Assignment-1 Report

Algorithm:

We consider 7 fold cross validation in which we split the data into 7 parts out of which 6 parts will be training set and the other will be testing set. we do this 3 times so that all the three will be testing set once.to find the best k value we consider values from 1 to 10 and for p we do it from 1 to 5.

minkowski distances are calculated from one point in testing set to all the other points of training set to predict the class of the given point based on the minimum distance for each k. now we compare the obtained result to the original value and find errors. Error is calculated as ratio of wrong predicted classes to total number of classes for each value of k and p. We finally get the best k and p values based on the minimum error

Result:

1.in 7 cross validation for each case correct predicted for each k and p are in the range 24 to 30 out of 30 Which means accuracy is 88.571 percent. Error range is 4.33 to 20 percent. Based on the result there will be 3 to 4 sets of k values will exist which will with respect to random shuffling or der of the training set and testing set.

2.

We have taken value of R as 20 (20-fold cross validation).The same algorithm is followed as mentioned above.There are totally 10 classes.

We have taken k - 7 , p - 2

The accuracy is around 80 percent. Error rate is also less .