Homework 3

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 has to be developed on a Jupyter Notebook. Each question needs to have at least a Code Cell (implementation) and a Markdown Cell (explanation and/or answer). The notebook developed, named as **<surname_homework_3>.ipynb** has to be sent via email at michela.papandrea@supsi.ch by sunday 24.11.2019.

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 building a classification model which is able to predict the quality of the wine (multiclass classification) based on its physicochemical values.

Approach

The idea is to apply different approaches and evaluate them, in terms of Accuracy, Precision-per class and Recall-per class. You can apply the hold-out validation methodology, selecting randomly 20% of a dataset for testing. Per each approach, plot also the Confusion Matrix and make some reasoning over it. Make a conclusion, proposing your best model.

Suggestions

Use Decision Tree and Random Forest classification algorithms. Try building different models and tune the algorithms parameters, in order to increase the performances of the trained models.