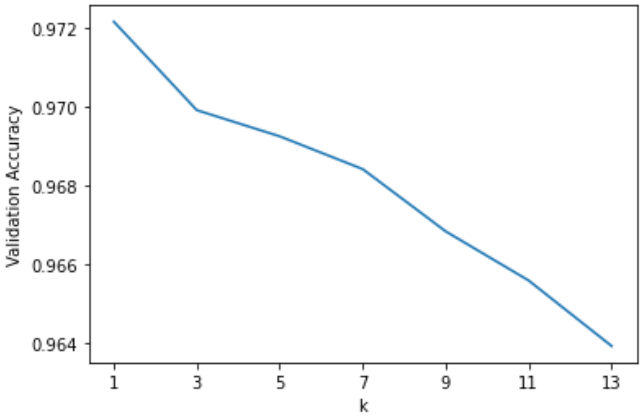
HW1- CS412 Report

In this report, the work has been made is to establish an optimal nearest neighbor number to be trained on the MNIST dataset.

The dataset includes 60000 training and 10000 test samples. Total training data has been splitted by 80% to be used as training and 20% to be used as validation data.

After shuffling the whole training data, I have taken the first 48000 rows to be remain as training, and the remaining 12000 rows to be used as validation set. Then, I trained the training data for different k values (1,3,5….13), stored the accurracy level and plotted in the figure as shown in (fig1) . As it can be seen easily, the best value for k is 1 that gives 0.972 accurracy level on validation set. That is why I have chosen k value as 1.



(fig1)

With the knowledge of best k value, I have trained the training and validation sets together to predict the test set. I have obtained the best results on the validation set with the KNN approach using a value of “1” for “k” parameter. The result of this model on the dataset is 96.91% accurracy.

The link for the notebook:
https://colab.research.google.com/drive/1q6ctShxcomgpasSzfXF61ptPB10_F-tm?usp=sharing