Assignment 2

1 Assignment for those who are achieving projects

You are in charge of programming four Matlab scripts.

- `PNB_createSet(name_of_file,K);` [1] This function creates a training set saved in a file called `training` and a testing set saved in a file called `testing`. K is an integer number greater than two. It is roughly equal to the size of the dataset divided by the size of the testing set. All samples are either in the training set or in the testing set. Both files created contain the following data. Note that samples having 0 as a label should not be put into the training set nor into the testing set.

    - `X` a matrix having the same number of columns than that of the dataset.
    - `Y` a column vector having the same number of lines than `X`.
    - `metadata` containing the following information.
        * `F` is the number of features.
        * `classes_l` is a cell containing a list of the names of each classes.
        * `C` is the number of classes.
        * `name_of_file` name of the dataset from which this dataset is constructed.
        * `n_l` it is a list of the same length as `Y`. Each component of `n_l` is the corresponding row-number in `metadata.X_with_null_pixels`. This is useful in `assignment5.pdf` to build the predicted label images.

- `info=PNB_train1();` [2] This is an empty function and `info` only collects the number of classes.

- `y=PNB_predict1(info,x);` [3] This is a dummy function, it predicts randomly a class.

- `PNB_score(name_of_file,K);` [4] This function computes the following metrics

    - average precision
    - overall accuracy
    - average accuracy
    - minimal accuracy
    - confusion matrix

All of these metrics are averaged over K randomly chosen training and testing sets. These metrics are stored in the following fields of a structure named `score`. The name of the fields are respectively `average_precision`, `overall_accuracy`, `average_accuracy`, `minimal_accuracy` and `confusion_matrix`.

The `.pdf` document is named `project_NB.pdf` and contains any relevant information.

2 Assignment for those who are reviewing projects

The goal is to build matlab functions that achieve some basic checks on the data provided along each project. Two files are to be delivered.

The first file is a `.pdf` document. Its name is `reviewer` followed by a number and an `b` indicating that it refers to the second assignment. The first part of this document explains what is tested by each test. The second part explains for each project what has passed and what has failed with precise values showing the problem. The third part is optional, it explains what supplementary information you would request from the projects and how this information could provide more valuable testing.

The second file is a `.m` script having the same name, it runs successively the different functions contained in this file that do the different testings.

As in the first assignment, the checks could similarly check the consistency of the different informations. The checks could also be more thorough by building a fake testing database and a fake predict function and testing whether the score values are correct. The fake testing database could be very small with two or three classes. The predict function may always give the correct answer or never or always the same answer.

---

[1] N stands for the project number.
[2] N stands for the project number.
[3] N stands for the project number.
[4] N stands for the project number.

3 Discussion

Your task is first of all to read all projects and check `Progress`. You should write a single `.pdf` document, named `discussionB.pdf` discussing how all projects have undergone this first step, the difficulties that have been overcome and those that remain challenging issues. You should then express your opinion as to whether I should come back on some specific issues. You may also add some specific comments to a specific project on *Discussions* [5] and some specific questions on *Questions*. You are also expected to write in *Questions* the answers to all other questions.

---

[5]Comments should be most respectful as any work needs attention, and regardless of it being possibly wrong, it is going to be useful to get a better understanding. So there can be no shame in being wrong.