



MASTER IN ENTREPRENEURSHIP
INNOVATION MANAGEMENT
IN COLLABORATION WITH **MIT SLOAN**

IN COLLABORATION WITH



UNIVERSITÀ DEGLI STUDI DI NAPOLI
PARTHENOPE

MASTER MEIM 2021-2022

Environment preparation

Lesson given by Prof. Paola Barra

University Parthenope of Naples

Overview

- Jupyter notebook
- OpenCV
- Pandas
- Scikit learn
- Supervised learning
- Practical examples

Jupyter notebook



Jupyter notebook is a web application that allows you to create documents that contain formatted text (using a simple markdown language) and executable code within the document, with output views and graphics

TO RUN USE THE FOLLOWING COMMAND: **“jupyter notebook”**

Jupyter notebook is installed by default in conda but is also installable with the command:
- pip install jupyter

OpenCV



Open Computer Vision Library: collection of open source algorithms for Computer Vision and Image Processing.

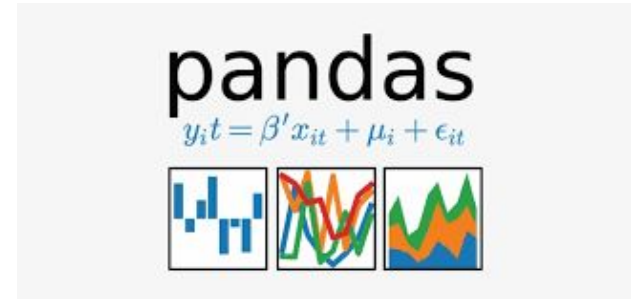
Available for Windows, Unix and MacOS X environments, the following command automatically installs the latest version of OpenCV available.

```
pip install opencv-python
```

Import THE LIBRARY IN PYTHON WITH :

```
import cv2
```

pandas



Pandas is a library for manipulating data in sequential or tabular format. It is ideal for data analysis.

Main features of Pandas are:

- Loading and saving of table data, such as CSV, TSV, Excel files and database formats
- Simplicity in performing indexing and data aggregation operations
- Simplicity in the execution of numerical and statistical operations
- Simplicity in viewing the results of operations

The following command installs the latest version of Pandas:

```
conda install -c anaconda pandas
```

Matplotlib



Matplotlib is a comprehensive library for creating static, animated, and interactive visualizations in Python. Matplotlib makes easy things easy and hard things possible.

- Create [publication quality plots](#).
- Make [interactive figures](#) that can zoom, pan, update.
- Customize [visual style and layout](#).
- Export to [many file formats](#).
- Embed in [JupyterLab and Graphical User Interfaces](#).
- Use a rich array of [third-party packages](#) built on Matplotlib.

Scikit-learn



Machine Learning with Scikit-Learn

Is the best Machine Learning library in the world.

Is important that you are familiar with python and statistics.

You can load dataset using **pandas**. Is Built on Numpy, Scipy & Matplotlib.

There is a huge documentation on scikit-learn.org

Install it:

```
conda install -c anaconda scikit-learn
```

Supervised learning

X = data	y = target
70% TRAINING SET = X_train	y_train
30% TEST SET = X_test	y_test

Learn some properties from the training set data and test these properties in the test set.

It is a common practice to take the initial dataset and divide it into two training sets and test sets, generally 70% training set and 30% test set.

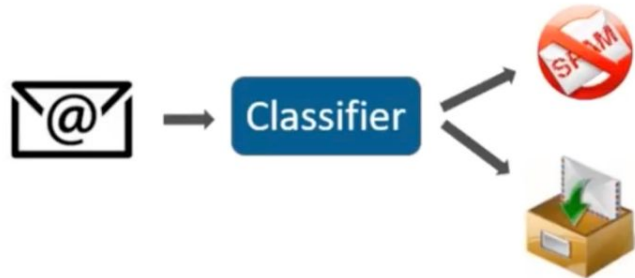
Training set: data already labeled.

Test set: data without class.

Supervised learning: Classification VS Regression

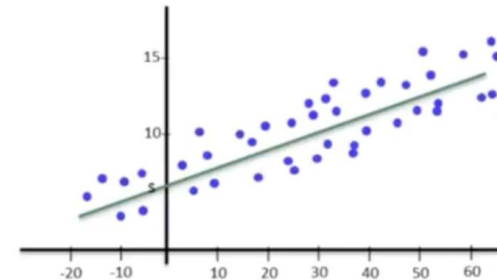
Classification is to associate the observation with a category, taken from a finite set of categories (or classes).

For example: classify a disease as "positive" or "negative" based on symptoms.



Regression is the prediction of a number taken from a continuous set of values.

Example: predicting the price of a house based on its characteristics.



Classification

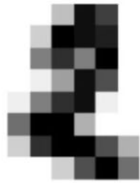
Training: 0



Training: 1



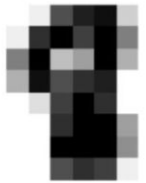
Training: 2



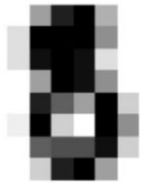
Training: 3



Prediction: 8



Prediction: 8



Prediction: 4



Prediction: 9



The data belongs to one or more classes, we must learn from the already labeled data to predict the class of the unlabeled data.

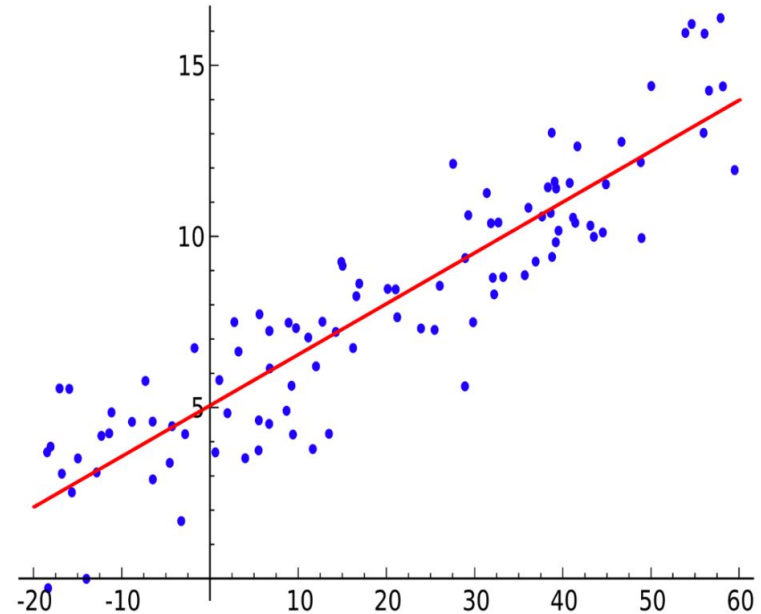
It consists in observing to which set of categories a new observation belongs on the basis of the information acquired from the training set.

An example of a classification problem in Computer Vision is trying to predict the handwritten number in this case the labels are 10 (0 to 9).

Regression

Regression differs sharply from classification, since the latter is limited to discriminating the elements in a certain number of classes (labels).

While in the regression the input is given and the system gives us a real output, therefore a discrete value.



Let's get our hands on the code





MASTER IN ENTREPRENEURSHIP
INNOVATION MANAGEMENT
IN COLLABORATION WITH **MIT SLOAN**

IN COLLABORATION WITH
MIT MANAGEMENT
SLOAN SCHOOL



UNIVERSITÀ DEGLI STUDI DI NAPOLI
PARTHENOPE

MASTER MEIM 2021-2022

Naive Bayes

Lesson given by prof. Paola Barra

University Parthenope of Naples



MASTER IN ENTREPRENEURSHIP
INNOVATION MANAGEMENT
IN COLLABORATION WITH **MIT SLOAN**

IN COLLABORATION WITH
MIT MANAGEMENT
SLOAN SCHOOL



UNIVERSITÀ DEGLI STUDI DI NAPOLI
PARTHENOPE

MASTER MEIM 2021-2022

Decision Tree

Lesson given by prof. Paola Barra

University Parthenope of Naples



MASTER IN ENTREPRENEURSHIP
INNOVATION MANAGEMENT
IN COLLABORATION WITH **MIT SLOAN**

IN COLLABORATION WITH
MIT MANAGEMENT
SLOAN SCHOOL



UNIVERSITÀ DEGLI STUDI DI NAPOLI
PARTHENOPE

MASTER MEIM 2021-2022

MLP

Lesson given by prof. Paola Barra

University Parthenope of Naples



MASTER IN ENTREPRENEURSHIP
INNOVATION MANAGEMENT
IN COLLABORATION WITH **MIT SLOAN**

IN COLLABORATION WITH



UNIVERSITÀ DEGLI STUDI DI NAPOLI
PARTHENOPE

MASTER MEIM 2021-2022

CNN - Image classification

Lesson given by prof. Paola Barra

University Parthenope of Naples