

POLIMI GRADUATE
SCHOOL OF **MANAGEMENT**

INTRODUCTION TO PYTHON - LAB SESSION

PERCORSO EXECUTIVE DATA SCIENCE AND BUSINESS ANALYTICS

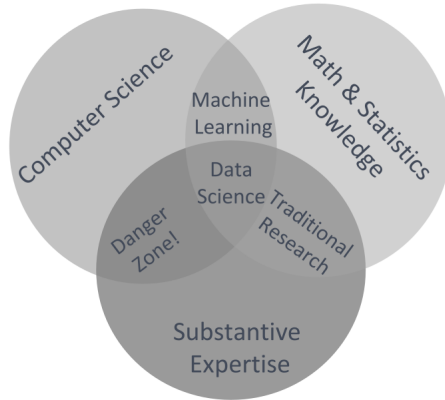
Mauricio Soto - mauricioabel.soto@polimi.it

MAURICIO SOTO

- ▶ Researcher at DIG-POLIMI
 - mauricioabel.soto@polimi.it
 - BL26 - 2.05
- ▶ Previous
 - Mathematical Engineering - U. of Chile
 - PhD. Computer Science - U. of Paris
 - Researcher - U. of Orléans, U. of Chile, UNIMIB
- ▶ Research interests
 - Graph theory
 - Evolutionary networks
 - Data representation (NLP)

DATA SCIENCE

- Drew Conway, 2010



@manudellavalle - <http://emanueledellavalle.org>

QUESTION TIME



<https://pollev.com/mauriciosoto>

ML PYTHON LABS

Feb 13	Setup & Introduction to Python Basics	Soto
Feb 15	Python and Data Science Libraries (Navigli)	Soto
Feb 21,22	Exploratory Data Analysis	Soto
Feb 22,27	Classification	Mor
Mar 8,9	Regression	Mor
Mar 13,15	Text Analysis	Mor
Mar 23	Assignment discussion	

Material: <https://github.com/mauriciosotogomez/BABD2023> [▶ Link](#)

SETUP: WHAT WE NEED

- ▶ Python ≥ 3.6
- ▶ Sklearn
- ▶ Jupyter Notebook
- ▶ Numpy
- ▶ Pandas
- ▶ Matplotlib
- ▶ Seaborn
- ▶ Orange



pandas

$$y_{it} = \beta' x_{it} + \mu_i + \epsilon_{it}$$



matplotlib

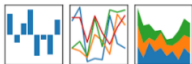
DATA SCIENCE & MACHINE LEARNING LIBRARIES



A library providing support for large, multi-dimensional arrays and matrices, along with a large collection of high-level mathematical functions to operate on these arrays.

pandas

$$y_{it} = \beta' x_{it} + \mu_i + \epsilon_{it}$$



A library for data manipulation and analysis. In particular, it offers data structures and operations for manipulating numerical tables and time series.

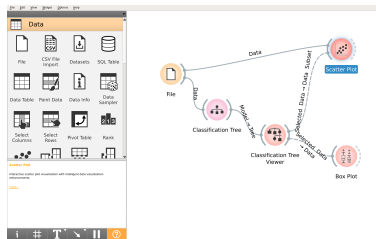


A library featuring various classification, regression and clustering algorithms including support vector machines, random forests, k-means..., and is designed to interoperate with the Python numerical and scientific libraries e.g. NumPy.

PROGRAMMING TOOLS

1. Orange

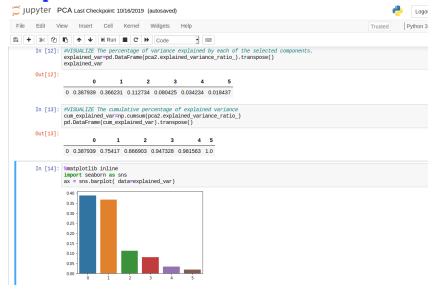
<https://orange.biolab.si/>



- ▶ Intuitive interface
- ▶ Fast development

2. Jupyter-Notebook (Anaconda)

<https://www.anaconda.com/>



- ▶ Advanced functions
- ▶ Customization

A library featuring various ML algorithms designed to inter-operate with the Python numerical and scientific libraries e.g. NumPy, Pandas.

<https://scikit-learn.org/stable/>

EXERCISE 1. FUNCTIONS

Create a function that given two integers m, n computes the greatest common divisor between m and n .

EXERCISE 2. NUMPY

1. Create a random array of length 100.
2. Sort your array.
3. Compute the mean, median and sample variance.

EXERCISE 2. NUMPY

1. Create a random array of length 100.
Hint: `np.random.rand()`
2. Sort your array.
3. Compute the mean, median and sample variance.

EXERCISE 3. PANDAS

1. Import the dataset *iris* as a DataFrame
2. Add the columns names (sepal length, sepal width, petal length, petal width).
3. Create a new column that contains the ratio between the sepal and petal length.
4. Add a new column named `target` with value 1 if the type is setosa and 0 otherwise.

THANK YOU