CS454-554 Homework 2: Nearest-Mean and Nearest-Neighbor Classification Spring 2022/2023

In this homework, you will implement the Nearest-Mean and the Nearest-Neighbor Classifiers (*k*-Nearest-Neighbor with *k*=1) on the Iris dataset (https://en.wikipedia.org/wiki/Iris_flower_data_set).

You are provided with two dataset files. As their names imply, **train_iris.csv** file will be used for training and **test_iris.csv** file will be used for testing. Each row of the files corresponds to one instance. Each instance is four-dimensional and there are three classes. You have 75 instances for training and 75 instances for testing. The first 4 columns of each row are the input while the fifth column contains the code of the class that the instance belongs to.

You should use the training set to train your classifiers, and the test set, unused during training, to estimate the performance of your classifiers.

1) For the Nearest-Mean Classifier, report the four-dimensional coordinates of the three means.

2) Calculate and report the 3x3 confusion matrix on the test set for both classifiers. The element C [i, j] of the confusion matrix corresponds to the number of instances that belong to class i but are assigned to class j. Report also the overall percentage of misclassified instances.

This homework is due **April 12th (Wednesday), 23:00**.

Your submission should include a short report of your findings and your source code.

You can use Python or Matlab. You are **NOT** allowed to use external libraries for classification (e.g., sklearn). You should code yourself the vectorial operations such as mean or distance calculations in terms basic addition, subtraction, etc.

Upload your submission as one **zip file** that includes a .**pdf** file of your report and a **.py** or a .**m** file for your code. Your executable code should be named NAME_SURNAME_HW2.py/m. If your code does not compile, you will receive 0 points from the coding part of the assignment!