Mnist Dataset modelling using KNN

-Shawn John

Approach:

First the dataset was loaded and a 75-25 train-test data split was carried out. Also 10% of the training data was used for validation. Next up the optimal value of k was chosen (in this case 3 for the given dataset) and the model was trained. The accuracy was determined and random 5 digits were chosen to test the model’s prediction.

1. How accurate is this method?

For a simple ML algo like KNN, the accuracy obtained in this model (99.26%) is pretty decent using 3 nearest neighbours to find the common class.

1. What metric did you use for distance?

The distance metric used was Euclidean distance.

1. Any ideas on improving accuracy?

One way to improve the accuracy of this model is to find the accuracies for different values of k given the total samples in the dataset. For the dataset this model was tried on, 3 was an optimal number. By choosing an odd number for the value of k we ensure that no stalemates occur.

Another way is to obtain a larger dataset or better quality dataset (however this is not always possible). Usually fine-tuning the hyperparameters will give you a better accuracy.

Output:











