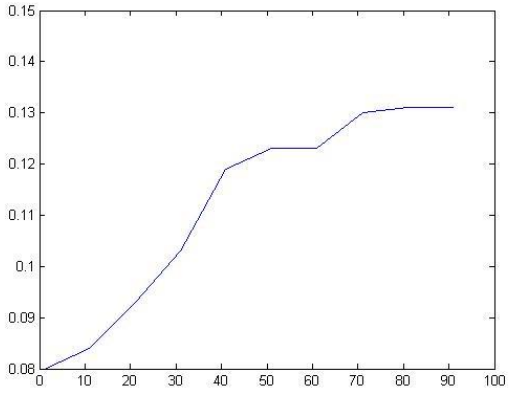
CS365 ASSIGNMENT 2

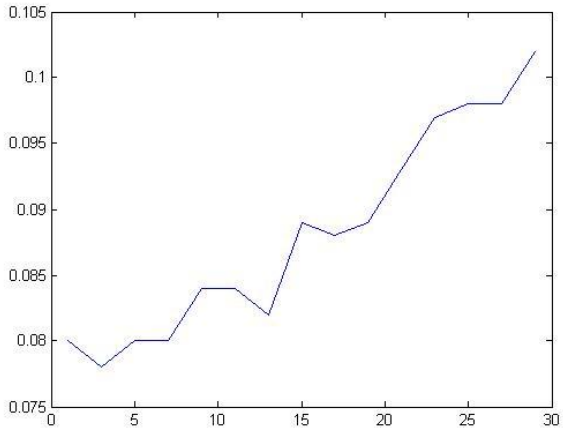
A. K-NN BASED CLASSIFIER

KNN Classifier	Properties
Training Dataset Size	10,000
Test Dataset Size	1,000
Number of neighbors	1:10:100
Number of neighbors	1:2:30

First a broad estimate of performance of knn was checked by varying number of neighbors from 1 to hundred in increments of 10. This is the resulting plot-



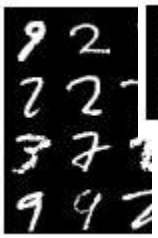
The x axis gives the number of nearest neighbors considered while the y axis gives the error. The error of 0.1 means 10% error.



Plot of nearest neighbors with finer grid between 0 and 30.

RESULT- The value of error increases as the value of k increases. The final value of k was chosen to be 5 and the minimum error was around 8%.

OBSERVATIONS- The value of optimal k also depends on the size of dataset that we have chosen for training as the same value for k would consider larger volume in smaller dataset as compared to larger dataset. Analyzed the error produced by knn for each group of digits-

Actual Number	0	1	2	3	4	5	6	7	8	9
Misclassified Numbers										

As it could be seen from table it performs badly in case classifying 4- 9 and 7-1. It is also not able to classify correctly if there is a lot difference in the thickness of digits.