Module-Lab 1: Probabilistic Machine Learning

Exercise 1

This Exercise is part of the Lab/Practical session of the ACP ("Aprendizagem Computacional Probabilistica"). Please, open the compressed folder <available in UC-Student "Material"> where you will find the following Dataset's files (in .txt format):
Y.txt, Labels.txt, Y2.txt, Labels2.txt

Each line in the file/dataset corresponds to an examples/entity/object thus, the dataset contains 1400 examples.

In this exercise the variable **Y** comprises the prediction of a given supervised classifier (eg, a SVM or a Neural Network), while the labels/**ground-truth** is given by the variable named *Labels*. This is a two/binary classification problem.

Note: Students can use Matlab or Python to coding.

Let's consider **positives** all the examples labelled by 1 (ie, digit = 1) while label = 0 represents the negative class.

a) Using the files *Y.txt, Labels.txt* ie, the variables **Y** and **Labels**, develop a code to open the provided files (eg, `load('Labels.txt');` in Matlab) and then compute:

- The number of examples per class
- The number of true positives (TP) per class
- The number of true negatives (TN) per class
- The number of false positives (FP) per class
- The number of false positives (FN) per class
- The respective rates ie, TPrate, TNrate, FPrate, FNrate

b) Develop your own code to calculate the following performance measures:

- Accuracy
- Balanced accuracy
- F1/F-score
- Precision
- Recall

Exercise 2

Repeat the exercise 1.a and 1.b but this time make use of the files *Y2.txt, Labels2.txt*, and respective variables, **Y2** and **Labels2**.

$$TPrate = \frac{TP}{TP+FN} \qquad TNrate = \frac{TN}{TN+FP} \qquad FPrate = \frac{FP}{FP+TN} \qquad FNrate = \frac{FN}{FN+TP}$$

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \qquad Balanced\ Accuracy = \frac{TPrate + TNrate}{2}$$

$$Precision = \frac{TP}{TP+FP} \qquad F_1\ score = \frac{2 \times TP}{2 \times TP + FP + FN}$$