ECE 456/556 - Pattern Recognition
Project #1
Due by class on Tuesday 8 February 2022

This assignment will introduce you to your first formal classifier, the Nearest Neighbors classifier. Nearest Neighbors is a simple but effective classification procedure. It is generally used with much larger dataset sizes than we have, but should still illustrate the mechanism and potential of this classifier.

Implement the NN classifiers:.

a)	1-NN with the eval1dat.txt dataset
b)	1-NN with the eval2dat.txt dataset
c)	3-NN with the eval1dat.txt dataset
d)	3-NN with the eval2dat.txt dataset
e)	5-NN with the eval2dat.txt dataset
f)	1-NN with the eval2dat.txt dataset using a condensed training set. – graduate students only
g)	3-NN with the eval2dat.txt dataset using a condensed training set. – graduate students only

Use the data from train.dat as the training set and the data from eval1.dat or eval2.dat as test sets. Remember to normalize all the data by the RMS values to provide approximately equal weight to all features. For each test set report the classification results (i.e., % correctly classified and % error). Also print out a confusion matrix:

```
                    Decision
              a c e m n o r s x z   error I
         a   9         1                1
         c     8 2                      2
         e     3 6                 1     4
         m        10                     0
         n            6 3 1              4
Input    o     2       8                2
         r     3         7              3
         s                 9 1          1
         x                 10           0
         z     3 1             6        4
error II     0 3 10 1 0 4 1 0 1 1    21

    Classification Results: 79% correct 21% error.
```

Each entry represents the number of samples (an integer with actual class corresponding to the row and decision corresponding to the column). In the bottom right corner, print the total number of errors. If the number of test points is 100, then the error in the bottom right corner is also the percentage errors, otherwise use the number of test points to calculate and report this as a percentage. You have 5-7 cases described above. Each should produce 1 confusion matrix.

To turn in:

A brief report describing your results. This report should be written in a professional tone. The report should start with a brief description of the algorithm in your own words. If you have any implementation comments/observations include them here. Then describe the training set and the test sets verbally (N datapoints, consisting of d moment features, c classes consisting of the characters 'a', 'c', …). Then include the 5 or 7 confusion matrices as figures and a table with the 5 or 7 error percentages each with TEXT surrounding them indicating what combination of classifier/training you were using. Describe how you handle voting ties. If you had ties in the voting, report how often and which letters were involved in the voting. Graduate students: Report how many from each class were still in your dataset after pruning. i.e., 3a's, 5c's,… Your condensing method is stochastic, so average over 10 or more runs. All reports should end with a few comments comparing the results between the datasets, and between the classifiers (1-NN vs. 3-NN vs. 5-NN, regular vs. condensed training sets). You have been given 5-7 cases to run. When describing your results, ask yourself why each case was assigned and in your report, analyze the results from this viewpoint. There should be approximately 3-5 pages of typed text **plus** however much space is needed for the confusion matrices. Submit your source code with the main classification routines.

ECE 456/556 Pattern Recognition
Spring 2022
Project #1
Due by class on Tuesday 8 February 2022
Grading Rubric

Filtering

_____   1-KNN Classifier

_____   3-KNN Classifier

_____   5-KNN Classifier

_____   Condensed Nearest Neighbor Classifier (ECE 556 only)

_____   Discussion of datasets

_____   Discussion of method(s)

_____   Discussion of results – general

_____   Discussion of results – comparison of K

_____   Discussion of results – comparison of datasets

_____   Suitable depth of investigation

_____   Suitable insight exhibited in comments.

_____   Well formatted report (intro/body/conclusion, figures w/ labels)

_____   Good presentation (sentence structure, flow of thoughts,

            referencing of figures)

_____   Good grammar

_____   **Grade**