NAME: ARAVIND SELVAN **ANDREW ID: ASELVAN**

Question 2.2:

The following summarizes the results,

K	Confidence Level	Accuracy	Low Interval	High Interval
10	95 %	0.7400	0.6540	0.8260
10	99 %	0.7400	0.6270	0.8530
2	95 %	0.7278	0.6628	0.7928
2	99 %	0.7278	0.6423	0.8132

Table (1)

From the table (1), it is observed that the confidence interval widens with the increase in partition value as there are less data to predict on which the error is calculated.

Question 2.3:

The following table below shows the testing and train error rate for different C values,

C	**Train Error**	**Test Error**
0	0.485	0.485
0.1	0.255	0.25
0.3	0.25167	0.245
0.5	0.25667	0.25
1	0.26333	0.25
2	0.26	0.255
5	0.26	0.255
8	0.25833	0.255
10	0.25833	0.255

Table (2)

The below plot [Figure (1)] shows the testing and train error rate (in y-axis) for different C values (in x-axis),

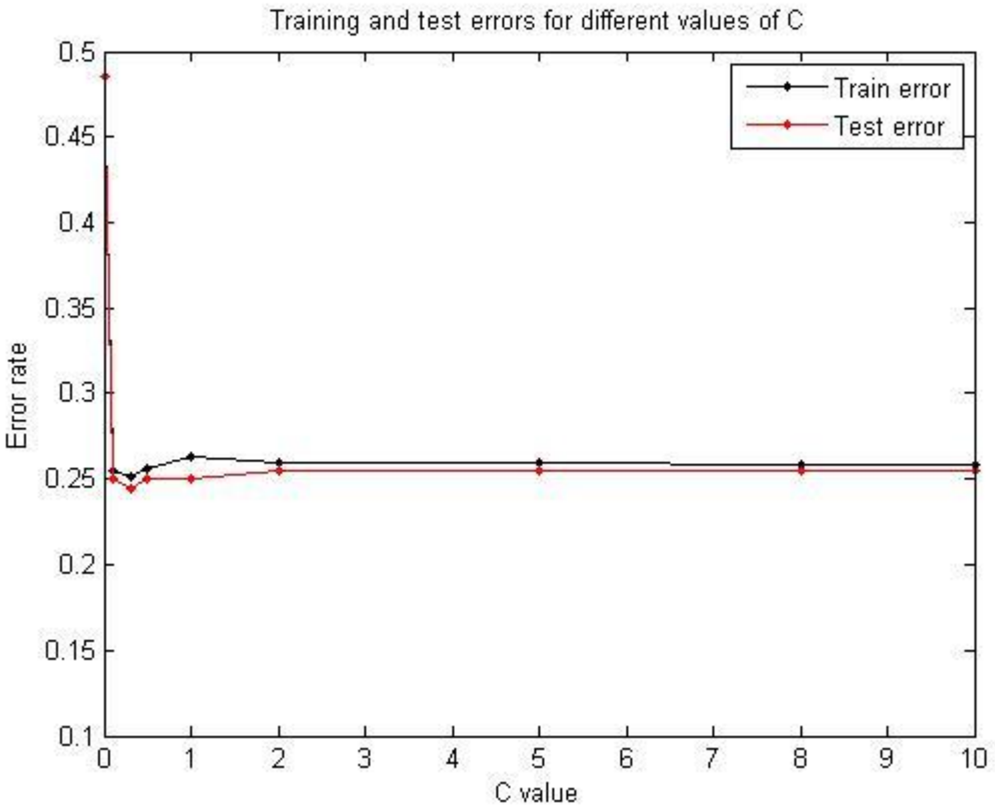

Figure (1)

From the above graph, the **C value = 0.3** gives the best values for training and testing error rate (0.25167 and 0.245 respectively); and therefore this value should be used to train the SVM.

Question 2.4:

The table (3) and table (4) shows the accuracy and p-value under one-tailed test and two-tailed test with k=10.

Type	Accuracy
Logistic Regression	0.9814
Neural Networks	0.9757

Table (3)

t-test:

Type	p-value
one-tail test	0.42
two-tail test	0.59

Table (4)

From the above two tables, the null hypothesis cannot be rejected as the p-value is high for both the one-tail test and two tail test. Thereby, it can be concluded that none of the models works convincingly better than the other.