DIPARTIMENTO
DI INGEGNERIA
DELL'INFORMAZIONE

UNIVERSITÀ
DEGLI STUDI
DI PADOVA

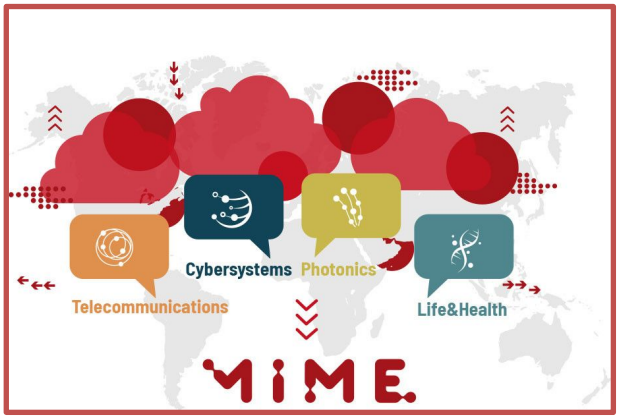
Machine Learning

Course Introduction 2022-23

Physics of Data and ICT for Internet and Multimedia
Federico Chiariotti (chiariot@dei.unipd.it)

DIPARTIMENTO
DI INGEGNERIA
DELL'INFORMAZIONE

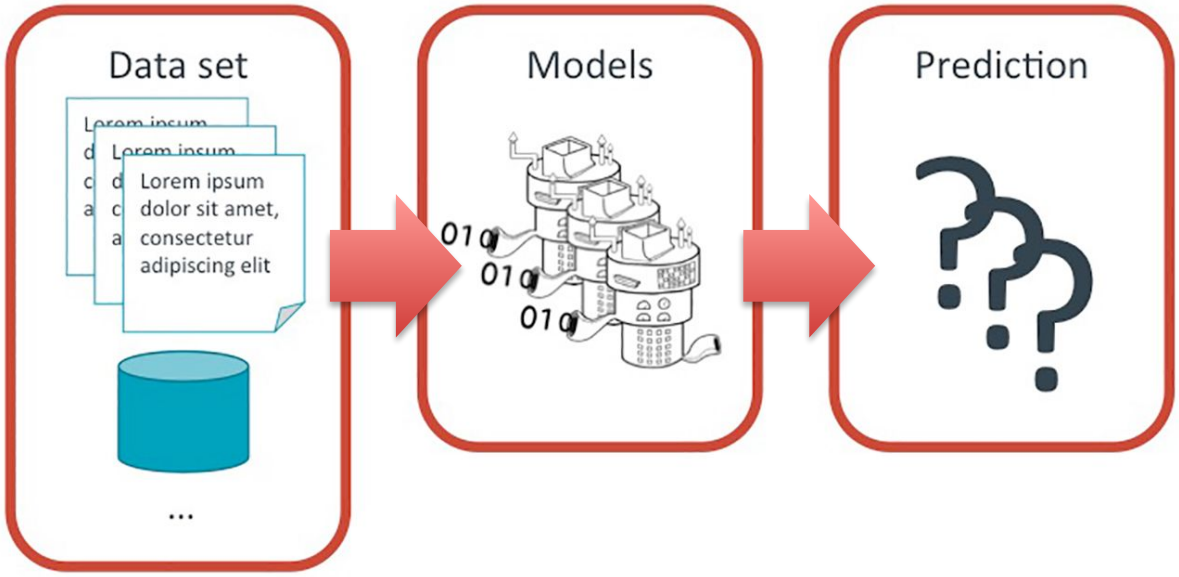
Machine Learning Course



Machine Learning (INP9087775/SCP8082660)

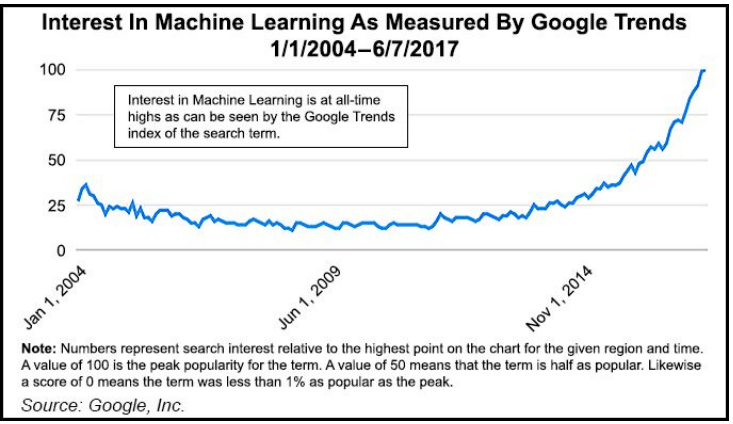
- This course is for *ICT for Internet and Multimedia* and *Physics of Data*
- The course is officially offered from the Physics department (even if lecture rooms and instructor from DEI)
- IF, IAM, IBM have different instructors/channels
- If you are from other physics/math courses notify the instructor
- 6 CFU (48 hours, 24 lectures) - in English
- This class is for student numbers that end in 0-4

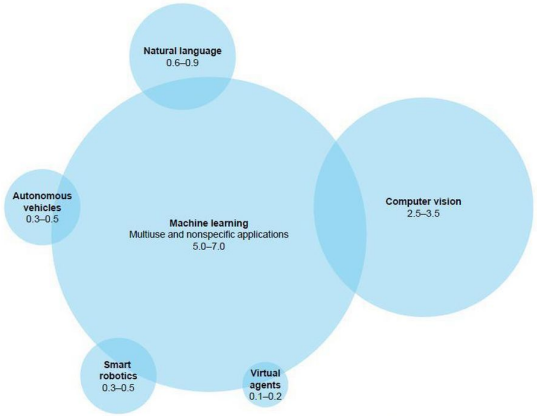
What is Machine Learning ?

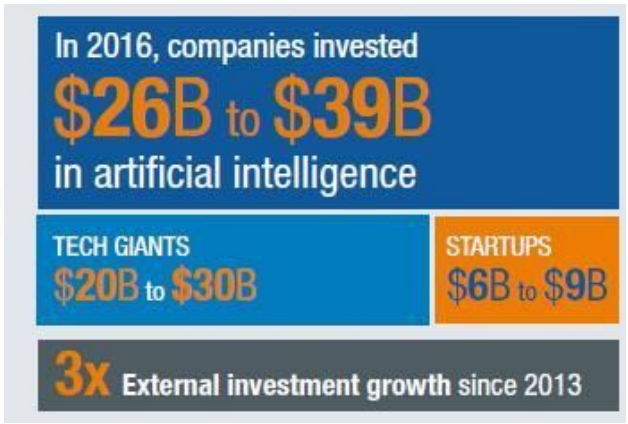


Machine learning (ML) is a set of methods that give computer systems the ability to "*learn*" from (*training*) data to make predictions about novel data samples, *without being explicitly programmed*

Why should you learn about it?

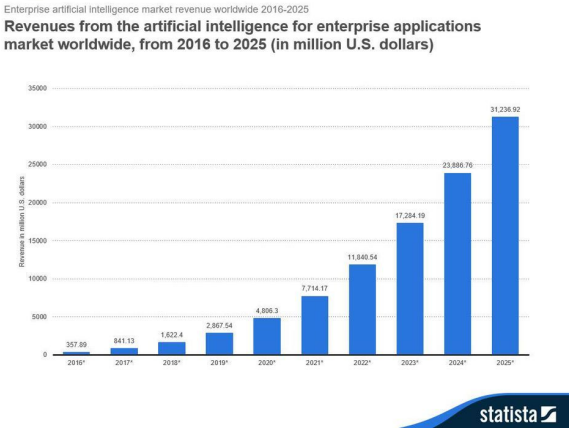


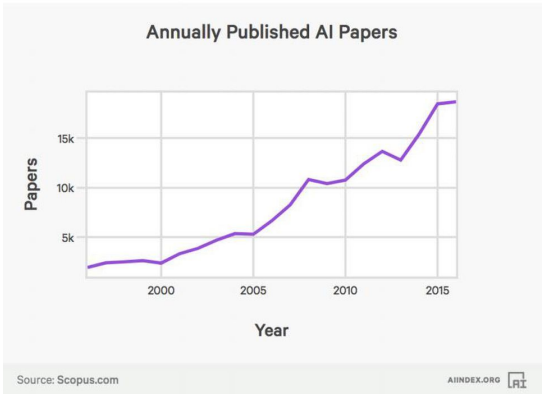


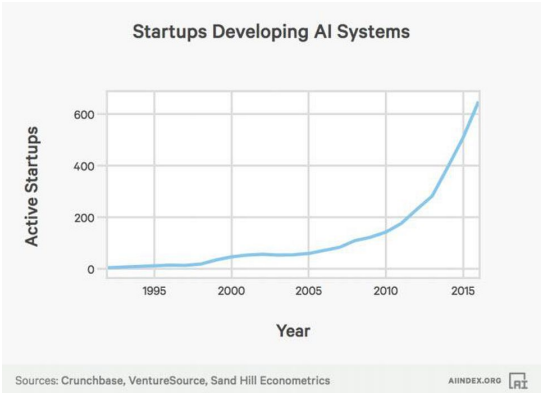


- Many applications of ML with outstanding results
- Huge investments from all around the world
- Availability of large amount of data and computational resources
- The global machine learning market was valued at **USD 2.40 Billion in 2019** and is projected to **reach USD 47.29 Billion by 2027**, growing at a CAGR 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

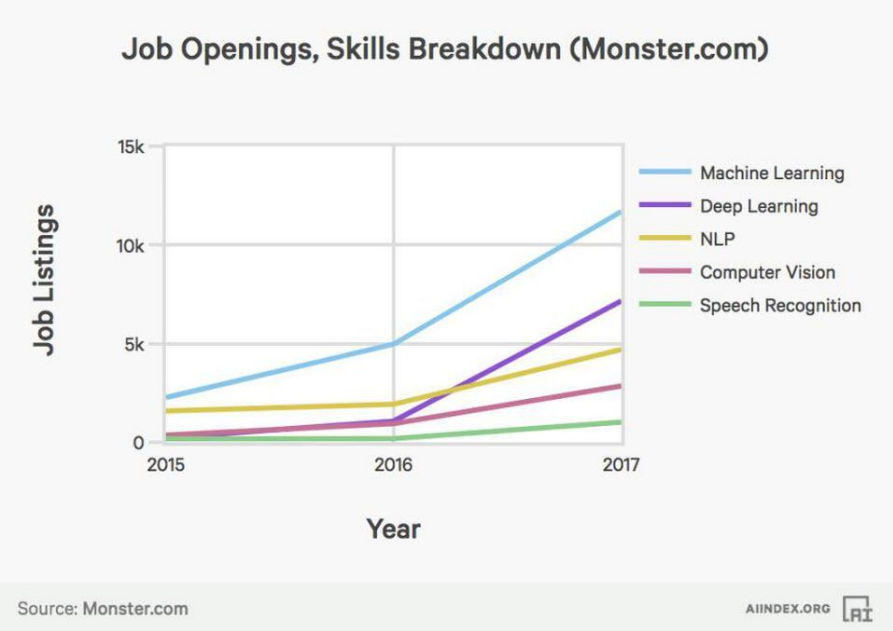


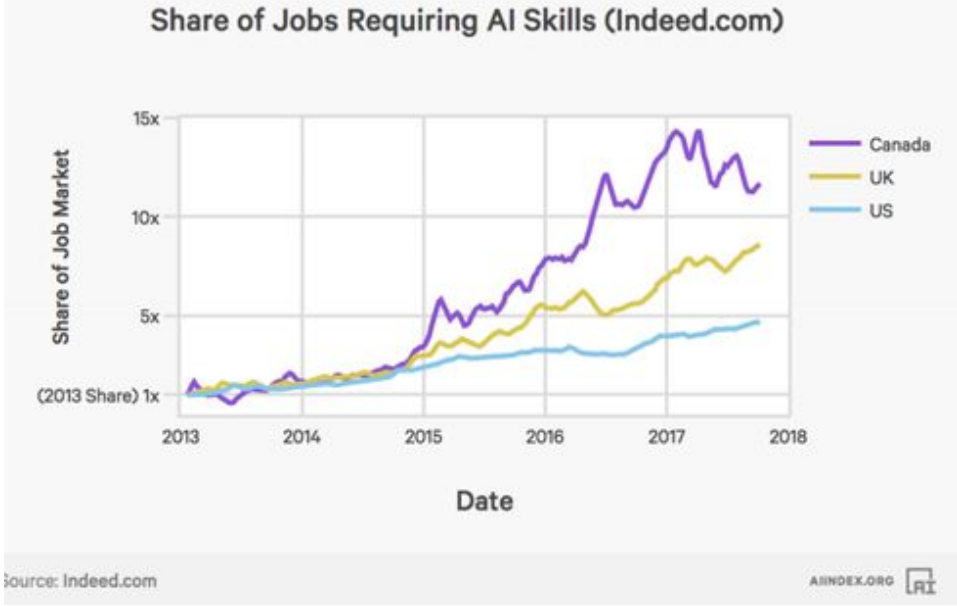




- The research on ML techniques is growing very fast both in the academia and in R&D departments of companies
- Both big companies and many fast growing start-ups involved in ML
- Larger and larger market for ML applications

Many Jobs in the Field



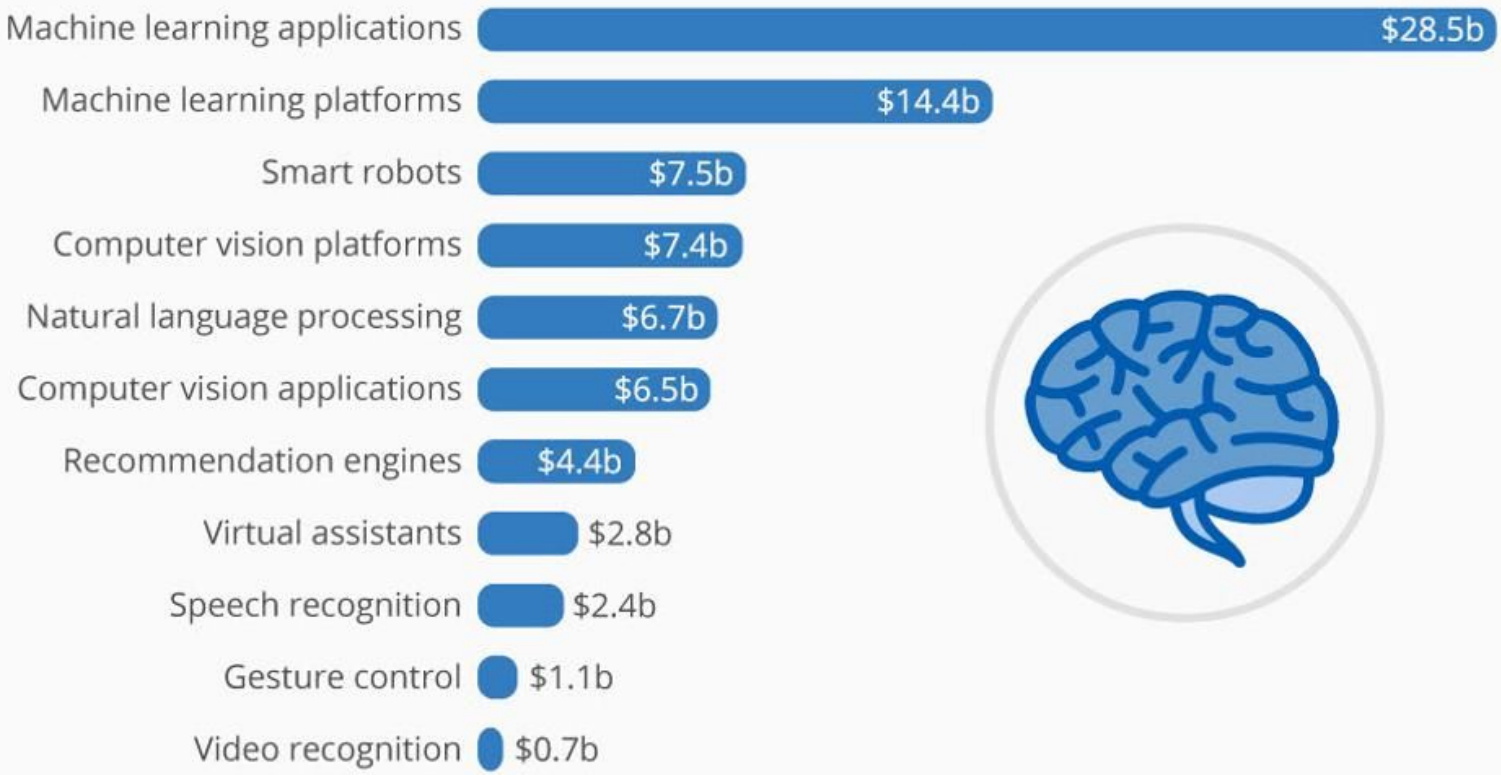


- Many different job opportunities both in Italy and abroad
- Companies struggle in finding ML experts
- According to a recent study machine learning engineering is the best paid job in the United States with an average salary of $146,085

Investments in ML

Machine Learning Tops AI Dollars

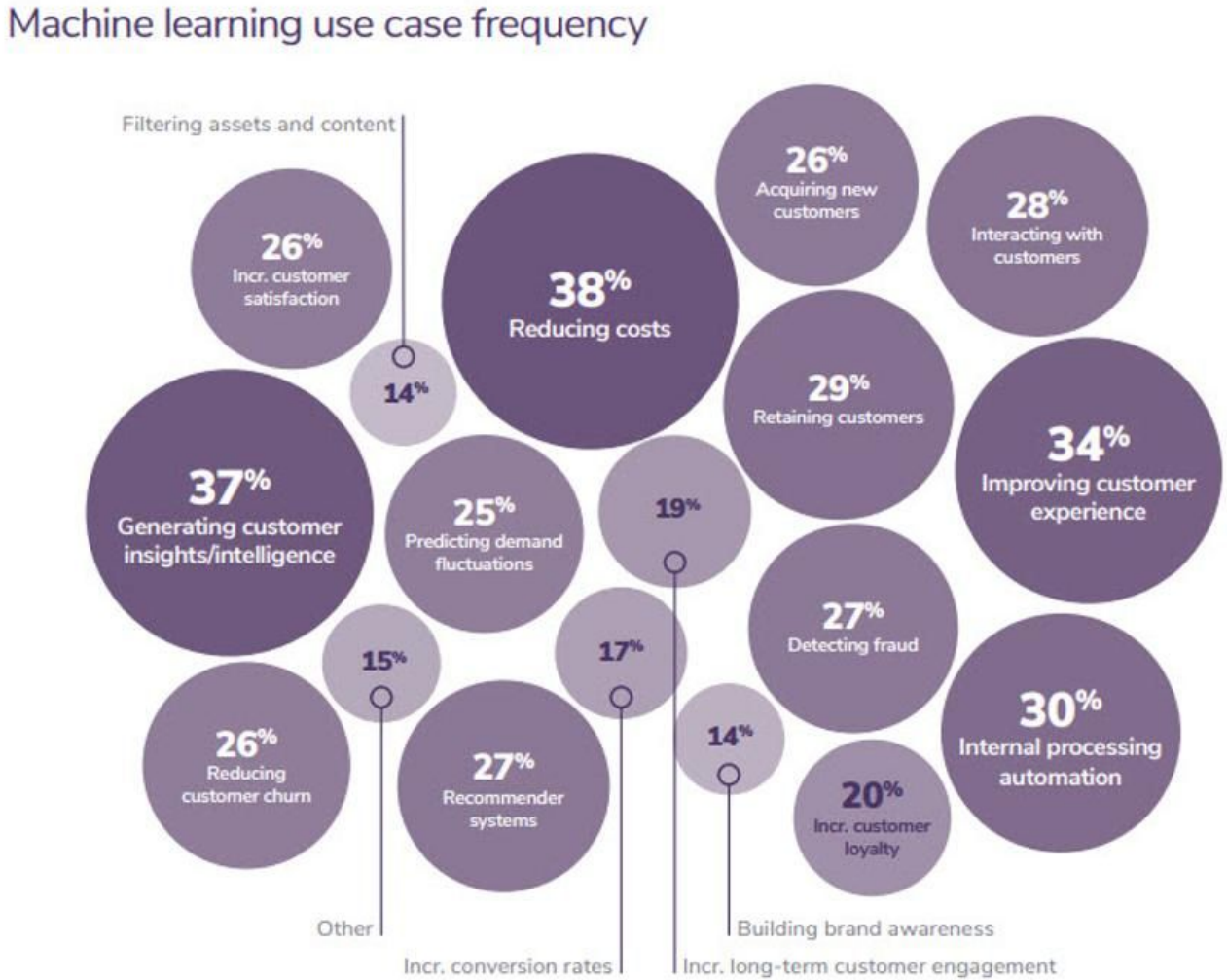
AI funding worldwide cumulative through March 2019 (in billion U.S. dollars), by category



Sources: Venture Scanner, Statista estimates

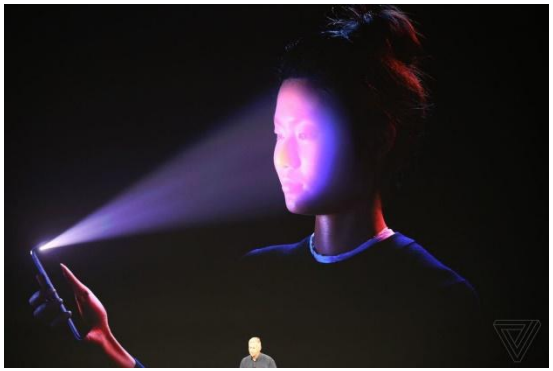
statista

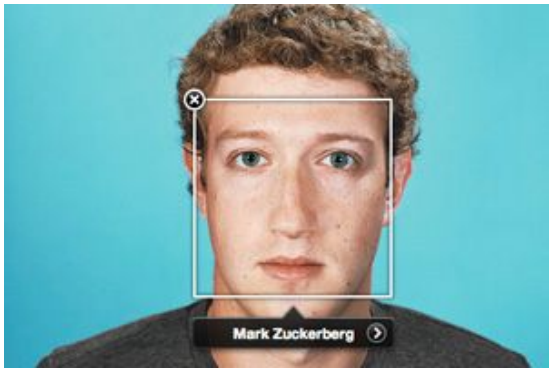
Where is ML used ?



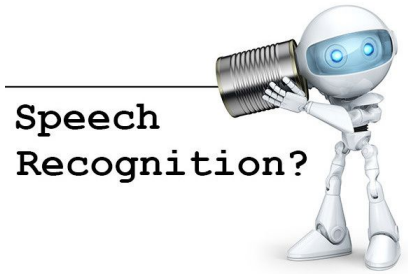
DIPARTIMENTO
DI INGEGNERIA
DELL'INFORMAZIONE

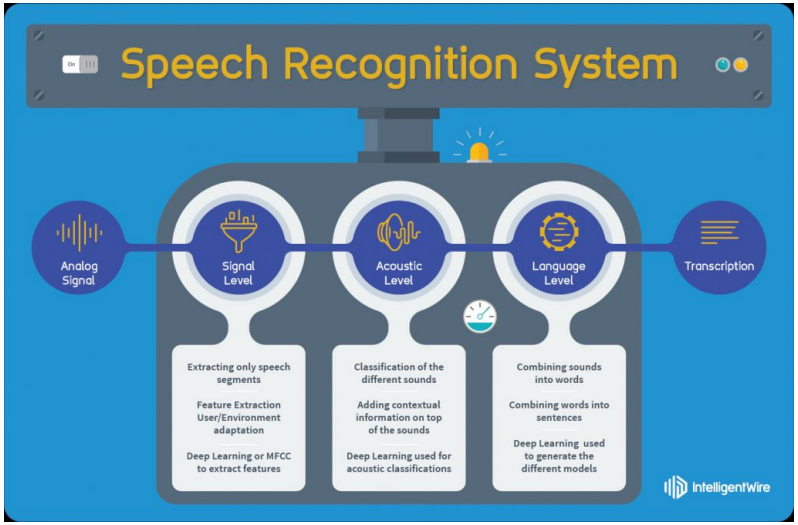
Many Applications: Pattern Recognition

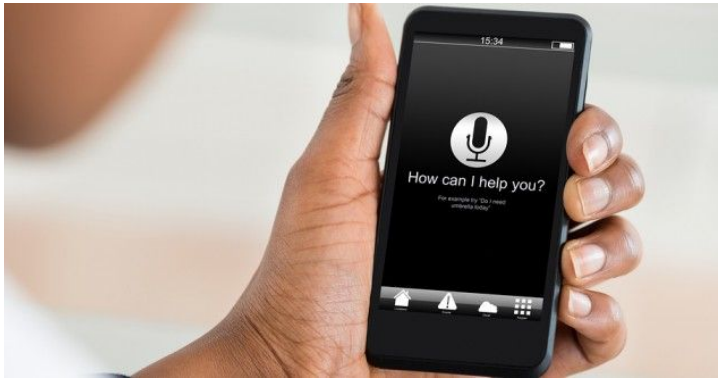




Many Applications: Speech Recognition







Google Home

amazon alexa

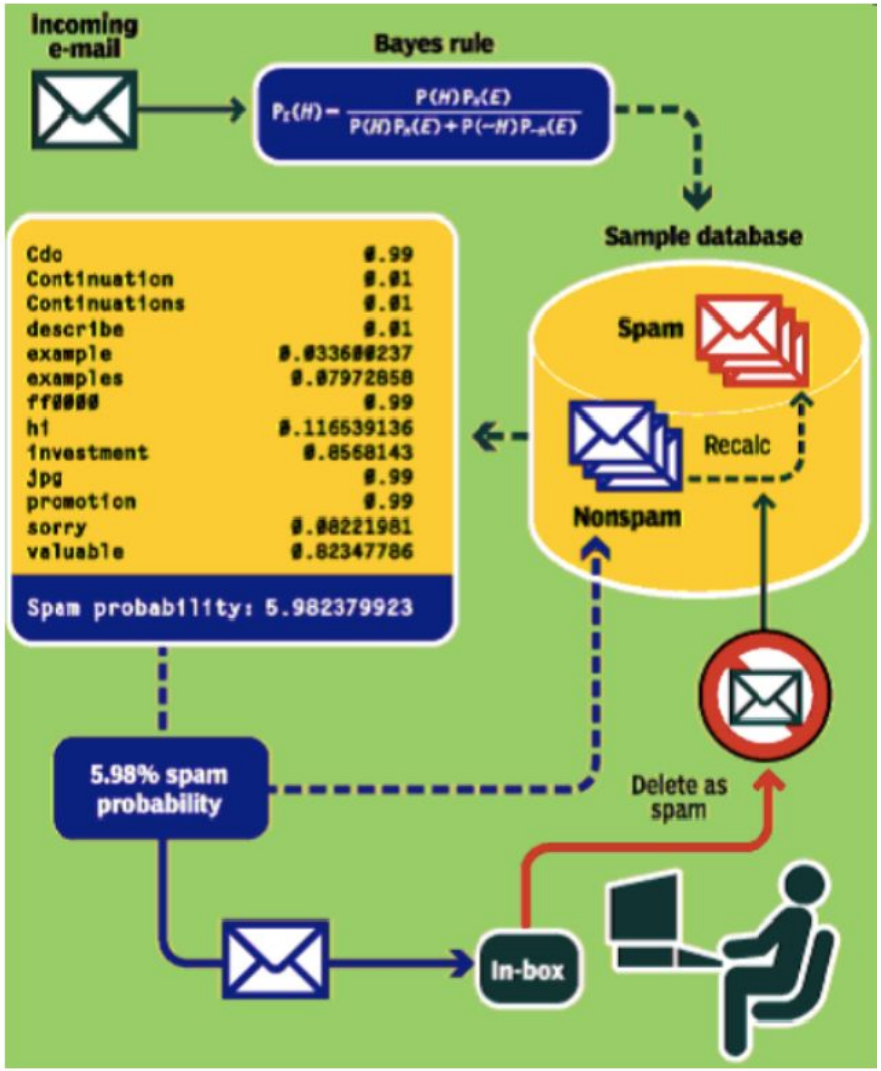
Many Applications: Spam Filtering

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

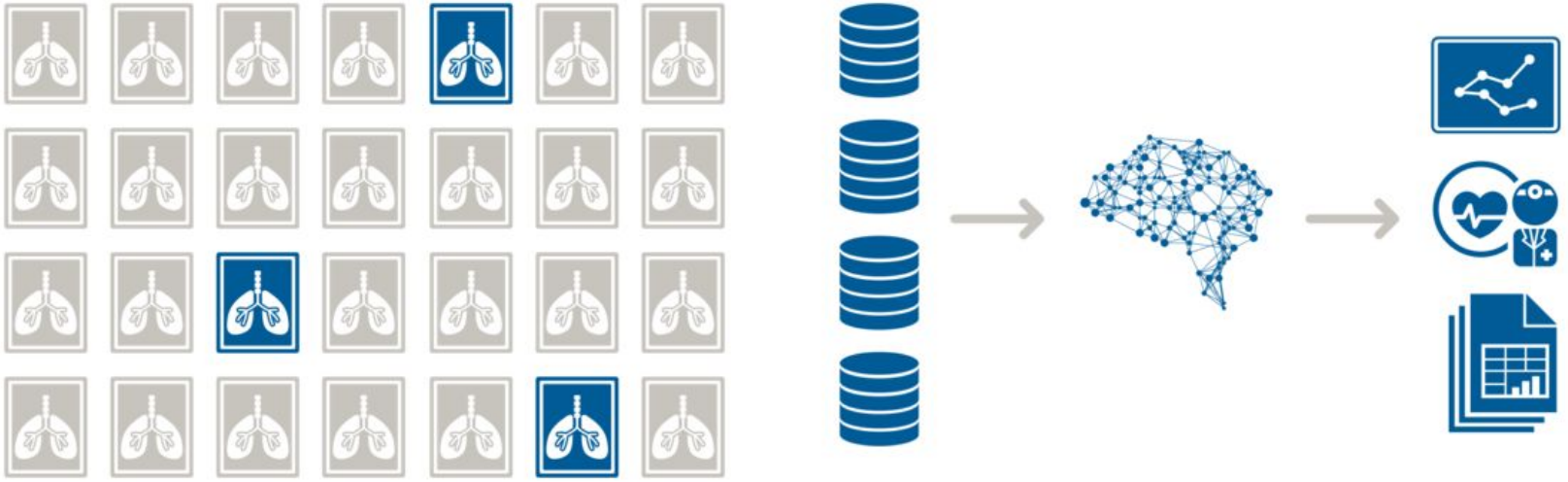
Challenge: There is no simple universal rule to define spam

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

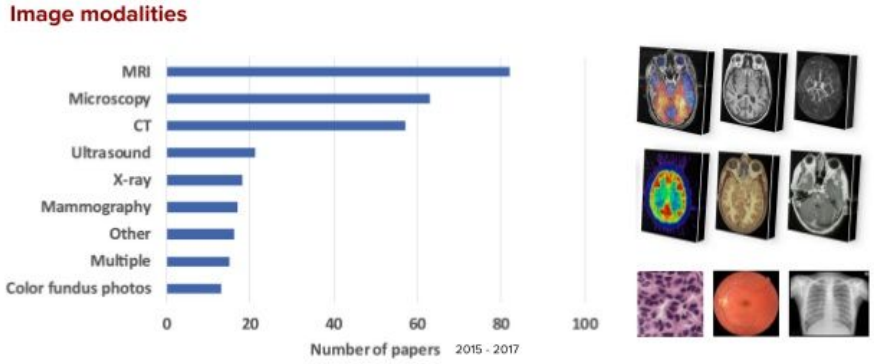
Challenge: Spammers evolve to counter filter innovations



Many Applications: Medical Imaging



Why deep learning for medical imaging?



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

Many Applications: Search Engines

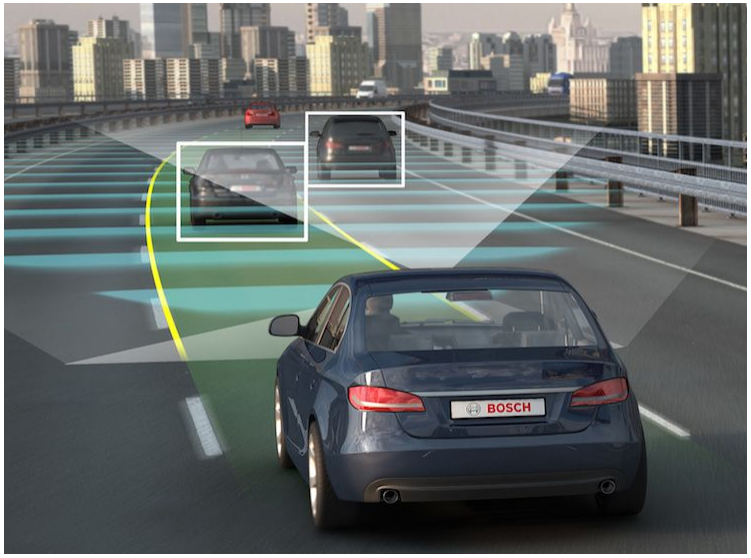
Google Goggles in Action

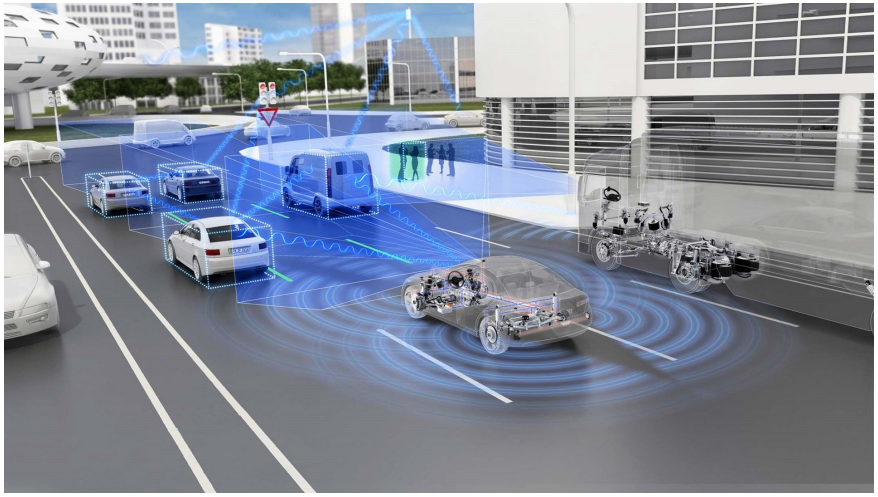
Click the icons below to see the different ways Google Goggles can be used.

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

Many Applications: Autonomous Driving





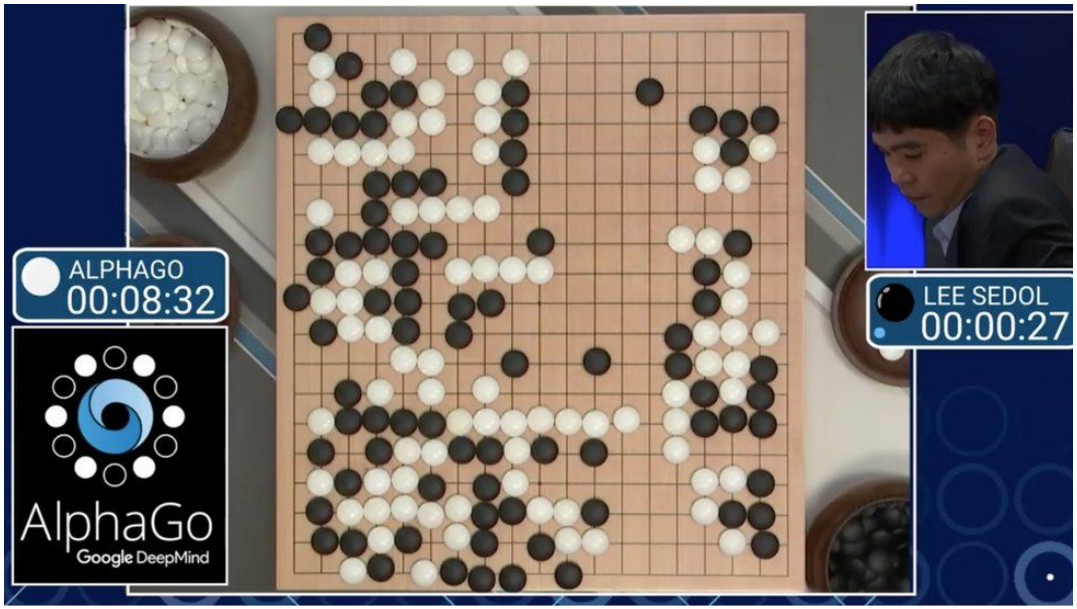
Source: A. Shashua, S. Seitz

Many Applications: Generating Patterns

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.

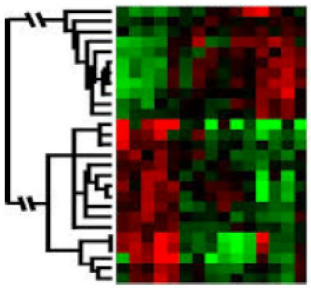
Reinforcement Learning

AlphaGo, a program developed by DeepMind (acquired by Google in 2015 for 500 M$), beat the Go world champion in 2017 (and its improved version subsequently beat the best chess and shogi programs)



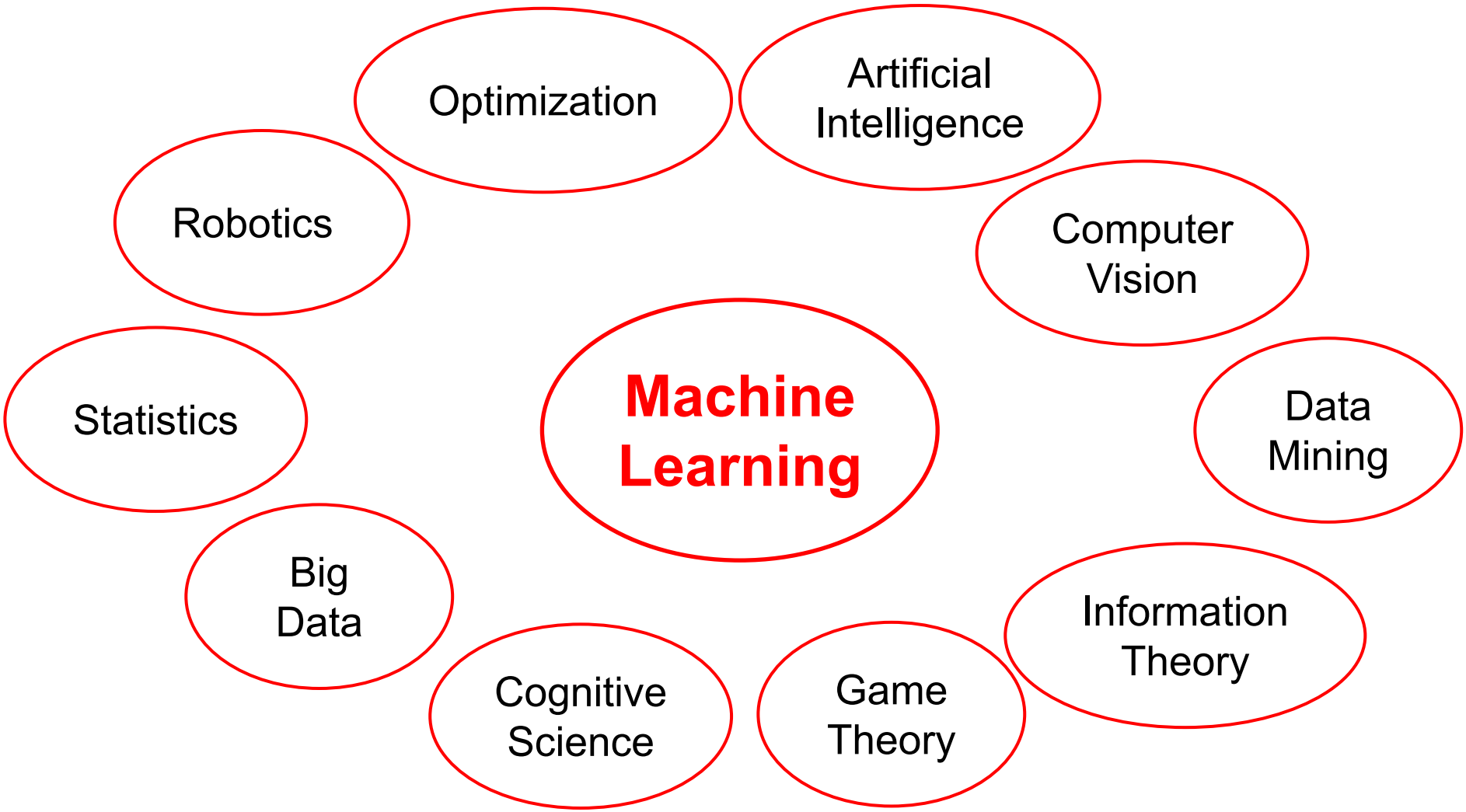
When is ML useful?



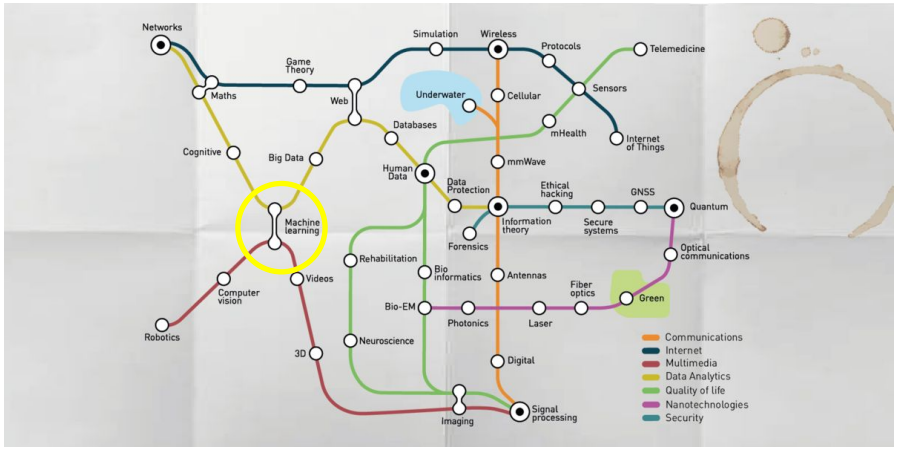

- Humans can't explain their expertise (speech recognition, image understanding)
- Tasks that humans can't solve, e.g., models are based on huge amounts of data (genomics, social media analysis)
- Models that must be customized or adapted (handwriting recognition, personalized medicine, spam filtering)
- 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)



	CFU	Semester
Digital Signal Processing	6	1
Machine Learning	**6**	**1**
Neural Networks and Deep Learning	6	1
Computer Vision	6	2
Reinforcement Learning	6	1
Human Data Analytics	6	1
Digital Forensics	6	2

… and ML techniques are used in many other courses….

Elearning

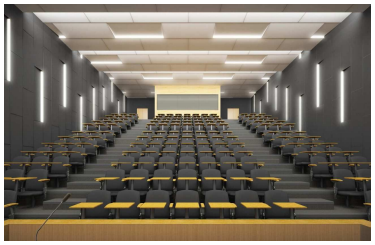
- All the material and information on elearning
- Use elearning to get the links for online streaming and recorded lectures
- Subscribe as soon as possible on

  https://stem.elearning.unipd.it
- You can login with your *unipd* account (please check that the course is correct and not the parallel one with the same name)
- Elearning will be used also for assignment delivery

DIPARTIMENTO
DI INGEGNERIA
DELL'INFORMAZIONE

Recorded Lectures

This year, the university is pushing in-person learning, so there will not generally be a Zoom link for following online. However, you still have two ways to get the material:

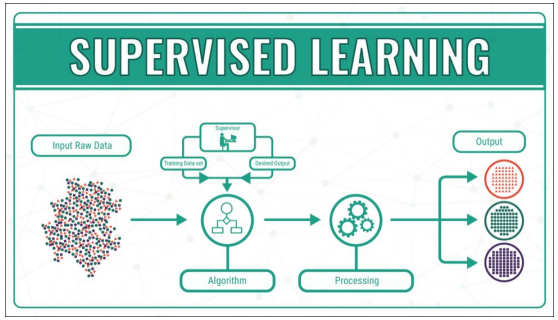
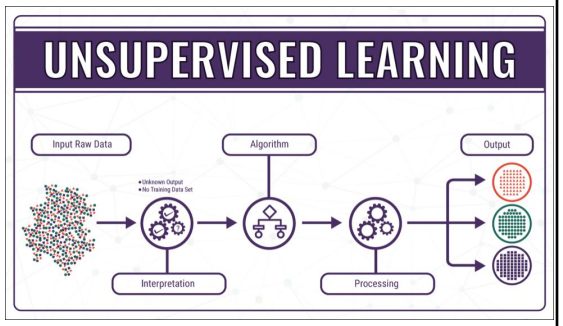


1. ***In classroom***: room Ae (Wednesdays and Fridays, 16.15)
2. ***Recorded Lectures***: The recordings of last year's lectures will be placed on elearning as a backup

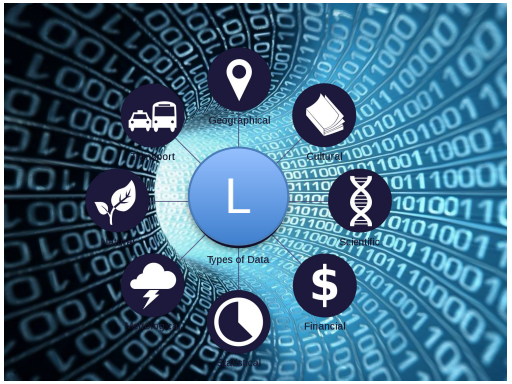
Given the COVID-19 situation, a switch to dual or fully online teaching could happen at any time for safety reasons:

remember to check your email for updates

Course Contents

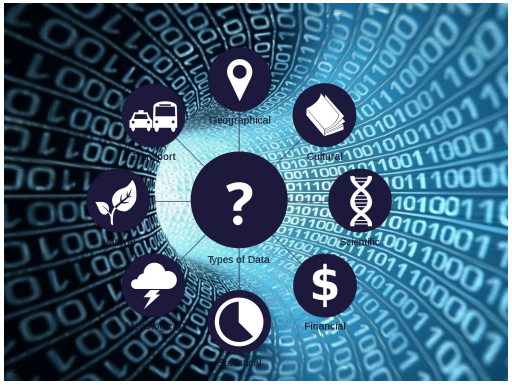
		
Supervised Learning	Unsupervised Learning	Laboratories

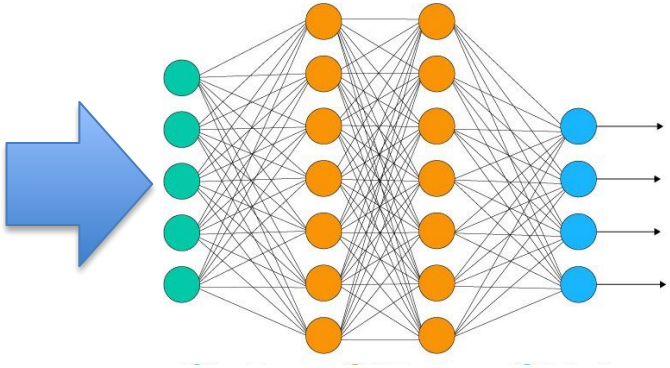
Supervised Learning



Training data

Training procedure



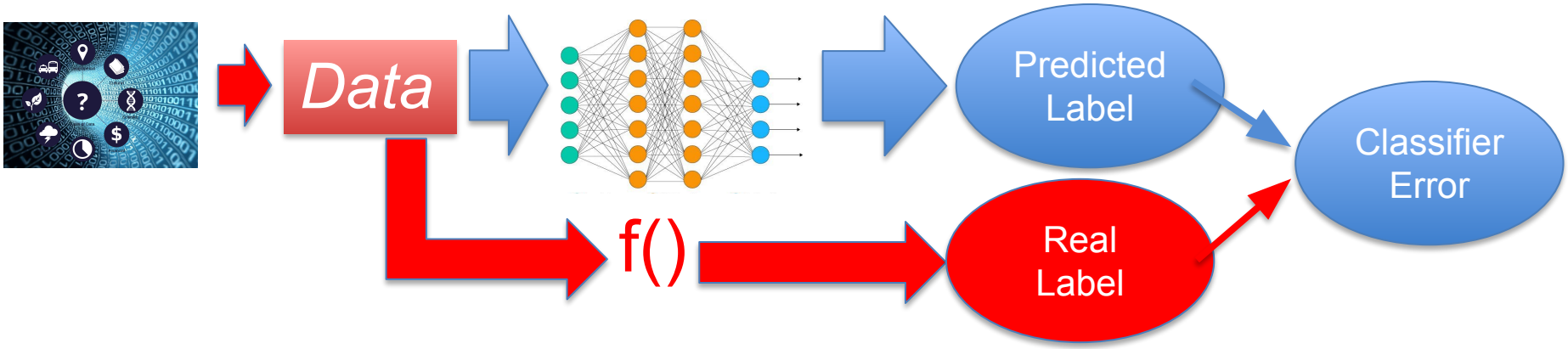


Predicted Label

Data to be analyzed

ML model (estimate parameters)

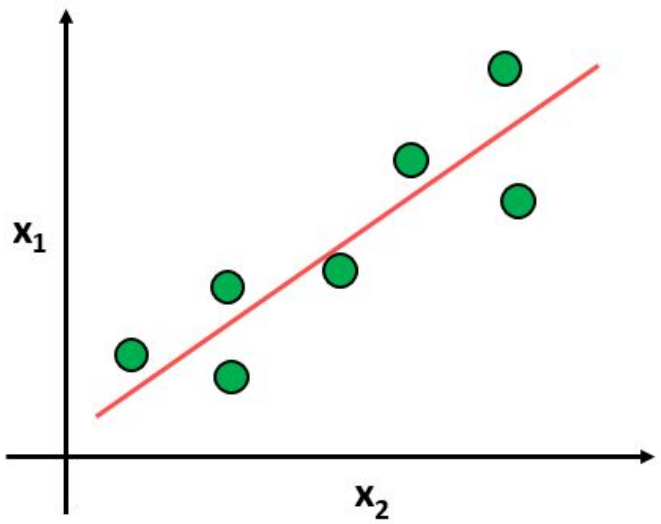
ML Basics

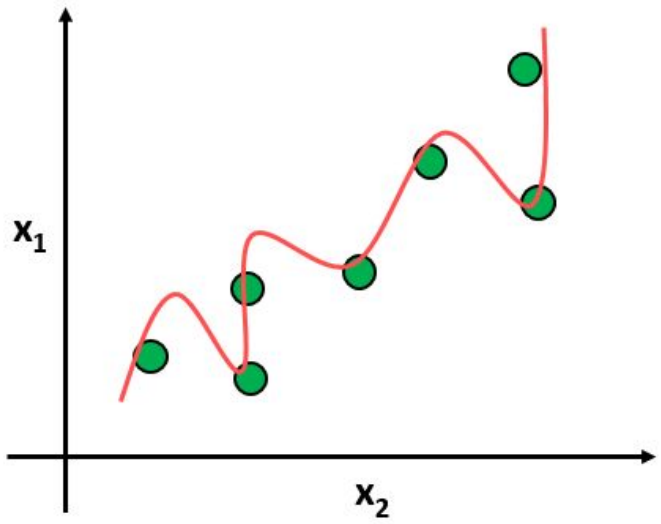


1. Introduction: Training and Test Data, ML models, losses
2. Theoretical foundations: probabilistic models and data representation
3. Regression and classification
4. When is a model good? Model complexity, bias complexity tradeoff/generalization (VC dimension, generalization error)

DIPARTIMENTO
DI INGEGNERIA
DELL'INFORMAZIONE

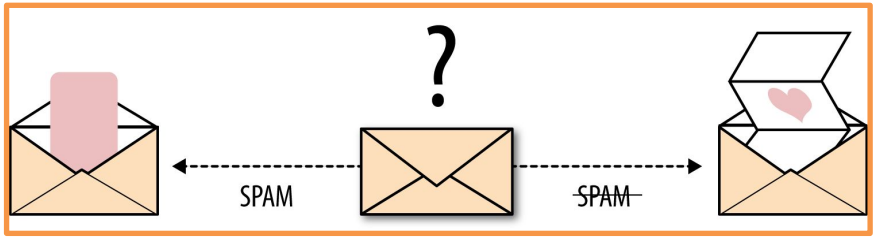
Validation and Model Selection

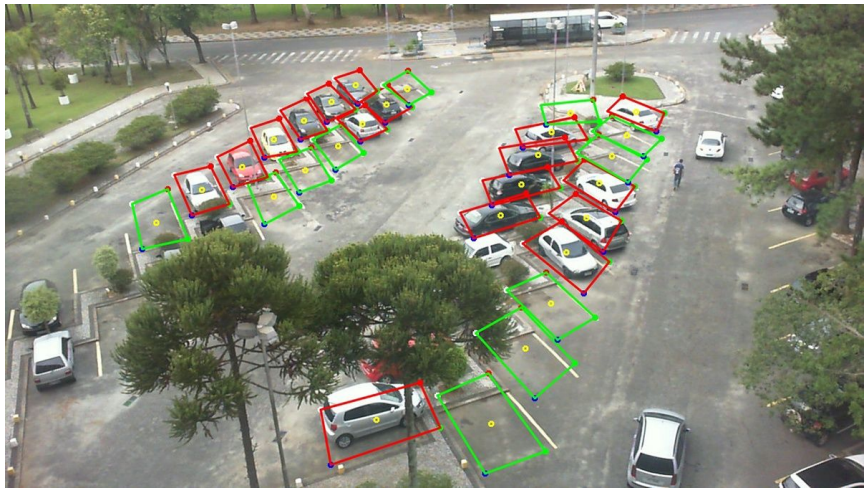


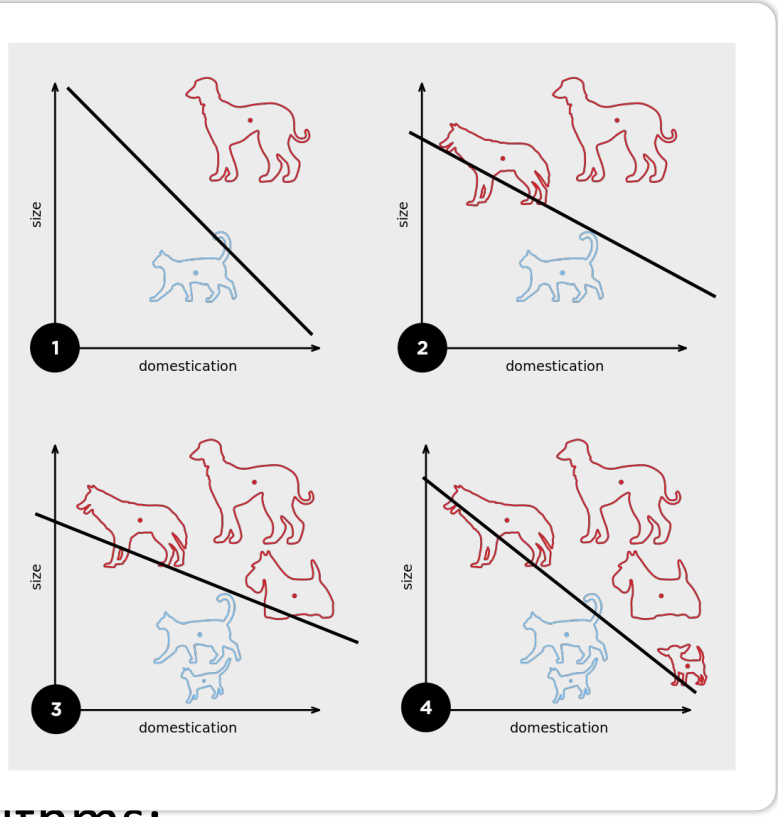


- Training and Generalization Error
- Bias-Complexity Trade-off
- Validation and model selection
- K-fold cross Validation
- Model complexity determination

Classification: Simple Strategies



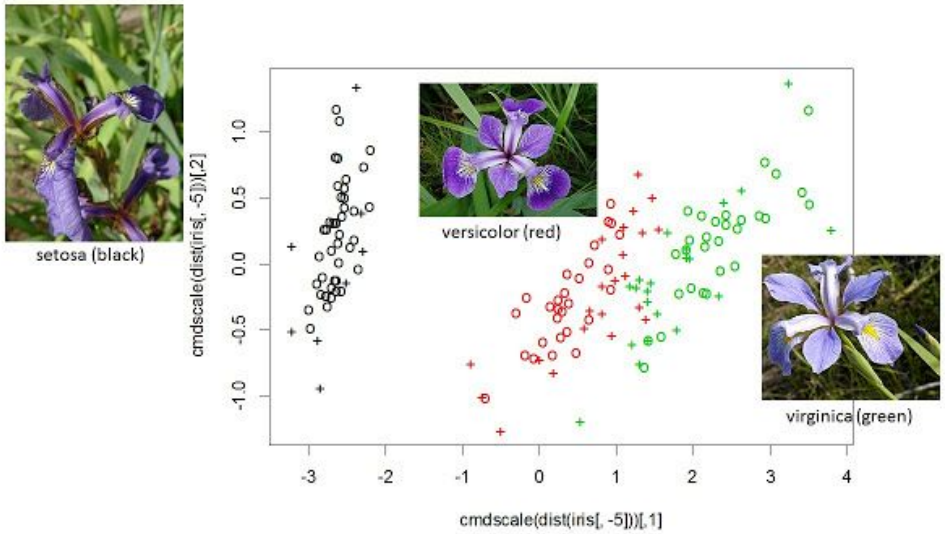


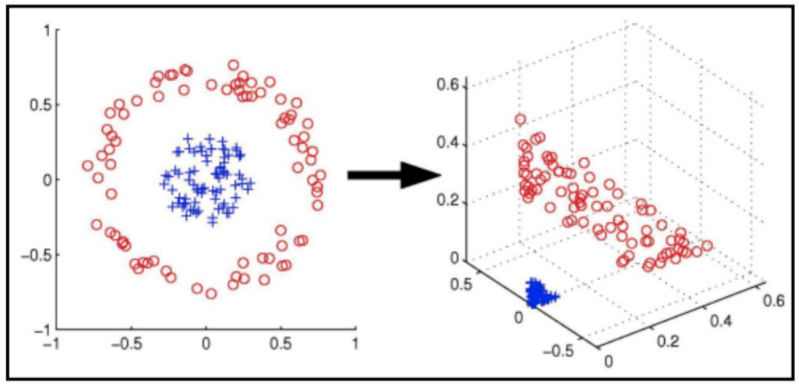


Start from simple classification algorithms:

- Linear classifiers
- Perceptron
- Logistic Regression / Maximum Likelihood Estimation

Classification: Support Vector Machines

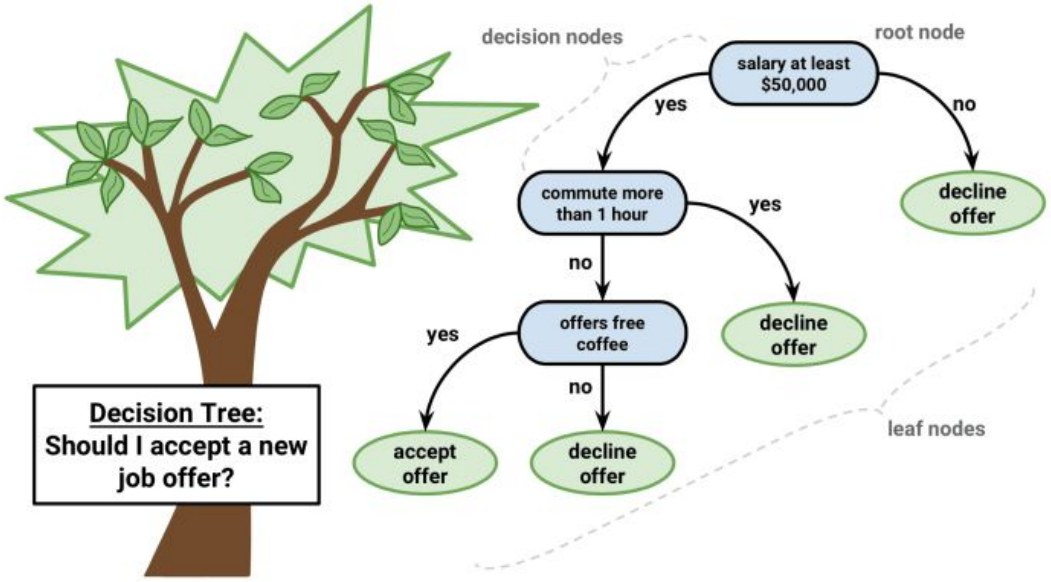


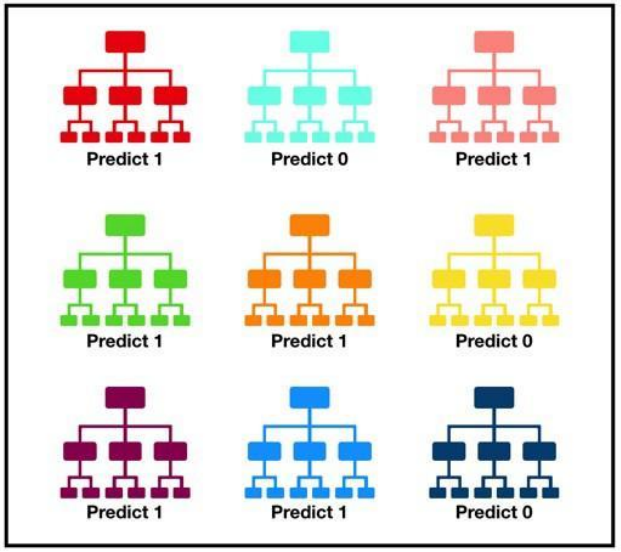


Support Vector Machines (SVM)

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

Classification: Random Forests



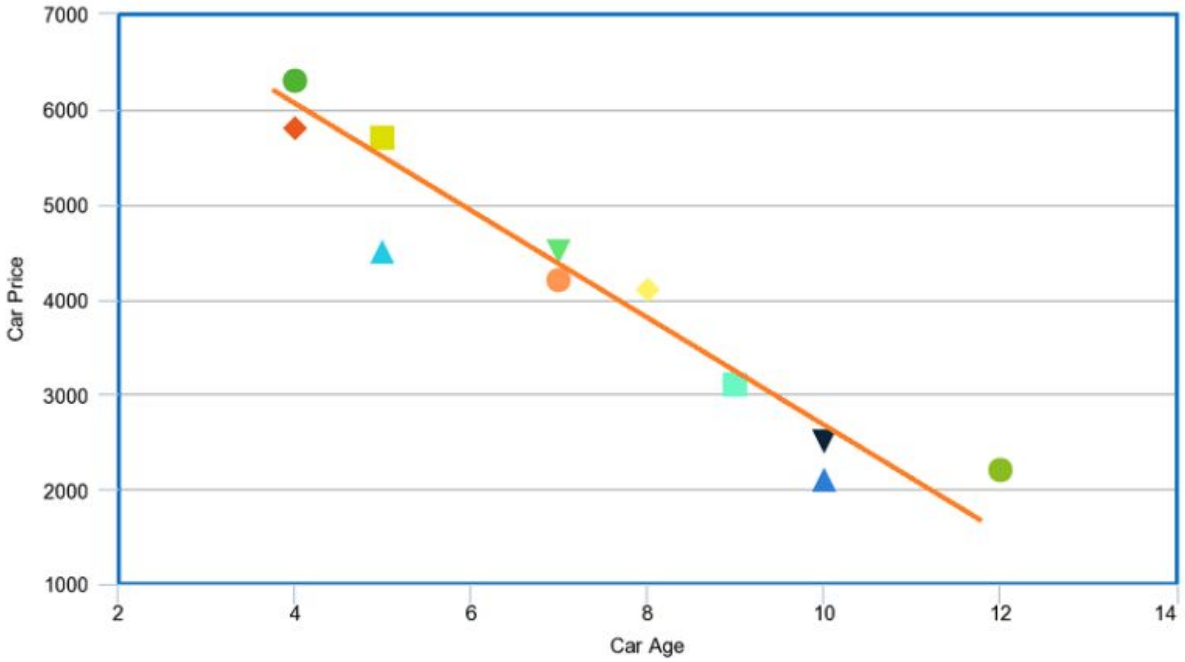


Tally: Six 1s and Three 0s
Prediction: 1

Random Forests (RF)

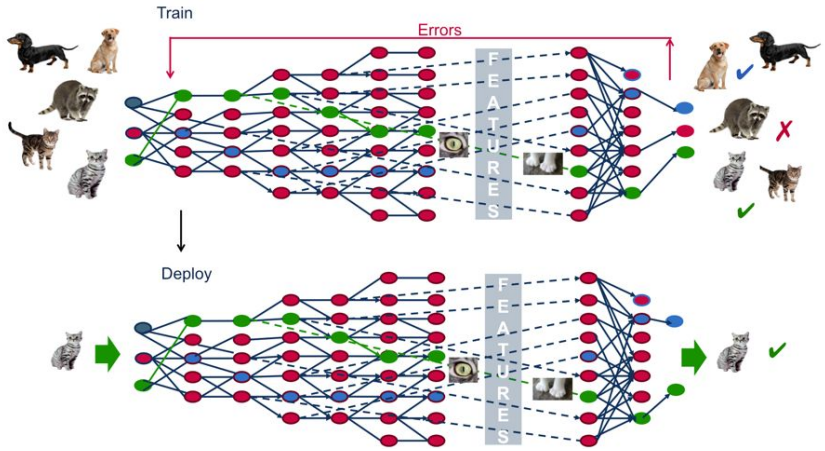
- Growing Decision Trees
- Classification with Random Forests
- Randomization techniques for RF

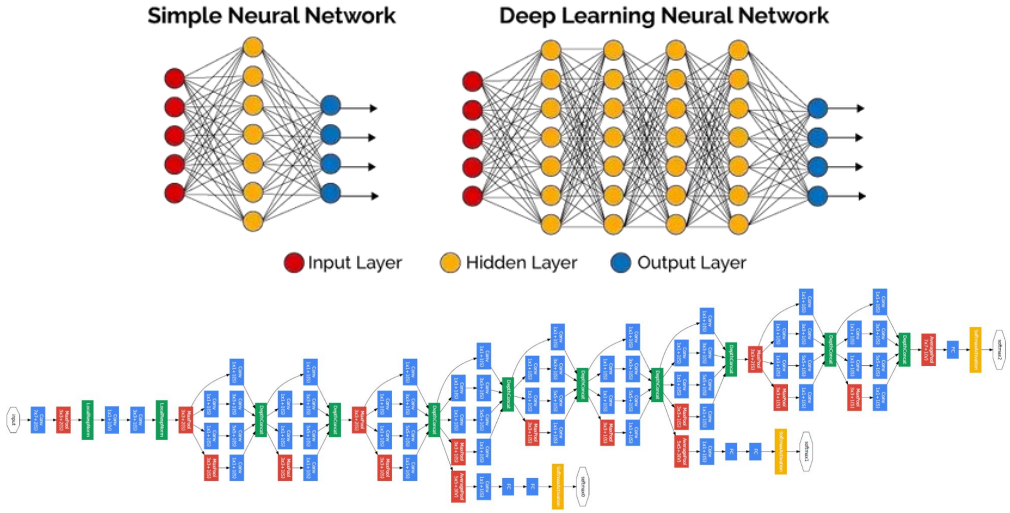
Regression



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

Neural Networks and Deep Learning

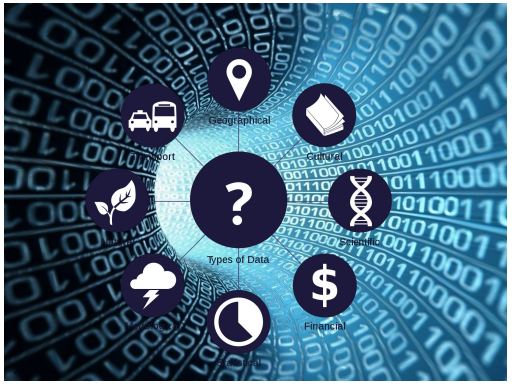




Neural Networks (NN) and Deep Learning

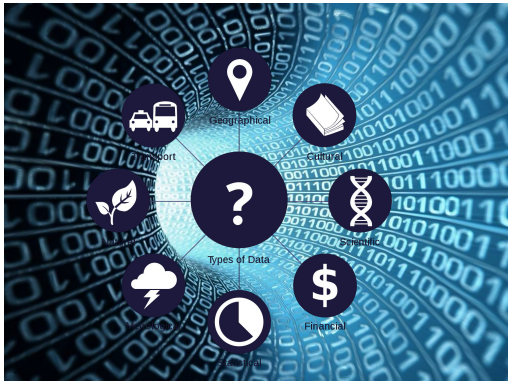
- Basic Neural Network model
- Multi-layer (deep) feedforward neural networks
- Convolutional Neural Networks (CNN)
- *Quick overview of Advanced models (RNN, GANs…)*
- Examples of applications

Machine Learning (unsupervised)

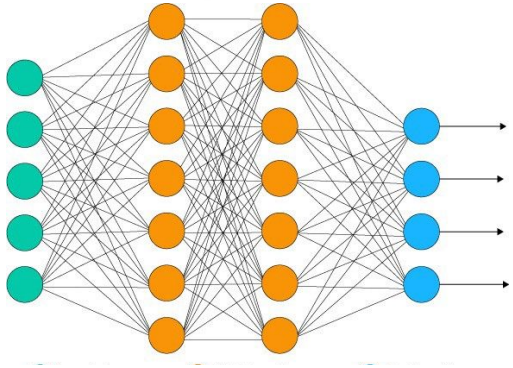


Training data
(without labels)

Training procedure

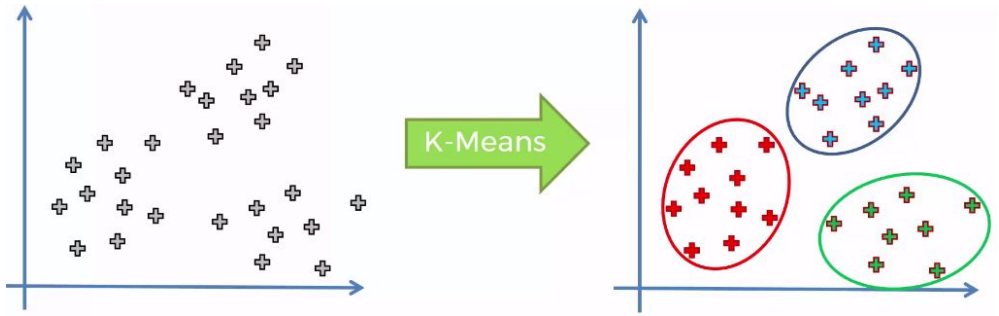


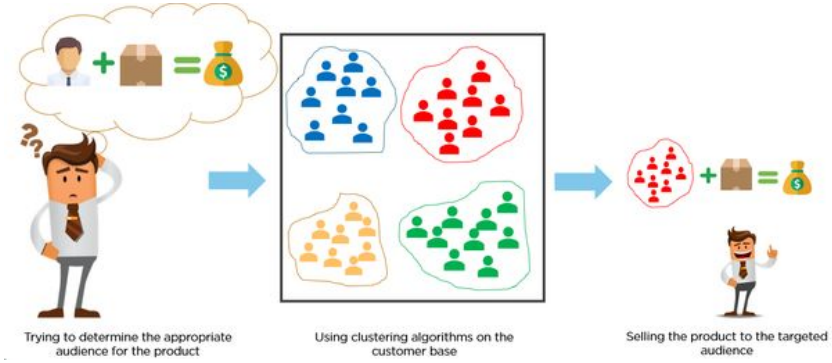
Data to be analyzed



ML model
(estimate parameters)

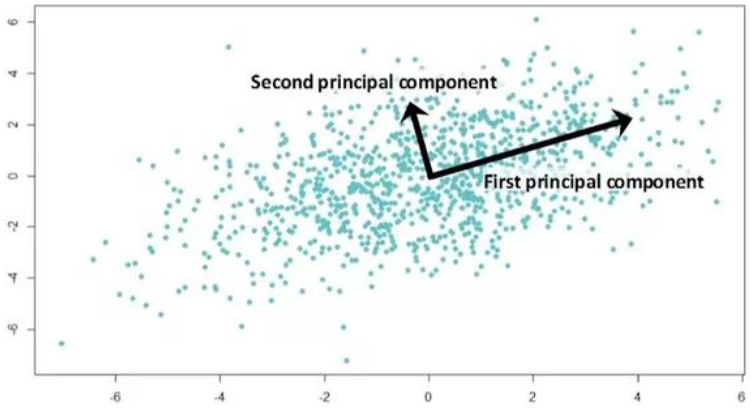
Clustering





- Basics of Clustering
- Linkage-based clustering
- K-means clustering

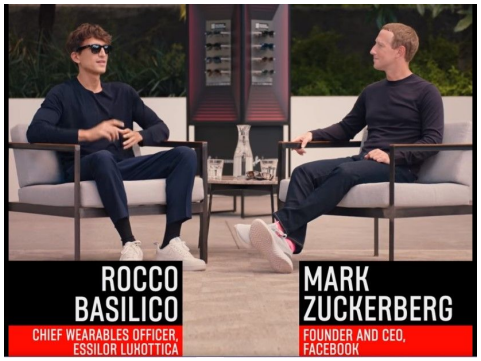
Dimensionality Reduction



Principal Component Analysis (PCA)

DIPARTIMENTO
DI INGEGNERIA
DELL'INFORMAZIONE

Labs: Python Programming



- We'll use Python + scikit learn
- Still many things to solve with Lab room availability, probably mix of department labs and/or take-home assignments with online support
  - All the work can be done from home with a standard PC
  - Install the software on your laptop
- *(if available on time)* case studies with smart glasses from Luxottica
- Libraries: scikit-learn, numpy
- Jupyter notebook (mix code and text, avoid separate report)
- Installation can be done through Anaconda

Laboratories

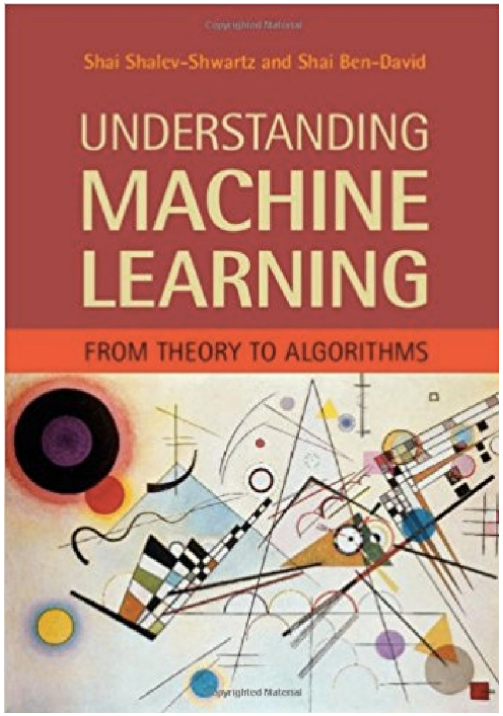
5 Labs:

1. **02 NOV** Introduction to Python
2. **16 NOV** Regression and Classification (HW1)
3. **30 NOV** Support Vector Machines (HW2)
4. **21 DEC** Neural Networks (HW3)
5. **18 JAN** Tutorial: Keras Deep Learning framework (*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	Delivery
1	15/11	29/11
2	29/11	13/12
3	20/12	09/01

** Tentative dates, will probably change*

- 3 Homeworks
- Two weeks period for each homework:

1. Homework is released
2. Support session (lab and/or Zoom)
3. Delivery deadline **(hard)** in approximately 2 weeks

- Up to 3 extra points for the homeworks (1pt for each homework)

Written Exam

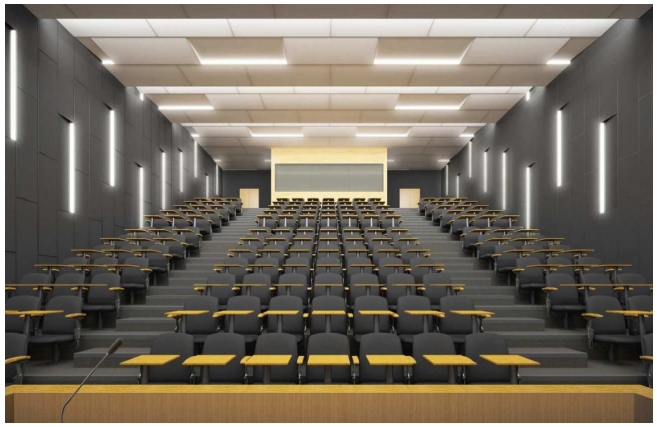
- Written exam in classroom at the end of the course
- No orals; No online exams
- Final mark is the written exam score + the homework score
- Can get to "*30*" without the homeworks but extra points help !
- Dates for the exams:

1. 24/01/2023
2. 09/02/2023
3. 28/06/2023
4. 07/09/2023
5. Extra session in September 2023 for Physics students (TBA)

Check the exam dates
No out-of-session exams

Exams will be in classroom only
No online exams

Lectures



- Wed 16.15-18.00 **Room Ae** + recorded
- Fri 16.15 - 18.00 **Room Ae** + recorded
- Classroom attendance is recommended
- Use the recorded lectures only in case of issues
- Labs: details will be announced when available

Labs: Setup your PC

- It is strongly suggested to ensure that you are able to develop and run the assignments on your PC
- We'll use Pyhton + scikit learn
- Simple tasks, any "standard" PC should be sufficient

Setup your home PC or laptop

For your PC:

- ❑ 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

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`

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 (comments / description)
- We'll use jupyter notebooks for the lab deliveries

Tutorial

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

DIPARTIMENTO
DI INGEGNERIA
DELL'INFORMAZIONE

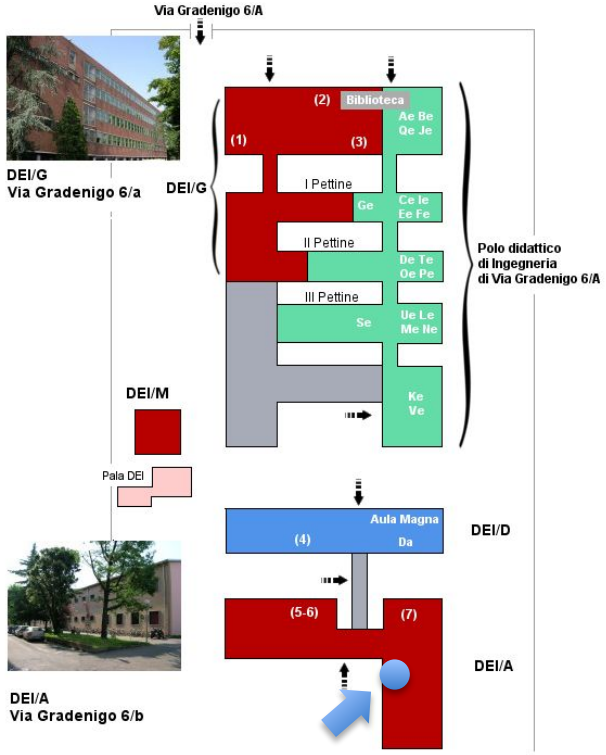
Contacts

Instructor

- Dr. Federico Chiariotti
- Email: chiariot@dei.unipd.it
- Office: TBA

Teaching Assistant

- Anay Deshpande
- Email: deshpande@dei.unipd.it

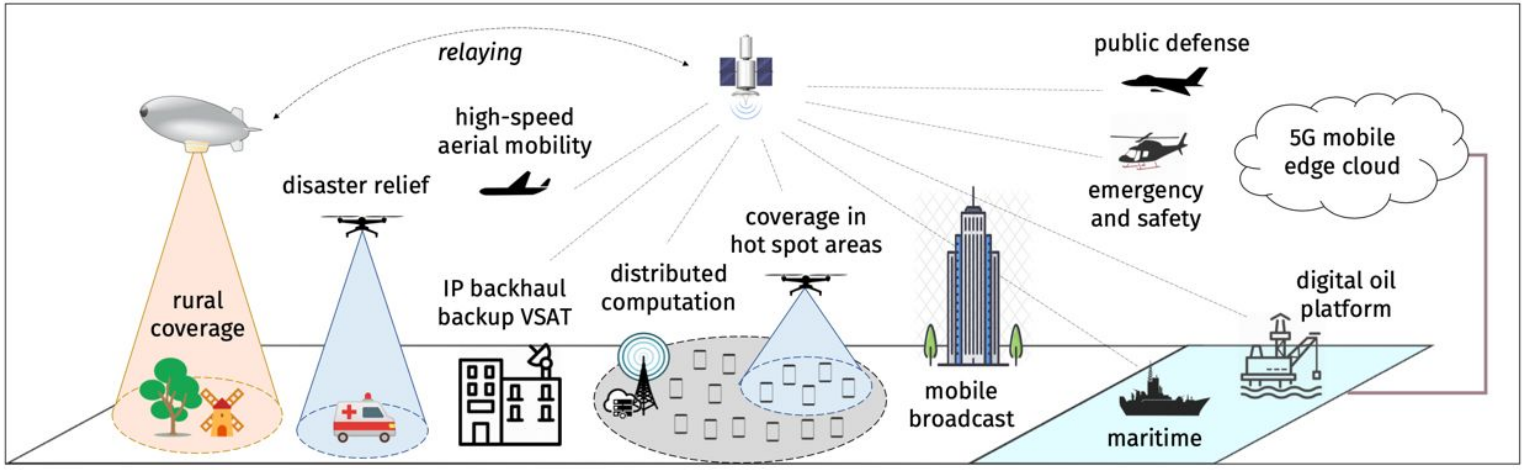


SIGNET project opportunities

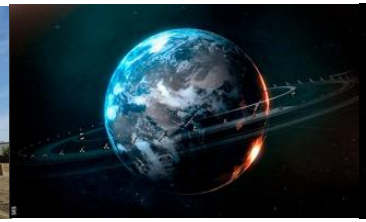
- Could be useful for thesis or course projects
  - The scope of the project can be tuned
- You can use ML in a real research context
- You can also propose your own ideas!

Non-Terrestrial Networks

- Complementing terrestrial infrastructures with aerial nodes (drones, satellites, high altitude platforms, etc.)
- ML optimization: how do we distribute comm/computation tasks?



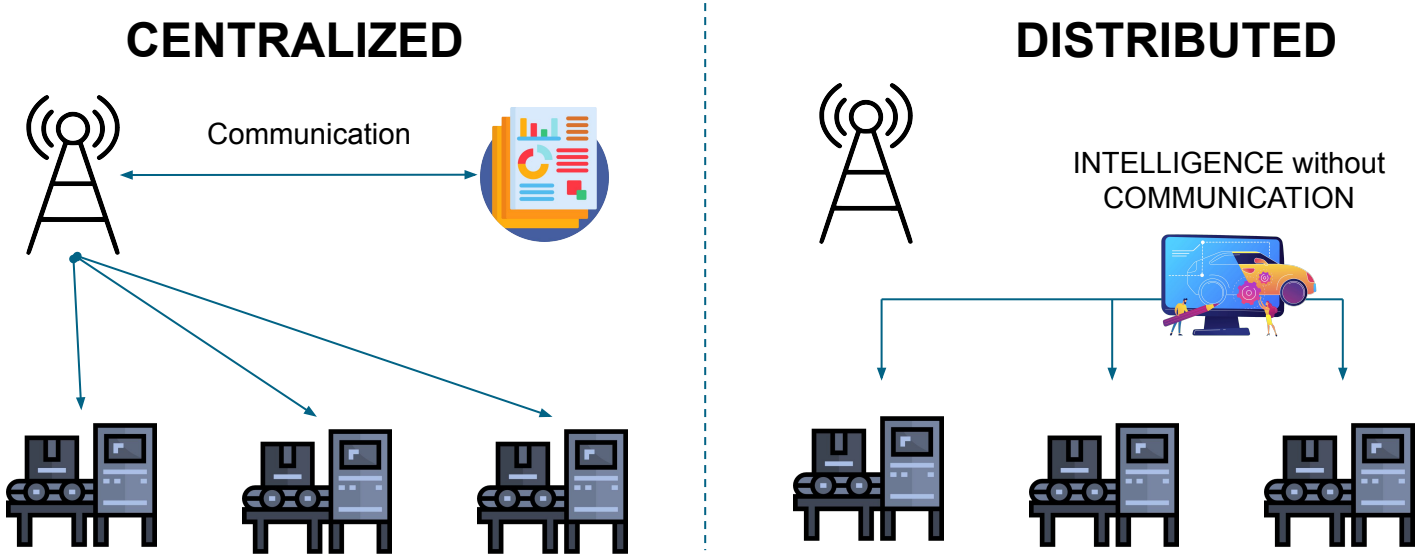



Industrial IoT

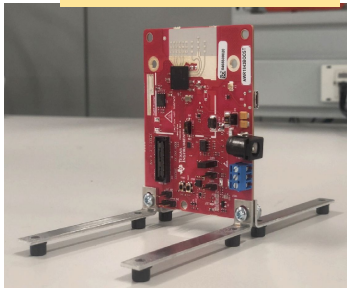
URLLC: extremely low latency services

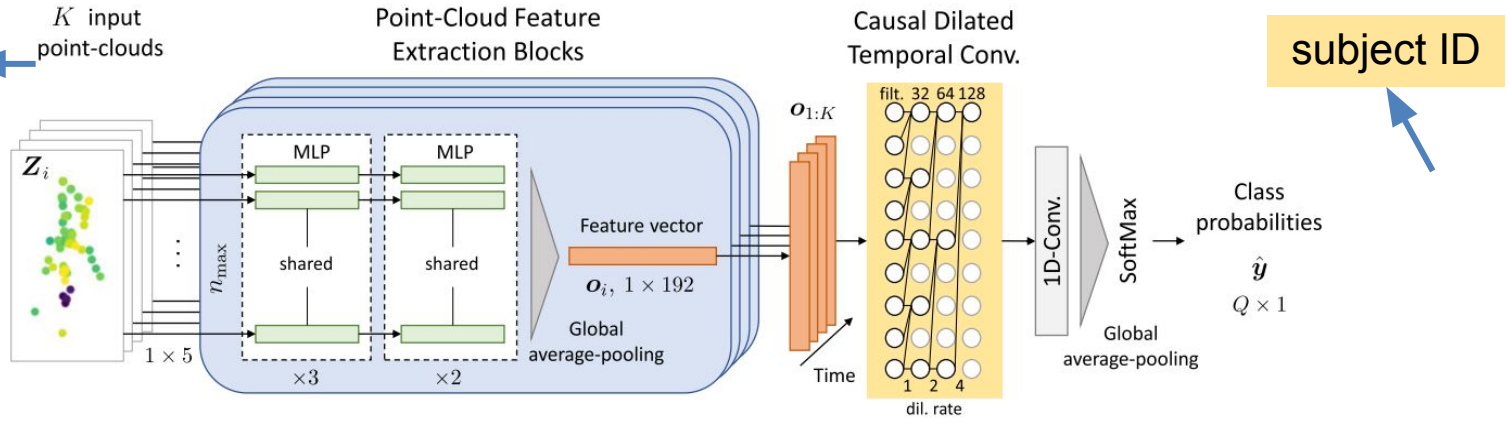
How do we distribute intelligence to meet the deadlines?

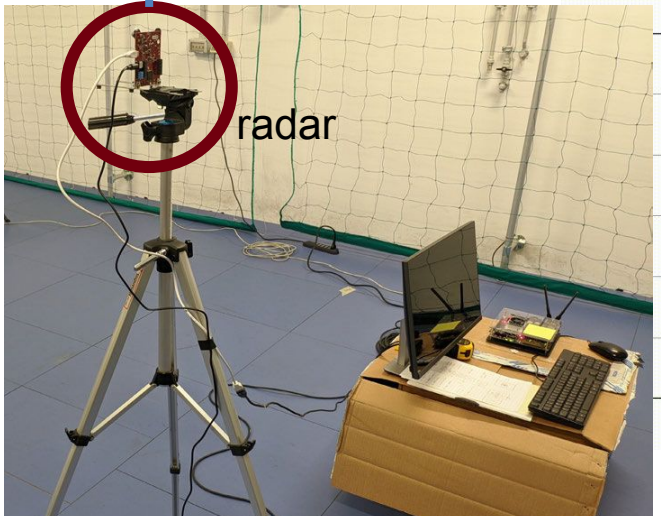


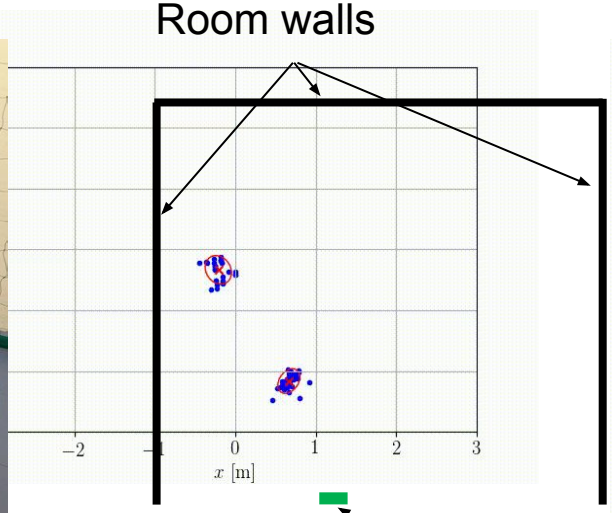
Radar identification

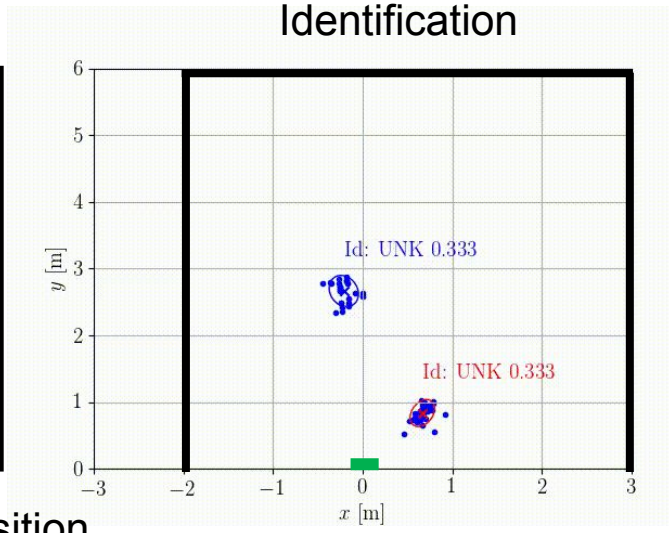
radar raw data



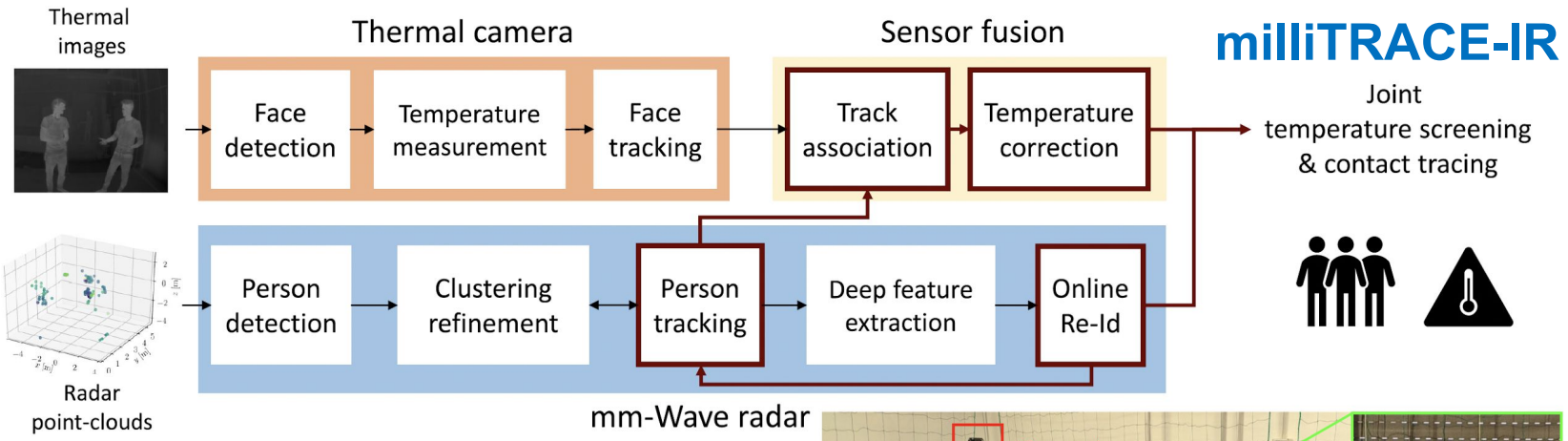


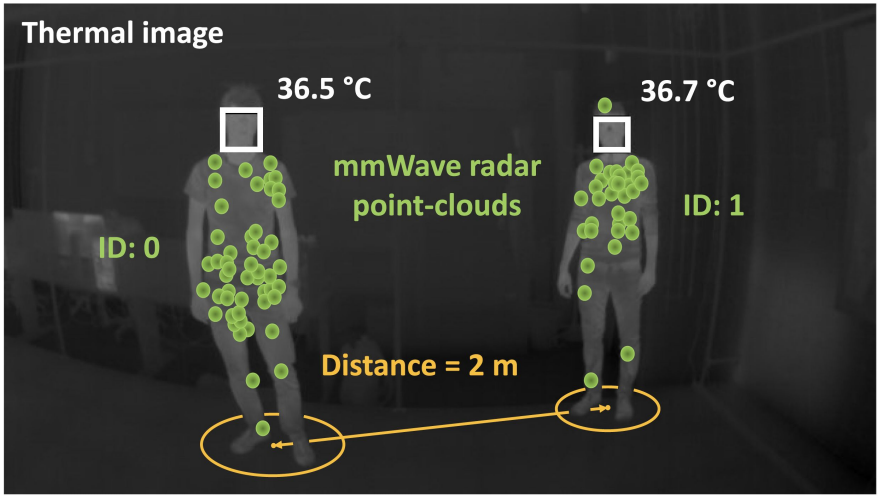


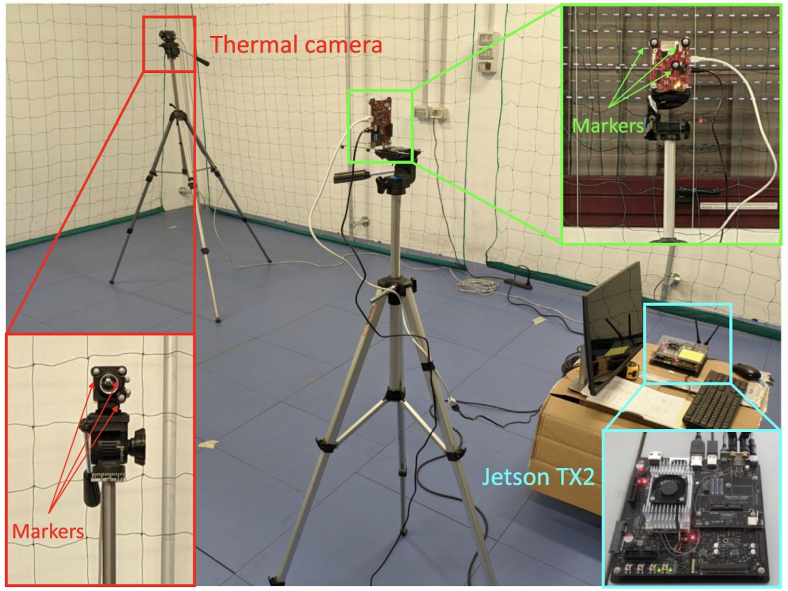




Temperature and contact tracing

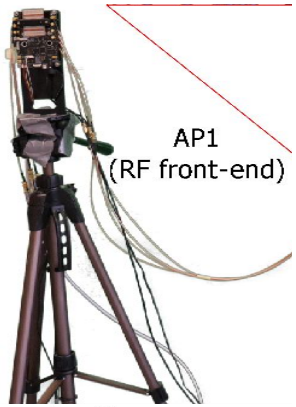


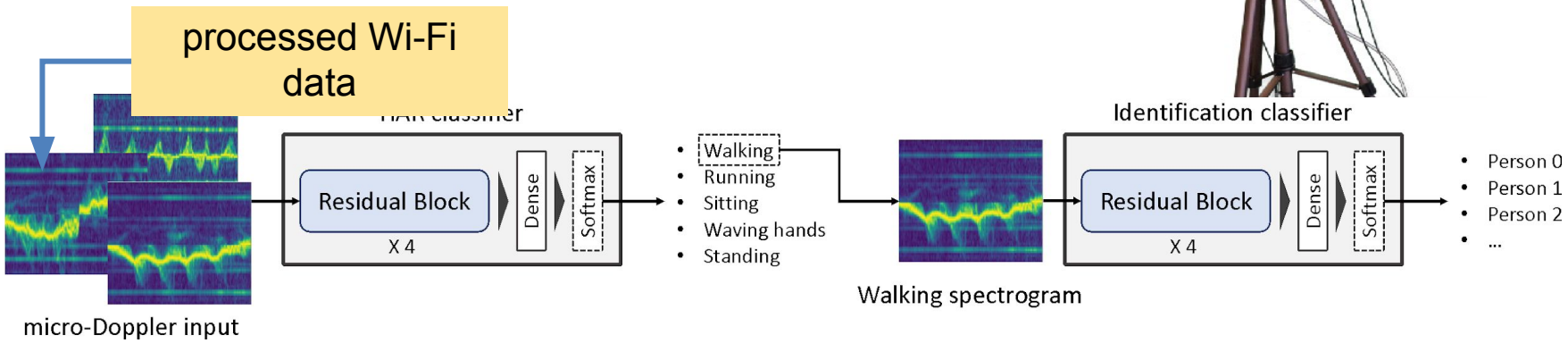


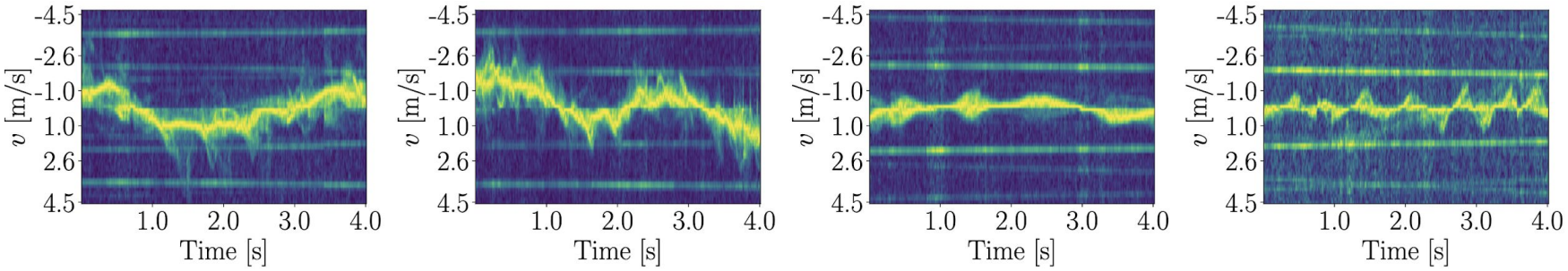


Wi-Fi sensing

- **RAPID** – indoor human detection and sensing through a testbed implementing the new IEEE 802.11ax Wi-Fi standard at 60 GHz



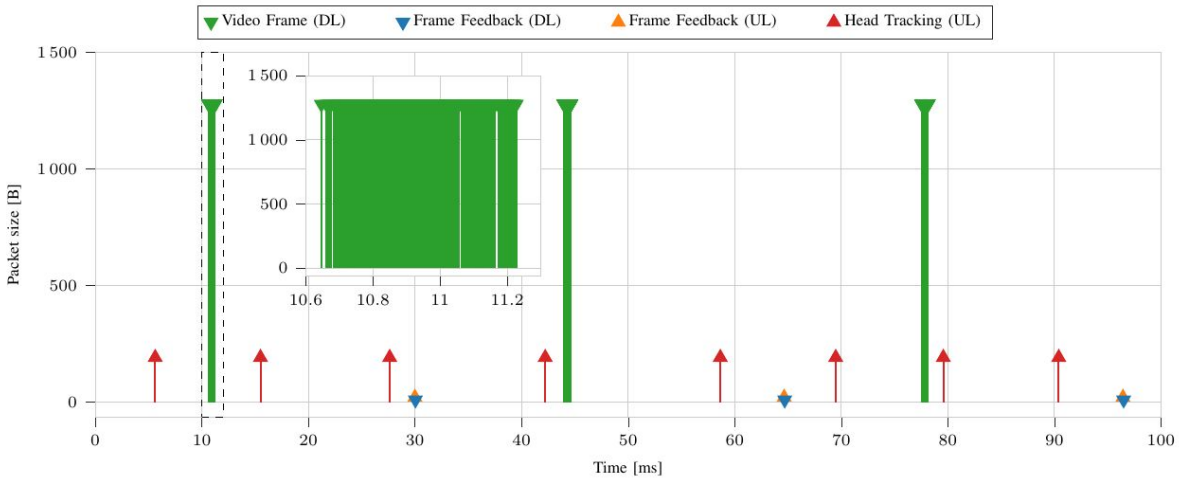




Walking. Running. Sitting down. Waving hands.

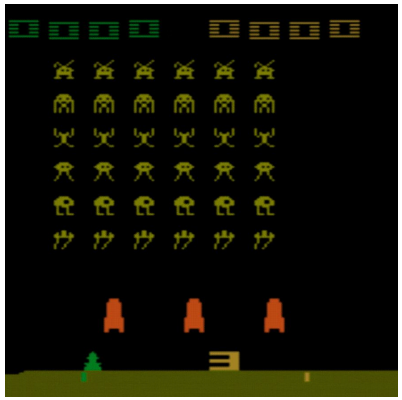
VR trace analysis

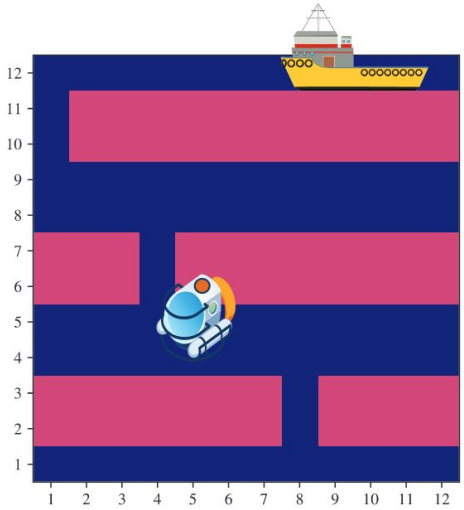
- How do we predict VR traffic?
- Frames depend on activity: what is the user doing?
- ML applied on traffic traces: capture and analysis tools

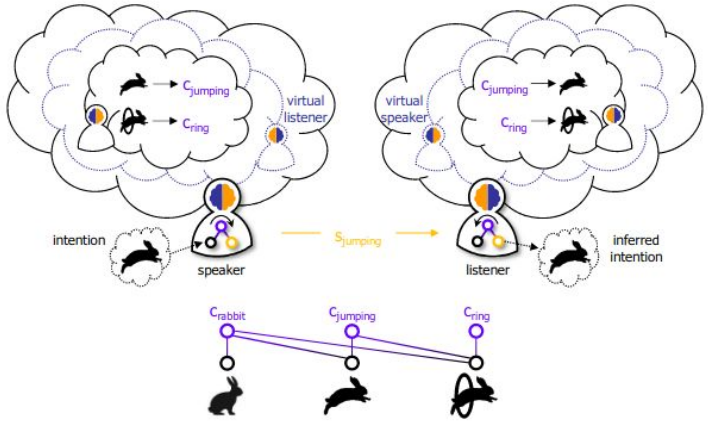


Semantic communications

- Adapting communications to only send the most relevant information
- Mix of ML styles: reinforcement, supervised, unsupervised
- Theory of mind: how do we model other agents?







Important logistic info

- Friday, September 30: **NO LECTURE**
- Wednesday, October 5: **LECTURE 2**
- Friday, October 7: **NO LECTURE**
- From the week starting on 10/10, we will have class regularly every Wednesday and Friday