Assignment 3

Statistical Learning – Fall 2018

Assignment Date: 1397/09/28

Due Date: 1397/10/15

Training

You have given two sets of data set for training. Train_DB1.csv and Train_DB2.csv.

Each row in these two files showing one sample: the first column is the label (ID) and the rest (F1 to F19) are the features associated with this point.

The data has been collected two times that resulted in Trai_DB1 and Train_DB2.

As you can each of the features take a value between -20. Just sometimes there are "0" values. These are the cases that we do not have a measurement for that particular feature.

The goal is that you build a model, that when we give it a vector ([F1 …. F19]) it predict the correct label (ID).

You can use any scheme that you want from very simple to very complex ….. .

Testing

There is a file called "Test_Data.csv" This files contains many rows that you should give each row to your model and predict the output. You can compare the

Note that:

1- You should not use the Test_Data.csv during the training phase.
2- You can check your prediction with the Test_label that I have provided (of course you cannot use them during training as well)
3- Note that "0"s in "Test_Datat" is not a correct value and it just show that you do not have reading there (like Train data)

What to submit

You should include:

1- A “complete” report. Make sure that to report the accuracy that you get with the “Test_Data” that I gave you. Also include the output file.
2- All your code.
3- A test code, that I can run it with a files as input (which will have a few rows like the Test_Data). After running the program should make a files as output (it should show the ID that you have predicted for each line of the input file, like Test_Label.csv)
4- A program that that gets “your output” file and another files (that I will use to test) with the same structure (but with the correct IDs), and produce the confusion matrix.

Grading

1- This assignment is mandatory for all (you can do it in group of two)

2- The first three groups who gets the lowest error will get extra mark.
First Group: if 1 person gets 1 point, if 2 persons each one take 0.75 point
Second Group: if 1 person gets .75 point, if 2 persons each one take 0.5 point
Third Group: if 1 person gets .5 point, if 2 persons each one take 0.35 point