Question **1**

Not yet answered

Marked out of 31

Produce some classification schemes for this dataset, according to the directions below.

The solution must be produced as a Python Notebook. The last column of the data file is the target of classification

The notebook must include appropriate comments and must operate as follows:

Task	Point(s)
1. Load the data from the file and show: the first few rows, the output of the `.describe()` function, the number of rows and columns (4pt)	4
2. Since the data contain nulls, eliminate the rows with nulls	4
3. Since one of the predicting attributes is *ordinal*, it must be converted into *numeric*, you can use the `OrdinalEncoder`	3
4. Split the data into *train* and *test*	1
5. Use two classification models of your choice (say: model 1 and model 2) execute the tasks below	
6. Model 1: find and show the best hyperparameter setting with *cross validation* on the training set, optimise for the best accuracy	4

7. Model 1: show the accuracy of classification and the confusion matrix on the test set For the confusion matrix use plot_confusion_matrix normalized in order to show for each class the *precision* (read carefully the documentation)	4
8. Model 2: find and show the best hyperparameter setting with *cross validation* on the training set, optimise for the best accuracy	3
9. Model 2: show the accuracy of classification and the confusion matrix on the test set For the confusion matrix use plot_confusion_matrix normalized in order to show for each class the *precision* (read carefully the documentation)	2

Quality of the code:
(6pt)

- *Include appropriate comments with reference to the numbered requirements*
- *Useless cells, pieces of code and non-required output will be penalized*
    - *Remove the code you use for testing and inspecting the variables during the development*
- *Naming style of variables must be uniform and in English*
- *Bad indentation and messy code will be penalized*

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

1. The notebook name must be ***emailusername.ipynb*** in lowercase letters
    a. E.G. if your email is mario.rossi45@studio.unibo.it the notebook filename will be mario.rossi45.ipynb
2. The first cell must contain the student first name, last name and email
3. The solution must directly access the data in the same folder of the notebook
4. Upload the notebook only to eol, any other way of submitting the notebook will be ignored

Cooperative work will be **heavily sanctioned**

The candidate can freely access any kind of materials

Maximum file size: 100MB, maximum number of files: 1

Files

You can drag and drop files here to add them.

◀ Informatics - Theory Jump to... Machine Learning Theory ▶