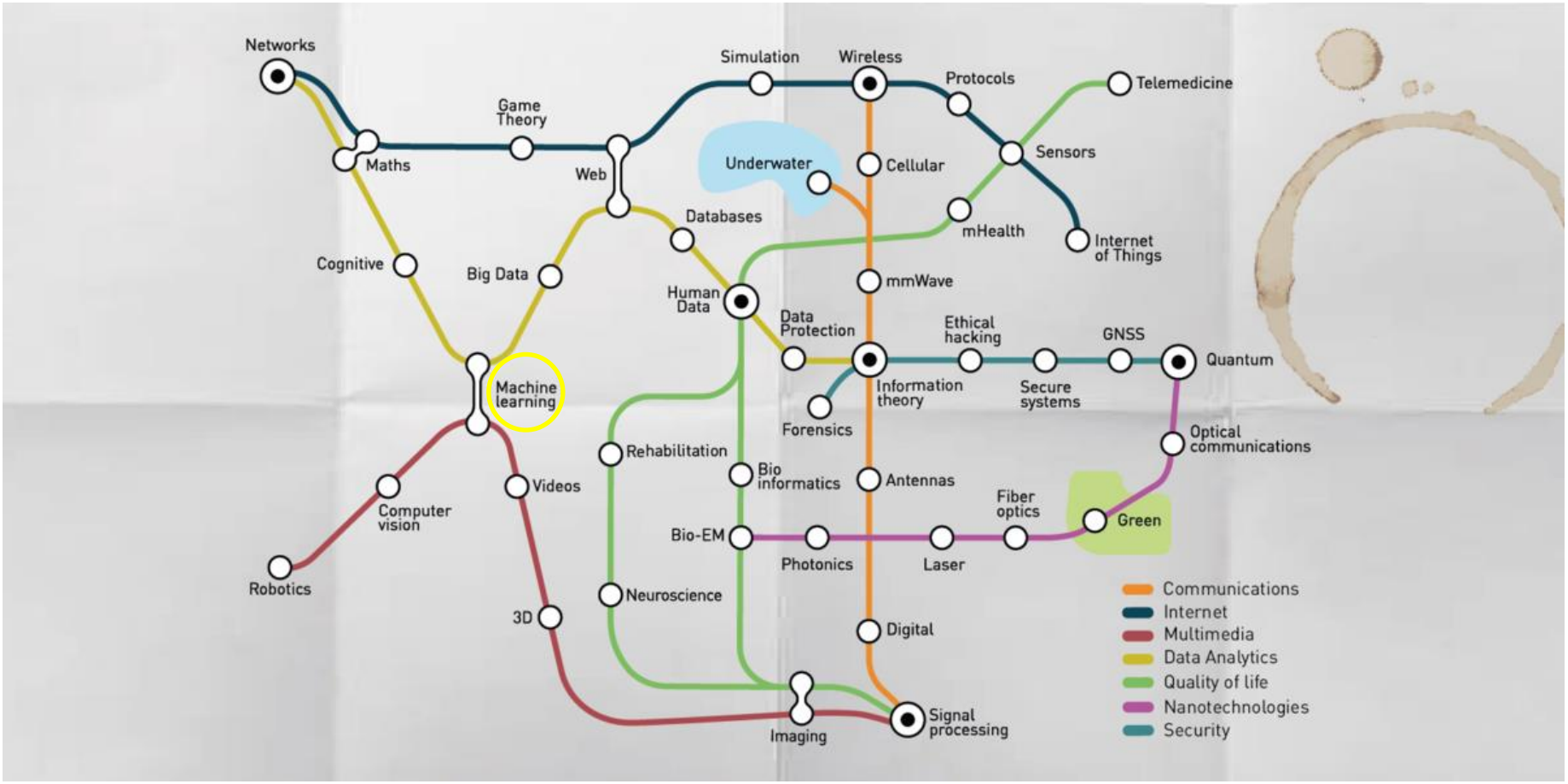

- Humans can't explain their expertise (speech recognition, image understanding)
- Tasks that humans can't solve, e.g., models are based on huge amounts of data (genomics, social media analysis)
- Models that must be customized or adapted (handwriting recognition, personalized medicine, spam filtering)
- Learning isn't always useful (there is no need to "learn" to calculate payroll)

Many Connections to Other Disciplines



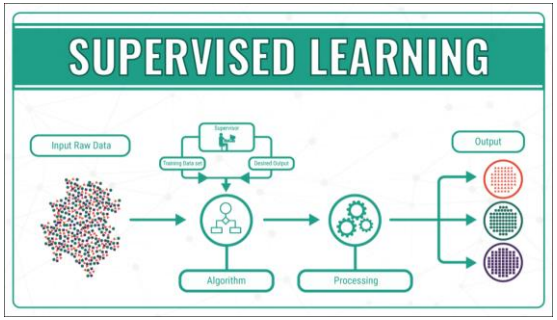
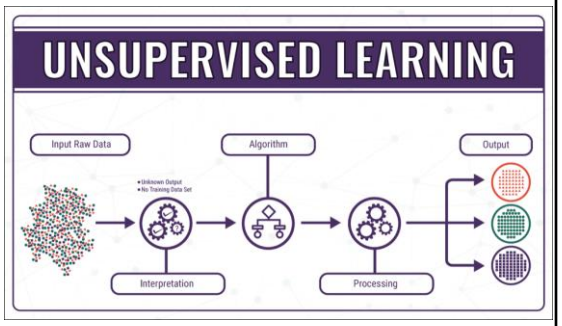
DIPARTIMENTO
DI INGEGNERIA
DELL'INFORMAZIONE

Related Courses (ICT)



… *ML techniques are used in many other courses*

Course Contents

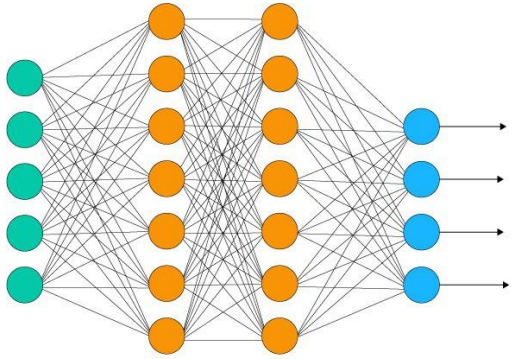
		
Supervised Learning	Unsupervised Learning	Laboratories

Supervised Learning

Training data

Training procedure

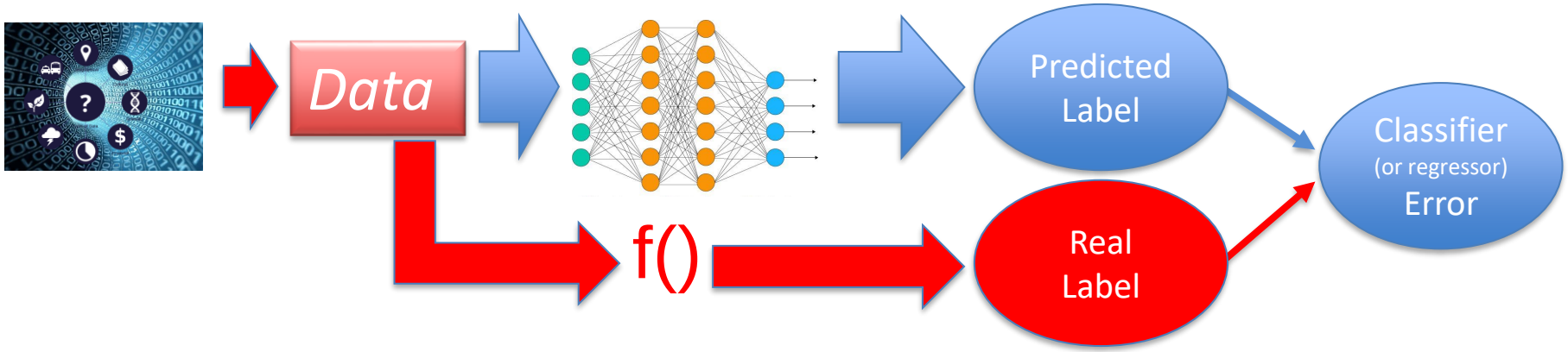
Data to be analyzed



ML model
(estimate parameters)

Predicted Label

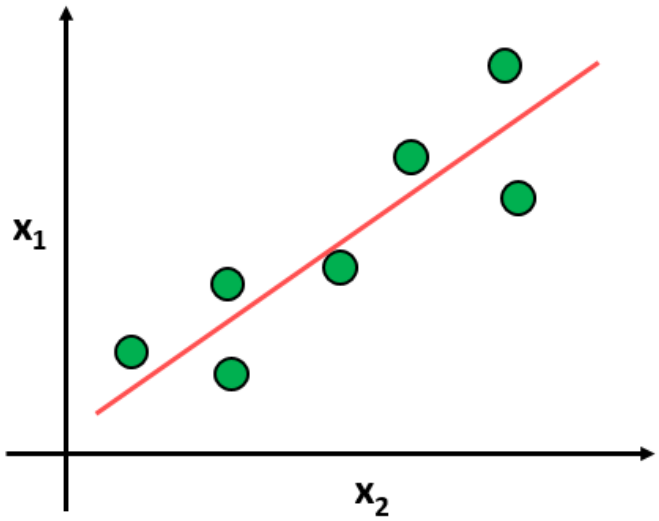
ML Basics

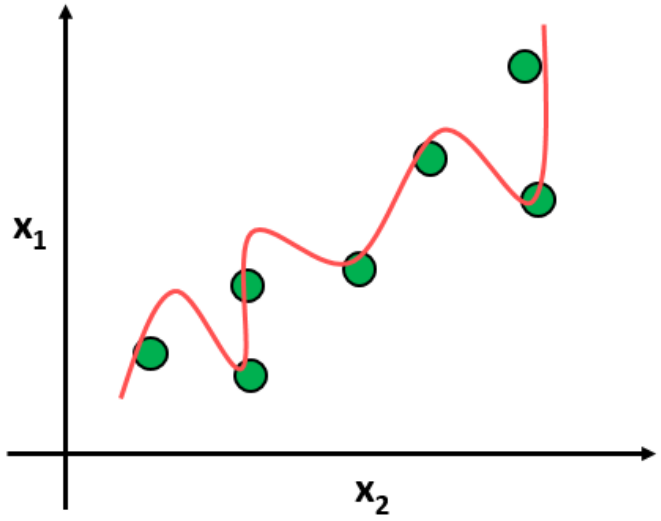


1. Introduction: Training and Test Data, ML models, losses
2. Theoretical foundations: probabilistic models, data representation
3. Regression and classification
4. When is a model good? Model complexity, bias complexity tradeoff/generalization (VC dimension, generalization error)

DIPARTIMENTO
DI INGEGNERIA
DELL'INFORMAZIONE

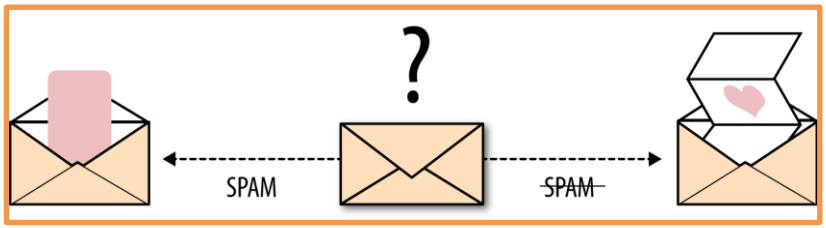
Validation and Model Selection

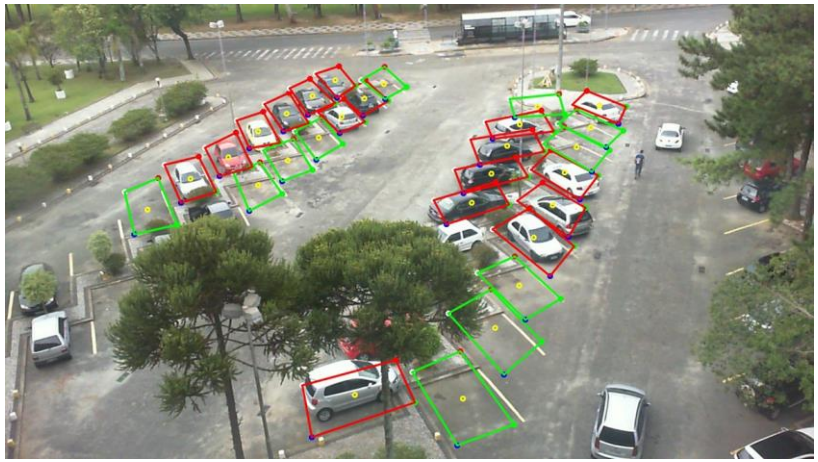


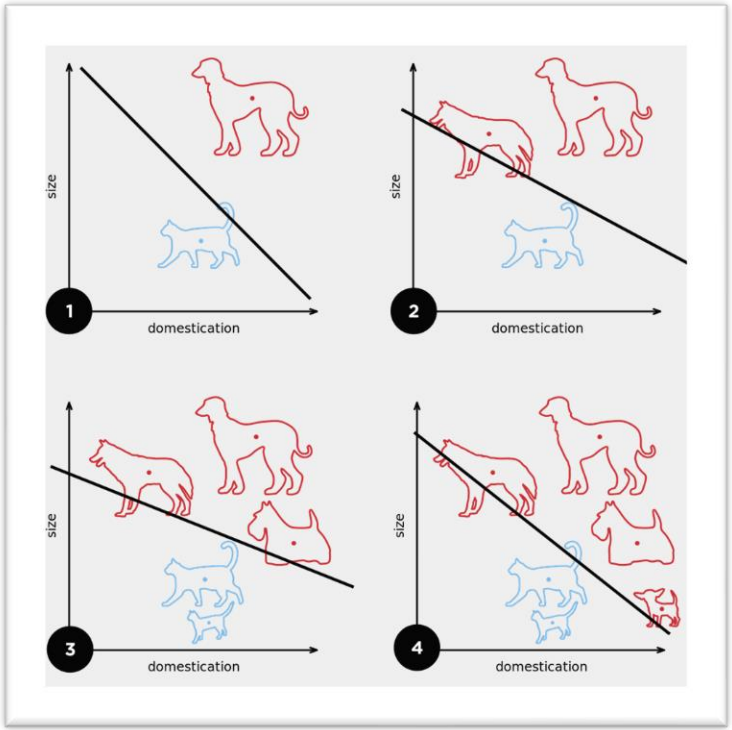


- Training and Generalization Error
- Bias-Complexity Trade-off
- Validation and model selection
- K-fold cross Validation
- Model complexity determination

Classification: Simple Strategies



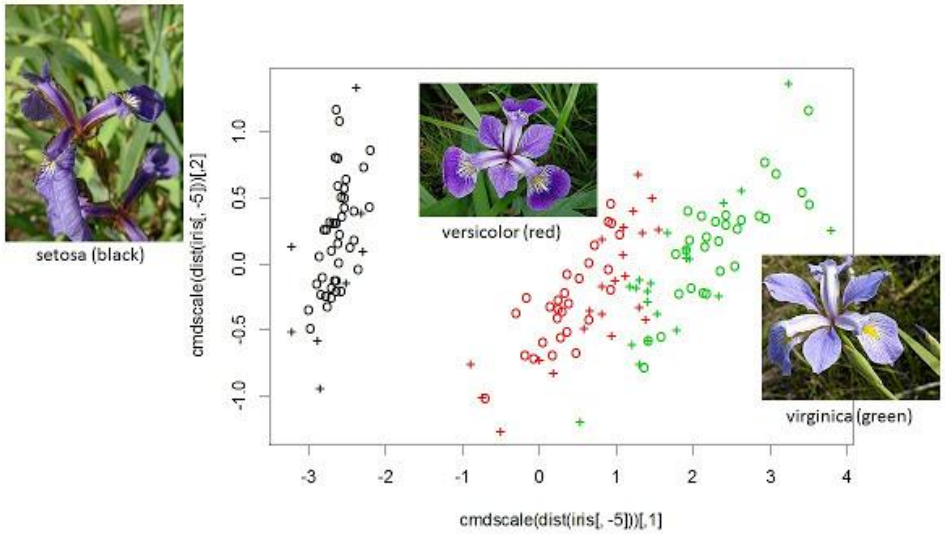


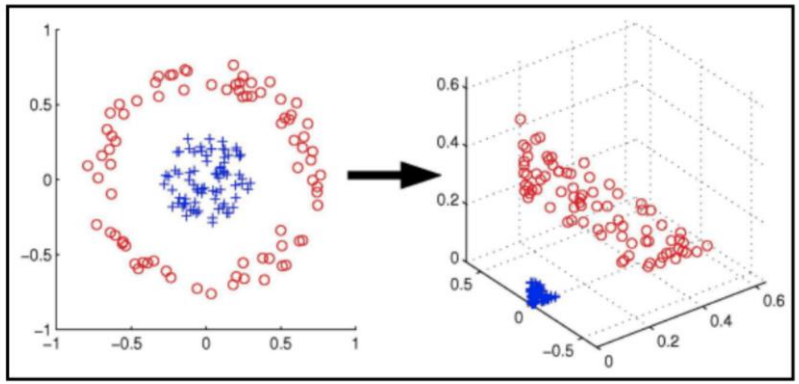


Start from simple classification algorithms:

- Linear classifiers
- Perceptron
- Logistic Regression

Classification: Support Vector Machines

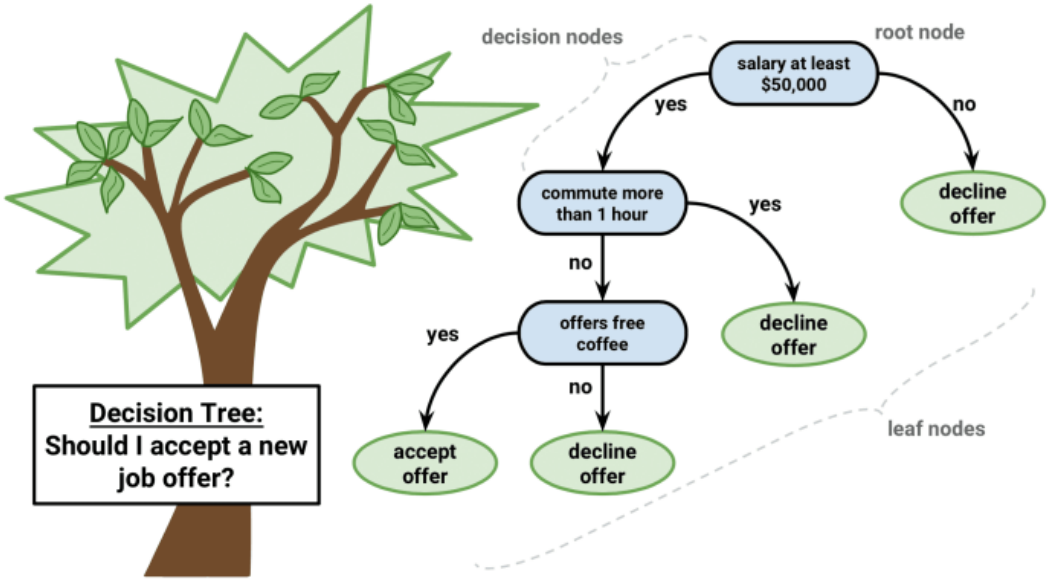


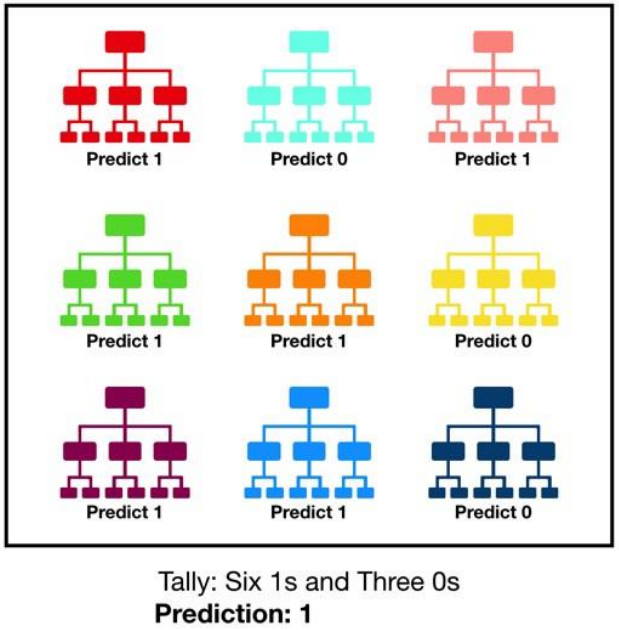


Support Vector Machines (SVM)

- Hard SVM (linearly separable data)
- Soft SVM (handle non-linearly separable data)
- The *kernel trick* (non-linear classification)
- Example of Applications

Classification: Random Forests

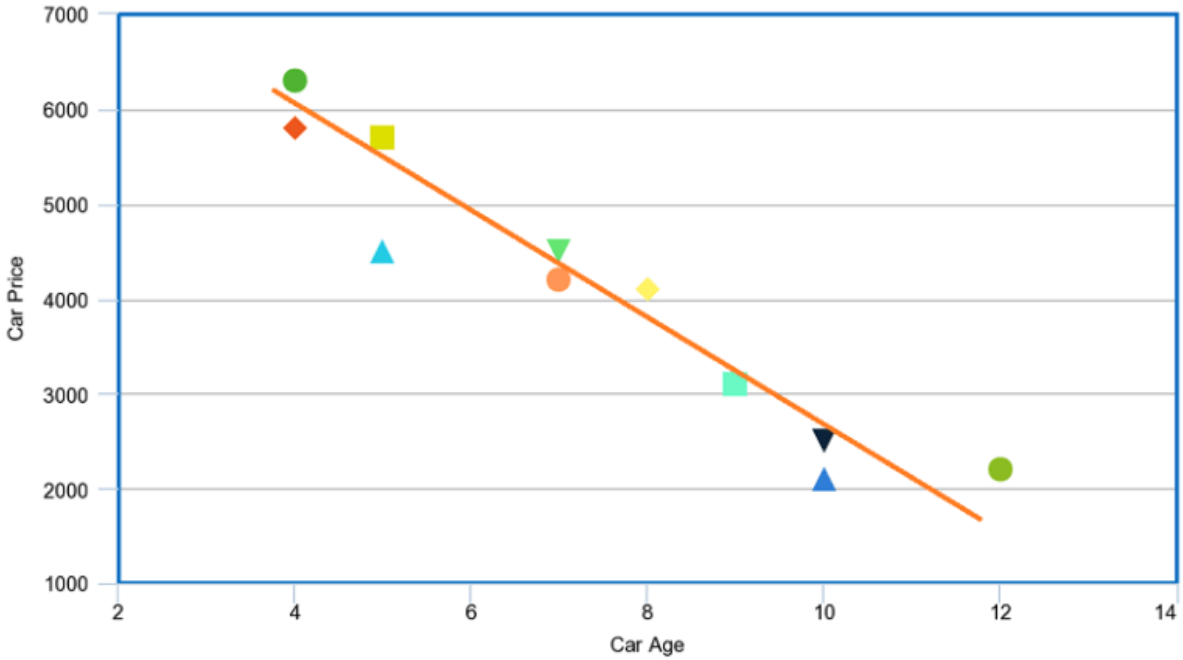




Random Forests (RF)

- Growing Decision Trees
- Classification with Random Forests
- Randomization techniques for RF

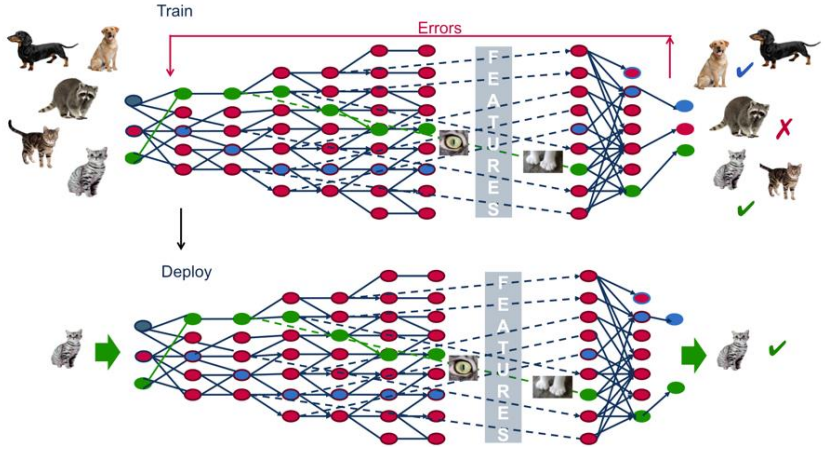
Regression

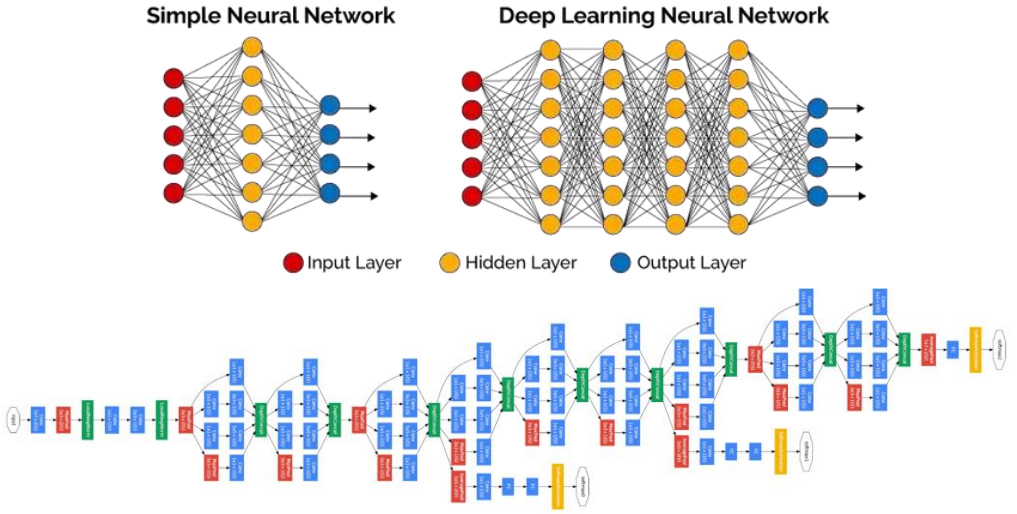


- Models for Regression
- Linear Regression (scalar and multivariate)
- Regularization techniques

DIPARTIMENTO
DI INGEGNERIA
DELL'INFORMAZIONE

Neural Networks and Deep Learning





Neural Networks (NN) and Deep Learning (DL)

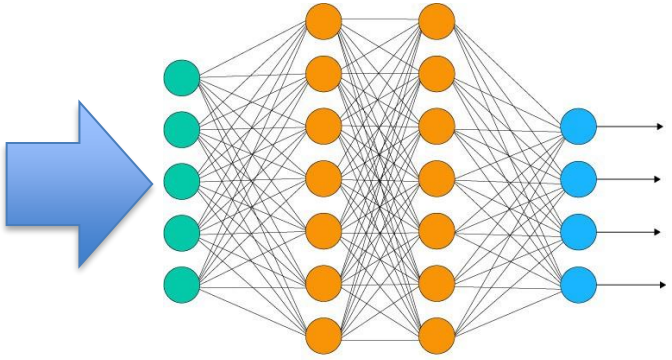
- Basic Neural Network model
- Multi-layer (deep) feedforward neural networks
- Convolutional Neural Networks (CNN)
- *Quick overview of advanced DL models*
- Examples of applications

Machine Learning (unsupervised)

Training data
(**without labels**)

Training procedure

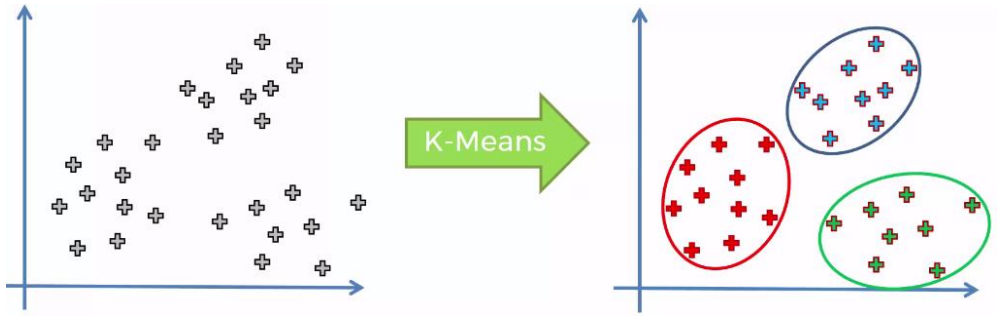
Data to be analyzed

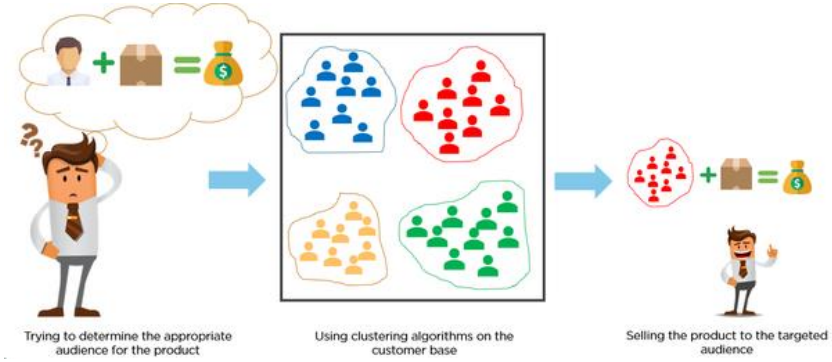


ML model
(estimate parameters)

Predicted Label

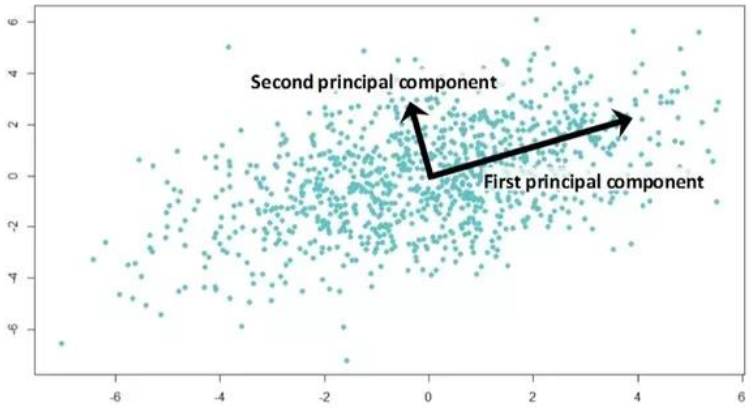
Clustering





- Basics of Clustering
- Linkage-based clustering
- K-means clustering
- *Lab on clustering*

Dimensionality Reduction



Principal Component Analysis (PCA)

Labs: Python Programming

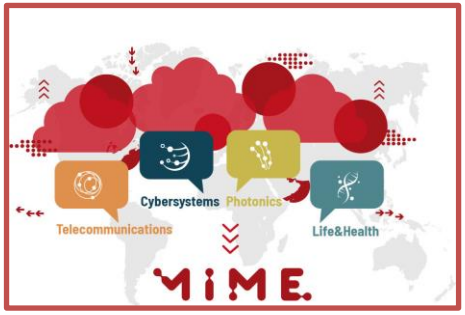
- We'll use Python + scikit learn
- Lab sessions in the computer rooms, but all the work can be done from home with a standard PC
- Install the software on your laptop
- Libraries: scikit-learn, numpy
- Jupyter notebook (mix code and text, avoid separate report)

DIPARTIMENTO
DI INGEGNERIA
DELL'INFORMAZIONE

Practical Info

Machine Learning (INP9087775/SCP8082660)

- This course is for *ICT for Internet and Multimedia* and *Physics of Data*
- The course is offered from the DEI department
- It is divided in two channels, this is channel "**B**"
  - Last digit of **ID number from 5 to 9**
  - If it is from 0 to 4 go to channel "A" with prof. Chiariotti
- IF, IAM, IBM have different instructors/channels
- ICT Life&Health: 12 CFU course together with DSP
- 6 CFU (48 hours, 24 lectures)
- Course in English



Elearning

- All the material and information on elearning
- Use elearning to get the links for online streaming and recorded lectures
- Subscribe as soon as possible on

  https://stem.elearning.unipd.it
- You can login with your *unipd* account
- Elearning will be used also for assignment delivery

DIPARTIMENTO
DI INGEGNERIA
DELL'INFORMAZIONE

Come to Classroom !

Lectures are in classroom only:

- Rooms Ae (Tuesday) and Me (*or Rn*) (Friday)
- Labs in rooms Te/Ue on Friday
- *Check the room for Friday on elearning, it changes week by week!!*

Recorded Lectures (emergency use):

The recordings of the lectures from 2021 will be placed on elearning, use them only if you can not come to the lectures due to critical issues. They are provided "as is"

Laboratories

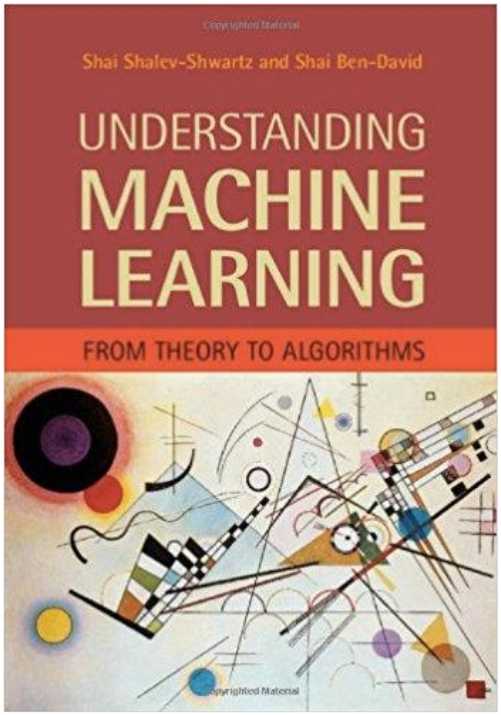
4+2 Labs:

1. Introduction to Python and scikit-learn
2. Regression and Classification (HW1)
3. Support Vector Machines (HW2)
4. Clustering (HW3)
5. Neural Networks (HW4)
6. Tutorial: Deep Learning (*optional*)

Books and Material

Main Book:

- Shalev-Shwartz, Shai; Ben-David, Shai, *Understanding machine learning: From theory to algorithms*, Cambridge University Press, 2014
- PDF available from the authors at http://www.cs.huji.ac.il/~shais/UnderstandingMachineLearning/copy.html



- Slides, tutorials, papers and other material on elearning
- *Come to the lectures and take notes !!!*

Homeworks

Homework	Released	Lab Session	Delivery
INTRO LAB		27/10	
1	31/10	10/11	16/11
2	17/11	24/11	27/11
3	28/11	1/12	18/12
4	19/12	22/12	10/1
DL TUTORIAL		19/1	

** Tentative dates, could change*

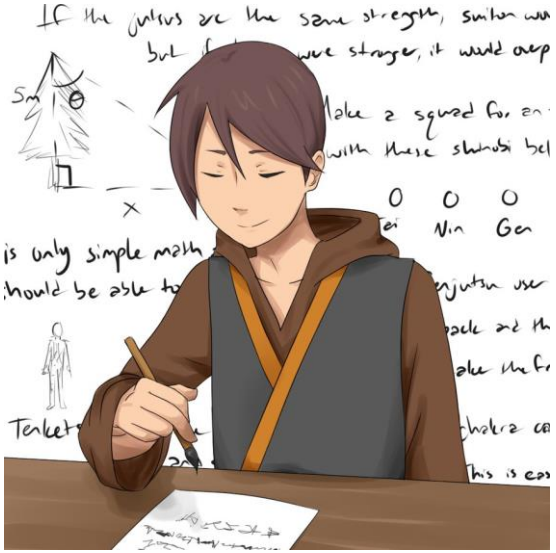
- 4+2 Labs, 4 Homeworks (first and last have no HW)
- Schedule for each homework:
  1. Homework is released
  2. Support session (in the lab)
  3. After ~2 weeks delivery deadline
- Up to 4 extra points for the homeworks (1pt for each homework)

Respect the deadlines!
No homework delivery in exam session

Written Exam

- Written exam in classroom at the end of the course
- No orals; ***No online exams***
- Final mark is the written exam score + the homework score
- Can pass without the homeworks but extra points help !
- Dates for the exams:

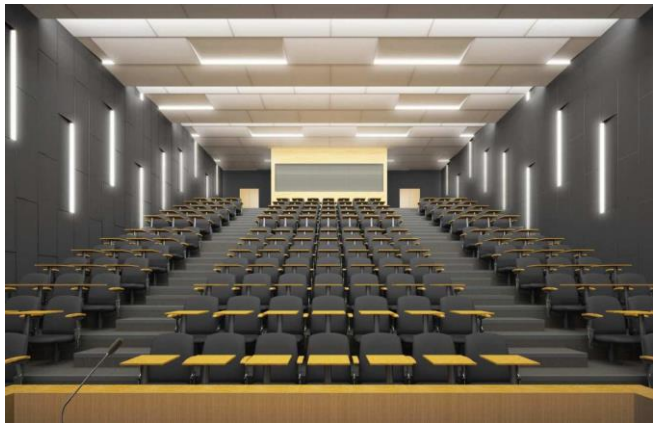
1. 30/1/2024 h 10.00
2. 19/2/2024 h 14.30
3. 28/6/2024 h 10.00
4. 12/7/2024 h 10.00
5. 10/9/2024 h10.00



Check the exam dates
No out-of-session exams

Exams will be in classroom only
No online exams

Lectures



- Tue 16:30 -18:00 Room Ae
- Fri 12:30 - 14:00 Room Me (or Rn,Te,Ue, check each week!)
- Labs on Friday in Rooms Te/Ue (for some labs also Da/Ge)
- Classroom attendance is recommended
- Use the last year recorded lectures only in case of issues
  - They are provided "as is" and could not correspond exactly to this year lectures (that however are very similar)

DIPARTIMENTO
DI INGEGNERIA
DELL'INFORMAZIONE

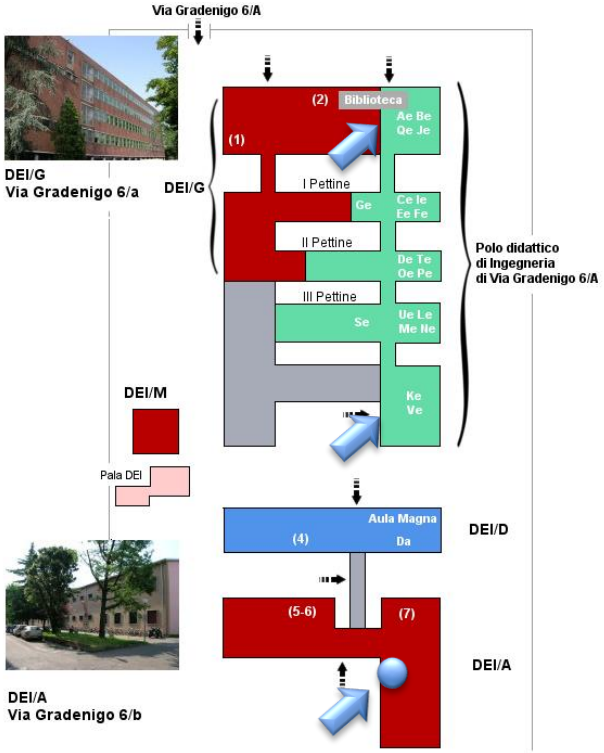
Contacts

Instructor:

- Dr. Pietro Zanuttigh
- e-mail: zanuttigh@dei.unipd.it
- Office: room 216 DEI/A
- Office phone: 049 827 7782

Teaching Assistants

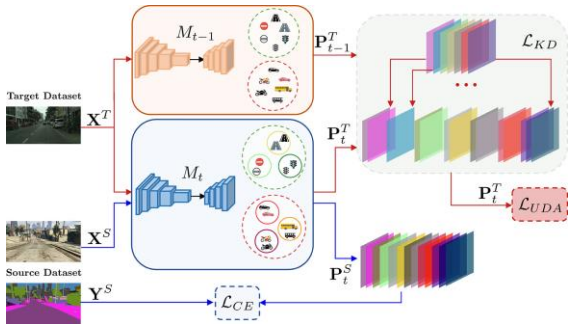
- Matteo Caligiuri
- Federico Lincetto
- Giulia Rizzoli
- Donald Shenaj



DIPARTIMENTO
DI INGEGNERIA
DELL'INFORMAZIONE

Thesis in the LTTM Lab

Machine learning for image and 3D Data Processing and Understanding



- Semantic segmentation from color and and depth data with deep learning
- Domain adaptation techniques for semantic segmentation
- Continual/Incremental learning
- Federated Learning
- Depth cameras data refinement and processing
- 3D reconstruction with Neural Radiance Fields (NeRFs)
- Stages in the image and video processing field are also possible
- Most problems: solved with ML !!

SONY sisma salvagnini SAMSUNG Addfor+ MICROTEC Altair REPLY 4EASYTECH NIDEK

Labs: Setup your PC

- We'll use *Python* with the `scikit learn` library
- It is strongly suggested to ensure that you are able to develop and run the assignments on your PC
- Simple tasks, any "standard" PC should be sufficient
- However lab's PC will also be available

DIPARTIMENTO
DI INGEGNERIA
DELL'INFORMAZIONE

Setup: Install on your home PC or laptop

Suggested approach (but others exist):

- ❑ Install Anaconda (with Python 3)
- ❑ Install scikit-learn (if not already installed by Anaconda)
  - Install scikit-learn with anaconda: `conda install scikit-learn`
  - or install with pip: `pip install -U scikit-learn`
  - It requires: Python (>= 3.4), NumPy (>= 1.8.2), SciPy (>= 0.13.3)
  - If required install the dependencies with pip or conda
- ❑ Install Jupyter notebook
  - ❑ With anaconda it is installed by default
  - ❑ Can be launched with : `jupyter notebook` or `jupyter lab`

Try a Tutorial to Learn the Basics of Python

Useful resources to learn the basics of Python programming:

- Look at http://cs231n.github.io/python-numpy-tutorial/
- You can find a Jupyter notebook version of the tutorial at:
  https://github.com/kuleshov/cs228-material/blob/master/tutorials/python/cs228-python-tutorial.ipynb
- A huge amount of material can be found on the web

How to use: Python

1. Launch with the python command from the bash/command prompt

```
[python36] C:\Users\root>python
Python 3.6.2 |Anaconda custom (64-bit)| (default, Jul 20 2017, 12:30:02)  [MSC v.1900 64 bit (AMD64)] on win32
Type "help", "copyright", "credits" or "license" for more information.  >>>
```

2. Write your source code and save in a .py file

Run the file: `python filename.py`

3. *Run with : `jupyter notebook` or `jupyter lab`*
   For the labs we'll use jupyter

DIPARTIMENTO
DI INGEGNERIA
DELL'INFORMAZIONE

How to use: Jupyter notebook / lab

- Run with : `jupyter notebook` or `jupyter lab`
  - Jupyter lab has some extra features
- Interactive environment inside the web browser
- You can run each block of code and see the output
- Can combine code and text (add comments / description)
- We'll use jupyter notebooks for the lab deliveries
- *Do not alter templates or data files (including names and locations)*