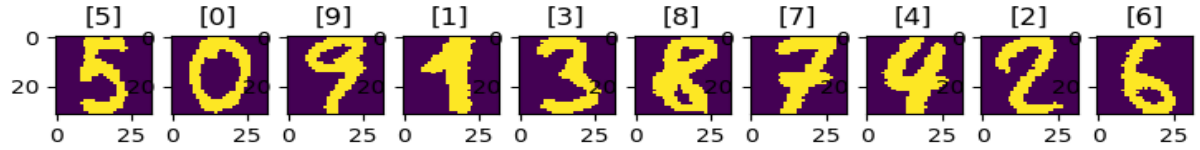
Programming Assignment -1

Q1:



Q2:

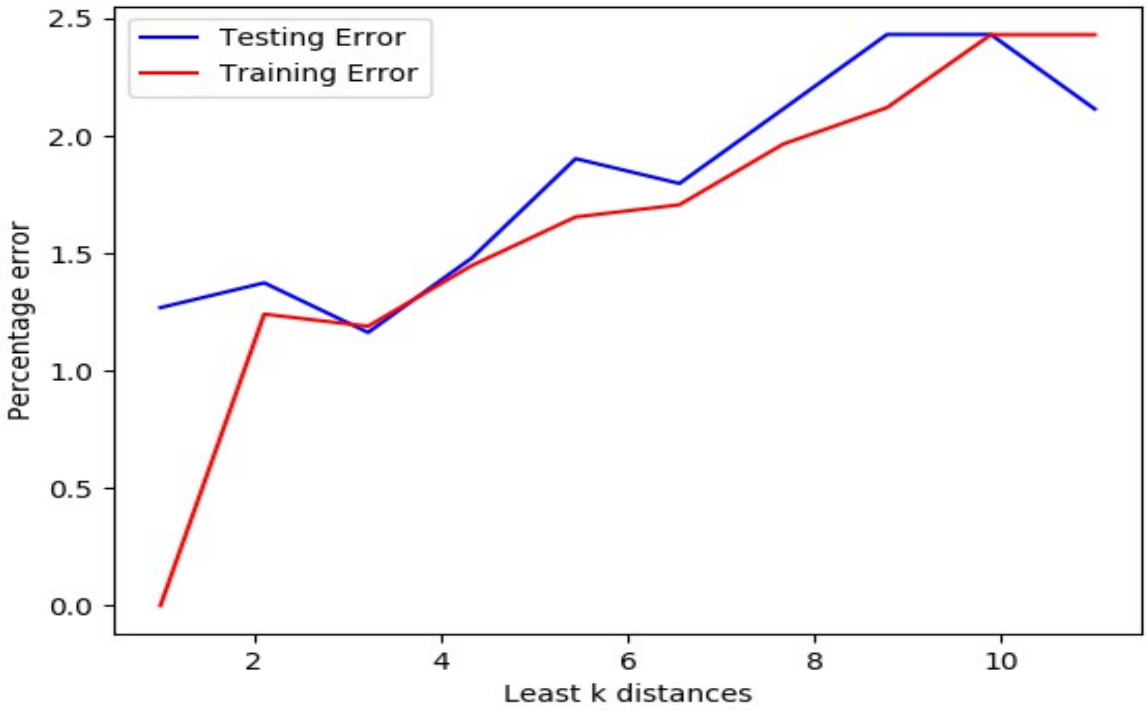
For Naive Bayes classifier using simple frequentist approach the following error rates are found:-

Training Error Rate: **5.89%**

Testing Error Rate: **7.08%**

Q3:

K	1	2	3	4	5	6	7	8	9	10
Training Error (%)	0	1.24	1.18	1.44	1.65	1.7	1.96	2.11	2.43	2.43
Testing Error (%)	1.26	1.37	1.16	1.47	1.9	1.79	2.11	2.43	2.43	2.11



Percentage error corresponding to k

Selection of best K: K = 3 works best for the given case, this is demonstrated in the above plot.

Trends and Difference in Testing and Training Error Rate plot:

One significant difference between error plots for training and testing data set is the zero error at K = 1 for the training data set. This is because, for K = 1, KNN compares the training image with itself and thus the minimum distance for one observation comes as 0 which is the least possible distance. For both the training and testing data-sets a minimum error is achieved at K=3.

After K = 2, both the training and testing error rates are approximately equal for different K values.

Model Averaging for Testing Error:

Testing Error = 1.70%

Comparison of KNN with NBC:

KNN	NBC
KNN has longer run-time	NBC has shorter run-time
KNN has better training and testing error rate	NBC has poorer testing and testing error rate
KNN has no assumptions associated with it	In NBC we assume every feature is independent of one another.
One model can take care of all types of values	Discrete and continuous distributions have to be modeled differently for NBC.

Ways to improve NBC:

1. We have introduced Laplace smoothing in Likelihood. This resulted in introduction of a generalization in the model and helps to take care of unforeseen circumstances.
2. In Naive Bayes we assume that the features are independent of each other for a given class which might not be the actual scenario. This can be changed if we have a prior knowledge about the correlation of the features.
3. We can use a Bayesian interpretation of NBC instead of the frequentist method used here. This will also improve the generalization error of NBC .

Q4:

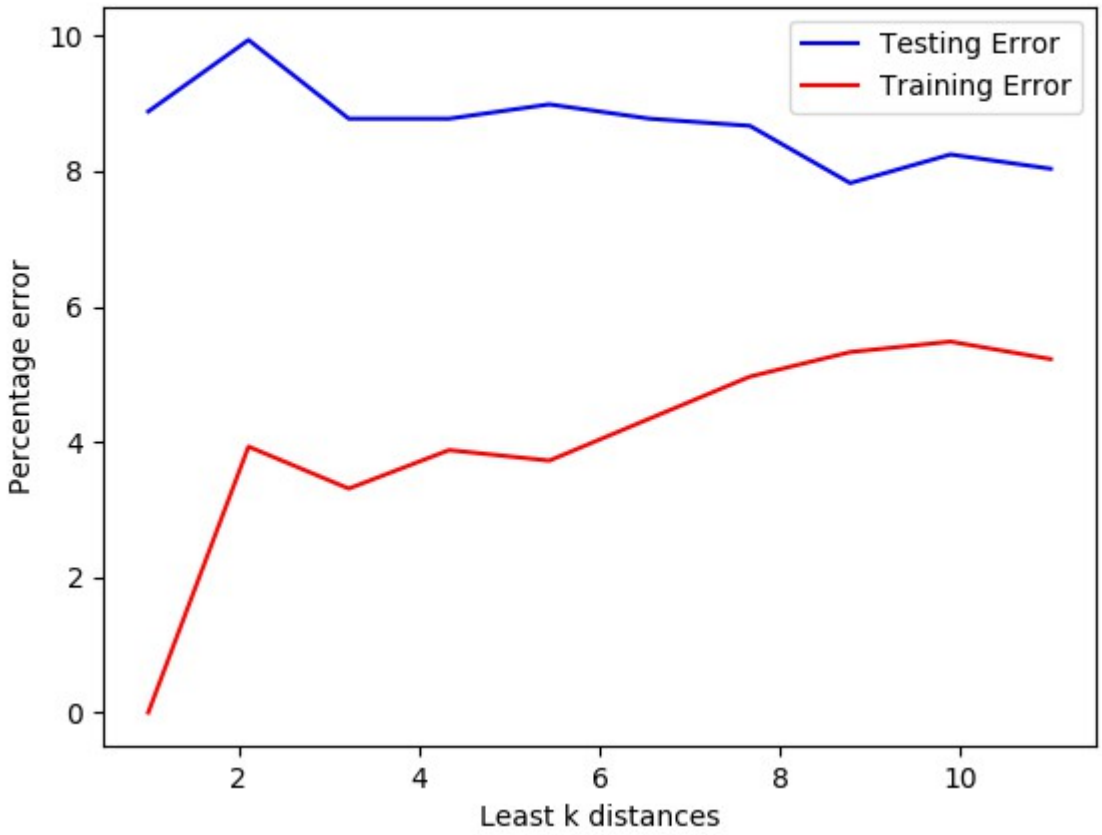
NBC:

On application of PCA we get the following results for a range of 1 – 9 components derived from PCA.

No. of components	1	2	3	4	5	6	7	8	9
Training Error (%)	60	40	24	20	15	12	10	8	8
Testing Error (%)	65	45	27	24	19	16	23	32	58

KNN:

K	1	2	3	4	5	6	7	8	9	10
Training Error (%)	0	3.92	3.3	3.87	3.72	4.34	4.96	5.32	5.48	5.22
Testing Error (%)	8.87	9.93	8.77	8.77	8.98	8.77	8.66	7.82	8.24	8.03



The main advantage of doing PCA is that it reduces the number of features thus reduces the complexity of the algorithm significantly.

As we can see from the graphs above, accuracy for both NBC and KNN has gone down which is a major drawback for PCA.