Machine Learning

Efrei Paris
Big Data & Machine Learning
M1

2020/2021

Machine Learning

Salim NAHLE

Organization:

- You can work on any Python environment
- 1 **PDF /HTML report file** is expected.
- It shall contain the code (executed, explanations and necessary screenshots). You can simply print your notebook into a PDF/HTML files.
- Please work in **pairs**! Each group (composed of 2 persons at most) shall submit one report. Do not forget to indicate your names in the report. The same pairs shall be maintained all the semester.
- The report shall be uploaded on the Moodle's page before **Tuesday 08/12/2020 at 11:45 pm.**
- Late reports are penalized (5 points/20 per day).

Consulting Project Use Case

You have been recruited as a Data Scientist consultant at a Finance company that proposes housing loans. When a customer applies for a housing loan, the company studies his demand to decide whether he/she is eligible or not.

You shall develop a decision-aid tool to automate the loan eligibility process. Your tool shall take as input the information provided by the customer while filling his/her online loan application form. These include customer's Gender, Marital Status, Education, Number of Dependents, Income, Loan Amount, Credit History, etc...

You are given a data set containing historic information about applicants (**features**) as well as their loan status (**output** given as binary variable indicating whether or not a loan was approved).

Lab expected outcome:

The objective or this lab is to implement Logistic Regression from scratch using gradient descent and the sigmoid function.

Apply on the banking loan dataset

Compare the obtained results with those obtained from Scikit learn implementation of logistic regression.