BABD

Masters in Business Analytics and Big Data

Python Basics

ML Python Labs

02.15	Setup & Introduction to Python Basics
02.16	Python and Data Science Libraries
02.21,22	Exploratory Data Analysis
02.23	Classification
03.01,02	Classification
03.14,15	Regression
03.16,17	Text
03.21,22	Text

Material: https://github.com/mauriciosotogomez/BABD2022 Link

Setup: What we need

- Python $\geq$3.6
- Sklearn
- Jupyter Notebook
- Numpy
- Pandas
- Matplotlib
- Seaborn

https://pollev.com/mauriciosoto

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.

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.

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.

Exercise 4.

1. Find a database of your particular interest.
2. Formulate a research question (statistic,prediction, classification, etc)