Machine Learning - Python Lab Exam – 07-01-2020

Fit a classifier for the included dataset.

The solution must be produced as a Python Notebook.

The notebook must include appropriate comments and must produce:

1. a pairplot of the data (see Seaborn pairplot) and a comment on remarkable situations, if any (**5pt**)
2. a classification model using a method of your choice with the schema "train-validation-test" exploring an appropriate range of parameter values (**5pt**)
3. the optimal parameter(s) (**5pt**)
4. a scatter plot of the test set using a pair of attributes of your choice with the class as colour (**5pt**)
5. ... and the good/bad prediction as the point style (**5pt**)
    1. hint: the seaborn scatterplot function allows a "style" parameter which is a vector of values; this can be obtained as a comparison between the true and the predicted target in the test set, see https://seaborn.pydata.org/generated/seaborn.scatterplot.html

Quality of the code (**6pt**):

1. The python cells must be preceded by appropriate comments
2. Useless cells and pieces of code will be penalised
3. Naming style of variables must be uniform and in English
4. Bad indentation and messy code will be penalised

Additional directions, the assignments not compliant with the rules below will not be considered

1. The notebook name must be ***machineNumber_lastname_firstname.ipynb***
    1. for example, if I am sitting on the machine lab42, my notebook will be 42_sartori_claudio.ipynb
2. The first cell must contain the machine number, the last name and first name of the student.
3. The solution must directly access the data in the same folder of the notebook

Cooperative work will be **heavily sanctioned.**

The candidate can freely access the manuals available online in:

scikit-learn.org

docs.scipy.org

pandas.pydata.org

matplotlib.org

seaborn.pydata.org

The candidate can freely access the teaching materials available in the course website, including the available examples of python notebooks.

The notebook must be uploaded in **both** *original* and *pdf* form, as two separate files.