Homework 4

Data Analysis and Classification 2019-2020
Classification

winequality-red.csv dataset (iCorsi)
winequality-white.csv dataset (iCorsi)
research paper: Cortez et. al. [2009] (iCorsi)

This homework could be developed as a Python project **OR** a Jupyter Notebook.

- Python project. Insert your assumptions, answers, discussions, decisions as comments within the code. The script developed should be named as **<surname_homework_4>.py** (in case of multiple files please number them as *surname_#_homework_4.py*)**.**
- Jupyter Notebook. Each question needs to have at least a Code Cell (implementation) and a Markdown Cell (explanation and/or answer). The notebook developed should be named as **<surname_homework_4>.ipynb .**

The produced code should to be sent via email at michela.papandrea@supsi.ch by sunday 8.12.2019.

This homework is based on the wine quality classification research paper and dataset shared on the course page.

What to do

The homework consists in the following parts.
1. Build a classification model which is able to predict the quality of the wine (multiclass classification) based on its physicochemical values. Use AdaBoost and Gradient Boost applied to decision trees, and compare the results with the ones obtained in HW3 (Decision Tree and Random Forest).
2. Build a classification model which is able to predict the quality of the wine (binary classification) based on its physicochemical values. Use SVM, and

compare the results with the ones obtained in HW2 - question 2 (Logistic Regression).

Approach

The idea is to apply different approaches and evaluate them, in terms of Accuracy, Precision-per class and Recall-per class. Per each approach, plot also the Confusion Matrix and make some reasoning over it. Make a conclusion, proposing your best model.

Suggestions

Perform model parameters tuning, when appropriate.