CSCE883 Homework1

Question1 KNN and Normalization/Standardization

Attached file Q1_train.csv Q1_test.csv are the training and testing dataset for a binary classification. The last columns of both files contain the labels.

Write a basic K-nearest neighbor classifier program to test the performance on the test dataset using the Euclidean distance. Try different K values and record the performances (accuracy) below.
Submit the code along with the tables below

K=	1	2	3	4	5	6	7	8	9
No. of correct predictions	461	461	461	461	461	461	461	461	461
Accuracy	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0

Now, add a normalization step to scale all attributes into the range of [0,1] and try your improved KNN algorithm on the test dataset and record the performance over K=1 to 9. Do the same for standardization.

Performance with scaled attributes via normalization

K=	1	2	3	4	5	6	7	8	9
No. of correct predictions (normalized)	460	460	460	460	460	460	460	460	460
Accuracy with normalization	0.9978	0.9978	0.9978	0.9978	0.9978	0.9978	0.9978	0.9978	0.9978
No. of correct predictions (standardized)	460	460	460	460	460	460	460	460	460
Accuracy with standardization	0.9978	0.9978	0.9978	0.9978	0.9978	0.9978	0.9978	0.9978	0.9978