**Exercise 4- Step two**

In the first exercise, we classified the point, “point” with values ​​of k = 1, and also k = 3. In the second part of the exercise, we will work with one file for learning, and another for classification. Before working with the large files, we will learn from the original file, myFile (including the first line), and classify the records in the second file, myFile\_test. What would be the tagging of the records in the second file with k = 3?

You can give values ​​in the second file by the sample code:

**with** open(**'mytrain.csv'**, **'r'**) **as** myCsvfile2:  
 lines = csv.reader(myCsvfile2)  
 dataWithHeader = list(lines)  
  
dataWithHeader[1][3] = **'F'**dataWithHeader[2][3] = **'M'  
  
with** open(**'mytest.csv '**, **'w'**, newline=**''**) **as** myCSVtest:  
 writer = csv.writer(myCSVtest)  
 writer.writerows(dataWithHeader)

Questions:

1. Classify the myFile\_test file with k = 3. Save the output to the same file

(myFile\_test) based on the code above.

1. Browse the mytrain and mytest file. Notice that there are about 200 total

records of which 100 exist in the mytrain file, and another 100 in the mytest file.

1. What accuracy do we get with k = 1? Save the tags as a file named

mytest1e.

1. What accuracy do we get with k = 7? Save the tags as a file named

mytest7e.

1. What accuracy did we get with k = 19? Save the tags as a file named

mytest19e.

d. Which value of k came out best?

e. Implement Manhattan distance in addition to the Euclidean distance.

Repeat the test on the mytest file with the same files and also with k =

1, k = 7, and k = 19. Save the file output to names: mytest1m, mytest7m

and mytest19m

f. Apply the Hamming distance in addition to the previous distances. Repeat the test on mytest file swiping with the same files and also k = 1, k = 7, and k = 19. Save the file output to names: mytest1h, mytest7h and mytest19h

g. Which combination of distance / value measure of k brings out the best

accuracy?

All 10 files must be submitted - one Python file and another 9 CSV files per model. It is recommended to submit everything as one zip file to the model.

Score:

40% - correct answers

40% - readable and clear code

20% - documentation