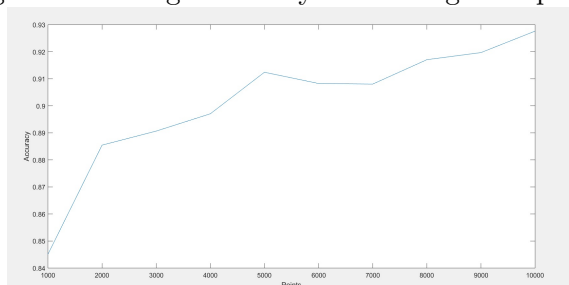


CMSC 678 – Fall 2016 — Homework 1

Q.7. Implement kNN in matlab for handwritten digit classification and submit all codes and plots:

- (a) For $k = 1$, change the number of training data points (30 to 10,000) to see the change in performance. Plot the average accuracy for 10 different dataset sizes. You may use command `logspace` in matlab. In the plot, x-axis is for the number of training data and y-axis is for the accuracy. Description of the graph : This is a points vs accuracy

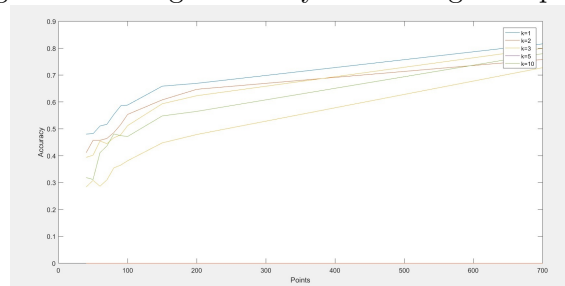
Figure 1: Average accuracy vs training data points.



graph in which the accuracies for data sizes varying from 30-1000 have been plotted. Depending on the data size, the accuracy varies.

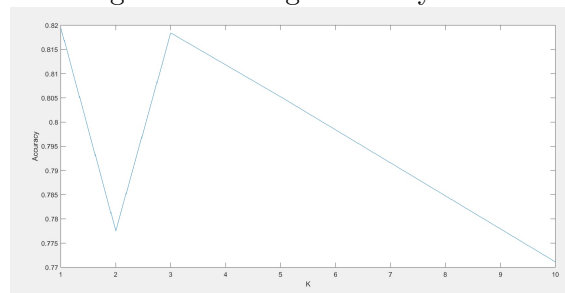
- (b) Show the effect of k on the accuracy. Make a plot similar to the above one with multiple colored curves on the top of each other (each for a particular k in $[1\ 2\ 3\ 5\ 10]$.) You may use command `legend` in matlab to name different colors. Description of the graph : The five graphs represent the accuracies for data sizes between 30-1000 when the k is respectively 1,2,3,5,10

Figure 2: Average accuracy vs training data points.



- (c) Choose the best k for 2,000 total training data by splitting the training data into two halves (the first for training and the second for validation). You may plot the average accuracy wrt k for this. Note that in this part, you should not use the test data. You may search for k in this list: [1 2 3 5 10].

Figure 3: Average accuracy Vs K



Description of the graph : The best value of K is when $K=1$. This accuracy graph is plotted when I consider just 2000 of the training data and out of that, the last 1000 is the test data.