HW 1
Q1 d)
Time [sec] (for loop): 1.8581528663635254
Time [sec] (np loop): 0.0010709762573242188

Q2 d)
Small petal length and petal width is likely setosa, medium petal length and pedal width is likely versicolor, and large petal length and petal width is likely virginica.

Small sepal width is likely setosa, medium sepal width is likely versicolor, and large sepal width is likely virginica. Large sepal length is likely setosa, and there is very little correlation between sepal length and versicolor or virginica

Q3 c)
For k = 1, Training accuracy = 100% and Testing accuracy = 90.3%

Q3 d)
K = 1
Training Acc: 1.0
Test Acc: 0.903

K = 5
Training Acc: 0.942
Test Acc: 0.925

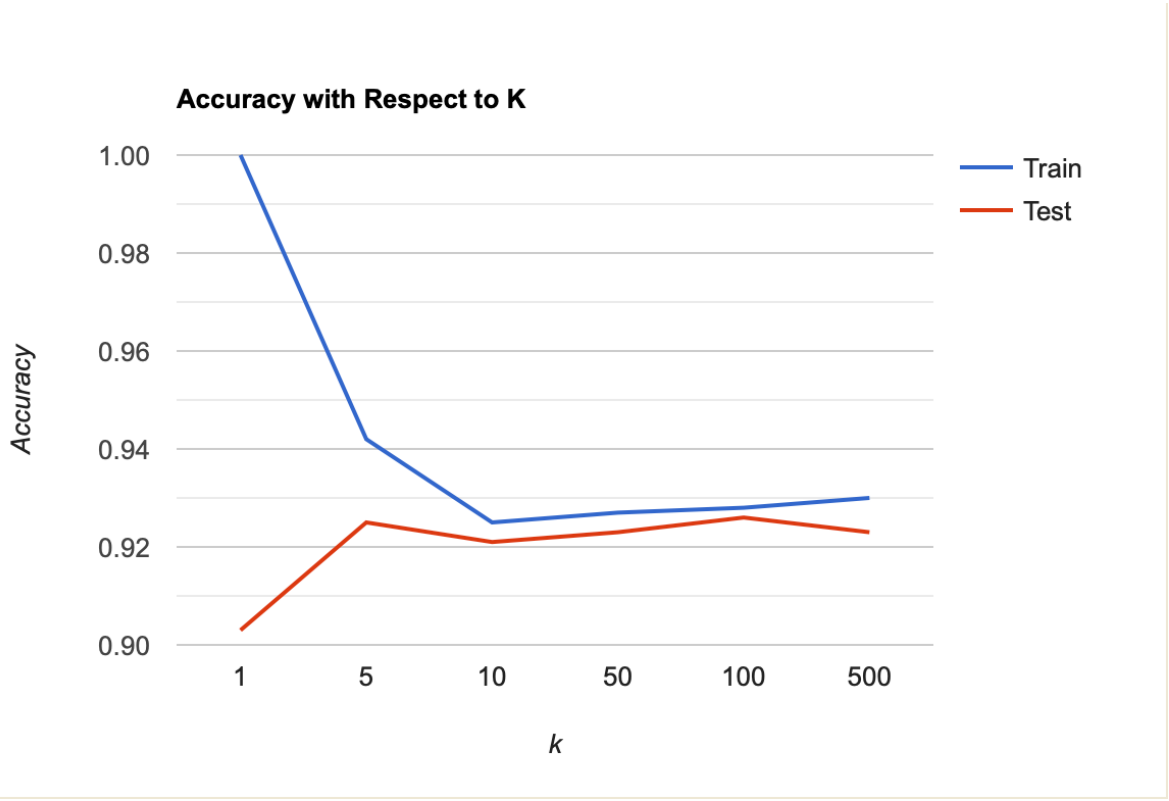
K = 10
Training Acc: 0.925
Test Acc: 0.921

K = 50
Training Acc: 0.927
Test Acc: 0.923

K = 100
Training Acc: 0.928
Test Acc: 0.926

K = 500
Training Acc: 0.93
Test Acc: 0.923

I selected these values of K since they cover most of the range of possible K values. Additionally, these values of K show the impact of choosing a small, medium, and large K.




Q3 e)
O(nd * k ln(k))
The computational complexity is linear with respect to n since we loop through each of the n element in the training set once. Additionally, we do one computation for each feature, d, so the number of features is also linear. Finally, the relationship between k, the number of neighbors, and the complexity is k log k because we sort the distance to select for the k most similar elements.

Q4 d)
K = 1
Test Acc (no-preprocessing): 0.8479166666666667
Test Acc (standard scale): 0.8791666666666667
Test Acc (min max scale): 0.8833333333333333
Test Acc (with irrelevant feature): 0.8604166666666667

K = 5
Test Acc (no-preprocessing): 0.8666666666666667
Test Acc (standard scale): 0.86875
Test Acc (min max scale): 0.8729166666666667
Test Acc (with irrelevant feature): 0.89375

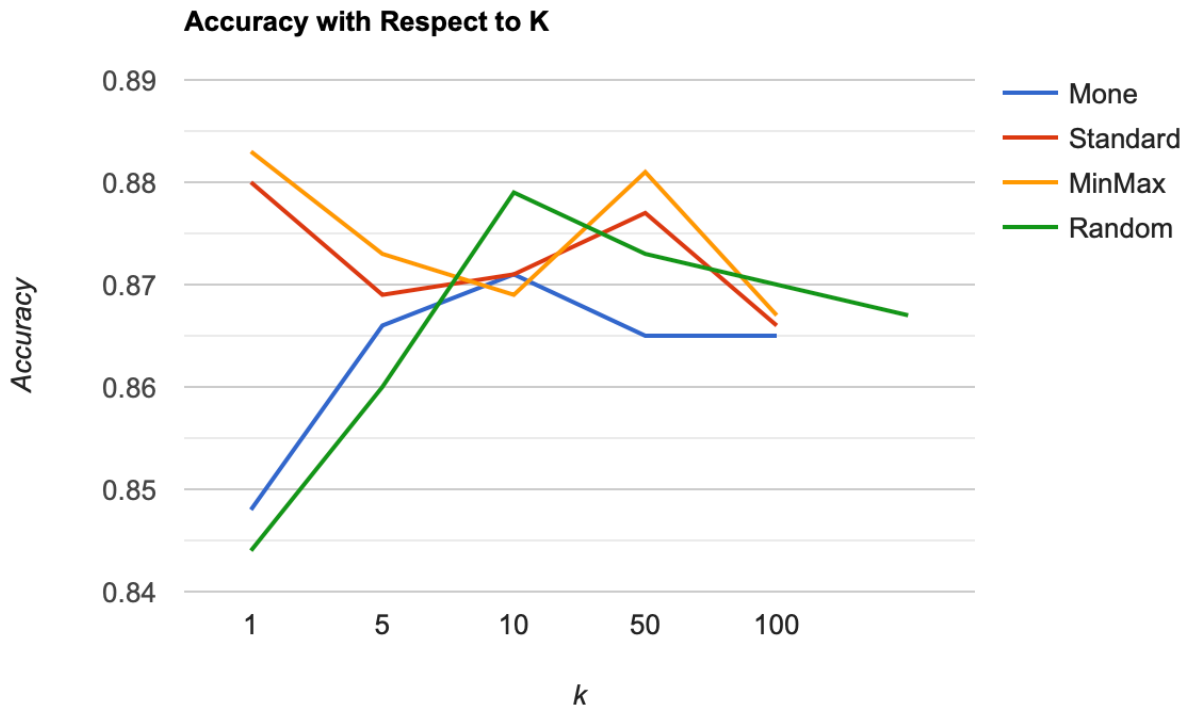
K = 10
Test Acc (no-preprocessing): 0.8708333333333333

Test Acc (standard scale): 0.8708333333333333
Test Acc (min max scale): 0.86875
Test Acc (with irrelevant feature): 0.88125

K = 50
Test Acc (no-preprocessing): 0.8645833333333334
Test Acc (standard scale): 0.8770833333333333
Test Acc (min max scale): 0.88125
Test Acc (with irrelevant feature): 0.8729166666666667

K = 100
Test Acc (no-preprocessing): 0.8645833333333334
Test Acc (standard scale): 0.8666666666666667
Test Acc (min max scale): 0.8666666666666667
Test Acc (with irrelevant feature): 0.86875

I selected these values of K since they cover most of the range of possible K values. Additionally, these values of K show the impact of choosing a small, medium, and large K.




The relative accuracies suggest no preprocessing techniques perform worse than standard scaling and minmax scaling. Additionally, MinMax scaling performed more accurately than Standard scaling for all values of K except for K = 10. Finally, the performance of the random features was random. In some cases, it was the least accurate, but in other cases, it was the most accurate. This implies the dataset varied significantly to irrelevant features in an unpredictable way.