WRITE - UP

For this Assignment, the train\_data was first loaded using numpy and its dimension and the dimension of each image were checked. Then, by using a for loop and reshape function, the image of each digit was printed. Next, by setting the range of x-axis from 0 to 9 and range of y-axis from 0 to 0.12, a normalized histogram was plotted with the use of the weight parameter in plt.hist function. Then, by finding the index of each digit in the dataset, the distance of each digit in dig\_list with every other digit in the dataset was calculated, and from the distance list, the datapoint with smallest distance was found and displayed along with the sample datapoint. Then, by finding the zeros and ones data in the dataset, the zero to zero, one to one, and zero to one distance were calculted and resulted in genuine and impostor distances, and a histogram about them was then plotted. Next, with the use of a for loop, the ROC Curve was plotted, and the two different error rates were shown. Then, a KNN was implemented, the distance between datapoint and every other data in the dataset was calculated, and the k smallest were chosen, next by equally voting, the final decision was made, and this was done by using 3 for loops. Finally, by separating the training\_data to train and label, and 3-fold cross validation was performed.

The accuracy along with the heatmap for the 3 trials are as follows:

Timeline

Description automatically generated

With each datapoint in the testing set, KNN was used to perform the prediction and the accuracy was calculted. From the heatmap that got plotted, we could see that the digit 5 is particularly tricky to classify.