ELL784 Machine Learning Assignment 3

Kashish Srivastava (2024AIB2289)

4th November 2024

1. Using Artificial Neural Network for Classification Task

Use the MNIST dataset, which contains images of handwritten digits (0-9) across 10 classes. Select 100 images from each class in the training dataset to form a new training set (1000 images total). Keep the original testing dataset unchanged.

a. **Multiclass Classification**:

We trained a standard Neural Network on the original training dataset to classify images across all 10 classes.

Neural Network is as follows:

Hidden Layer:

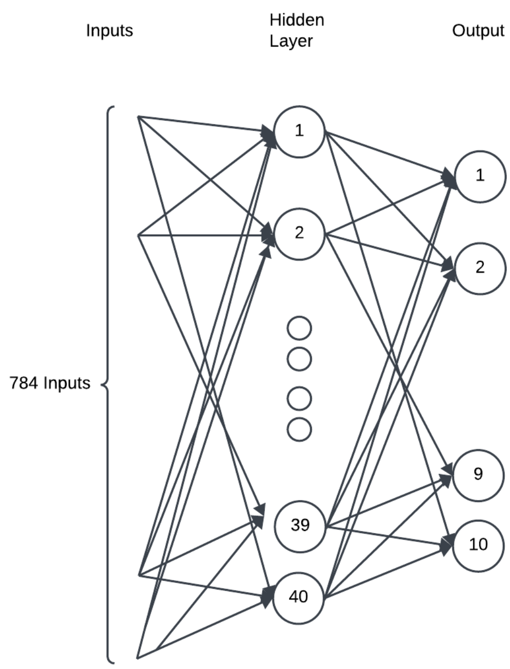
Consists of 40 neurons accepting 200 inputs from previous hidden layer.

Activation Function: Sigmoid

Output Layer :

Consists of 10 neurons corresponding to each possible output (0-9)

Activation Function: SoftMax

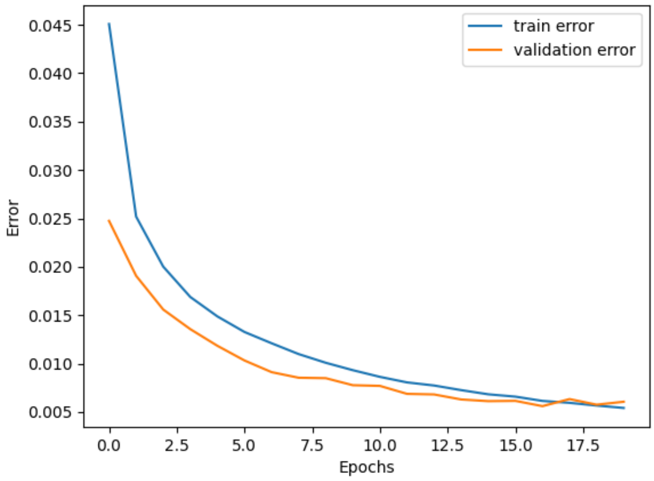


Artificial Neural Network Architecture

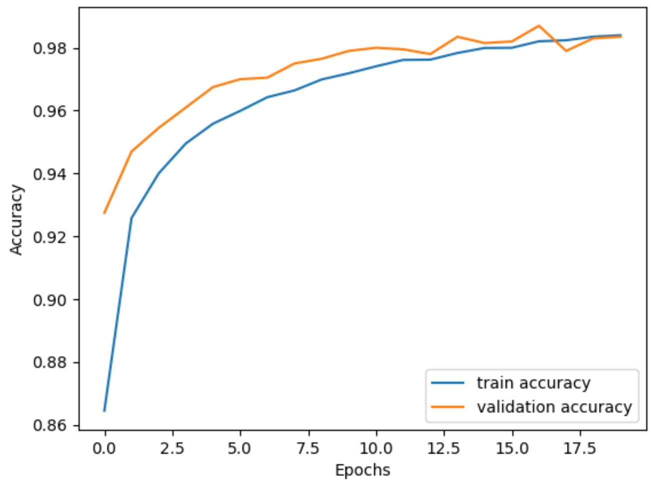
The ANN was trained on the dataset for **20 epochs** with l**earning rate = 0.05.**

Loss function used for the same is **Categorical Cross Entropy**.

Observations from the training the model are as follows:



Training/Testing Error vs Epochs



Training/Testing Accuracy vs Epochs

Metric	Value
Training Accuracy	0.9837
Training Loss	0.0056
Validation Accuracy	0.978
Validation Loss	0.0068
Test Accuracy	0.97435

2. **Binary Classification on pairs:**
   Using the 1000 selected training images, we created every possible pair of images. Labelled each pair to indicate whether both images belong to the same class (1) or different classes (0).

   Architecture for binary classifier:

   a. Inputs:
      Instead of using a single input image for classification we use 2 images and check for similarity between the images.

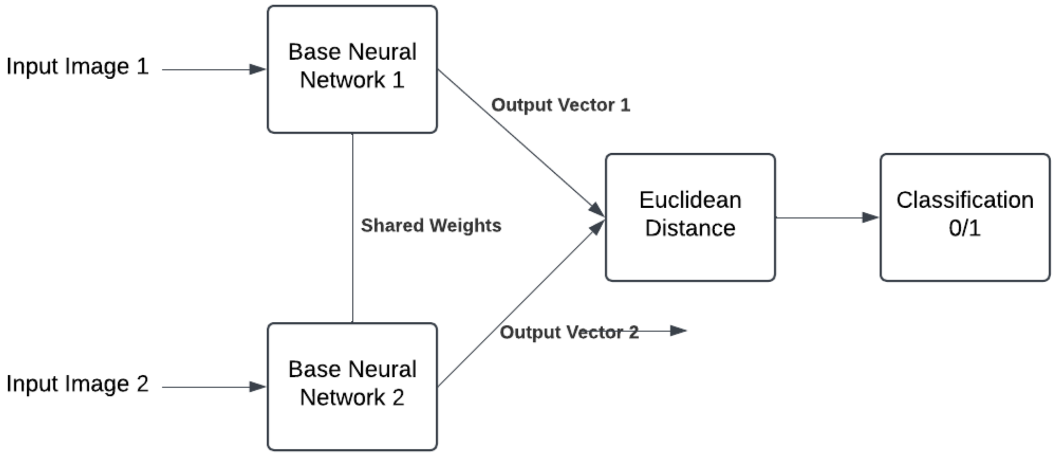
   b. Base Neural Networks:
      Base Neural Network refers to the Artificial Neural Network we trained in the 1st part of the assignment. We use the same network to obtain the output vector which are used in the next part of the classifier network.

   c. Similarity Criterion:
      To check for similarity between the pair of images (or rather the similarity between the output vectors calculated by the base neural network) we are using the Euclidean distance between the outputs of the sister networks.

$$\sqrt{\{G_W(X_1) - G_W(X_2)\}^2}$$

Euclidean Distance formula

   d. Classification:
      If the Euclidean distance between the pair is small enough, we will classify the pair as 1 meaning that both the images belong to the same class. Else, we will classify the input pair as 0, meaning that the pair belong to different classes.



Architecture of Binary Classifier

Base Neural Networks (Sister Networks) were trained using the same hyperparameters and original training/validation set as were used in the former part of the assignment.

New dataset had to created for training the binary classifier network by taking 100 samples from each class and made each possible pair that exists.

New dataset was forwarded to base neural networks and reshaped as 2 different images and performed forward pass on input.

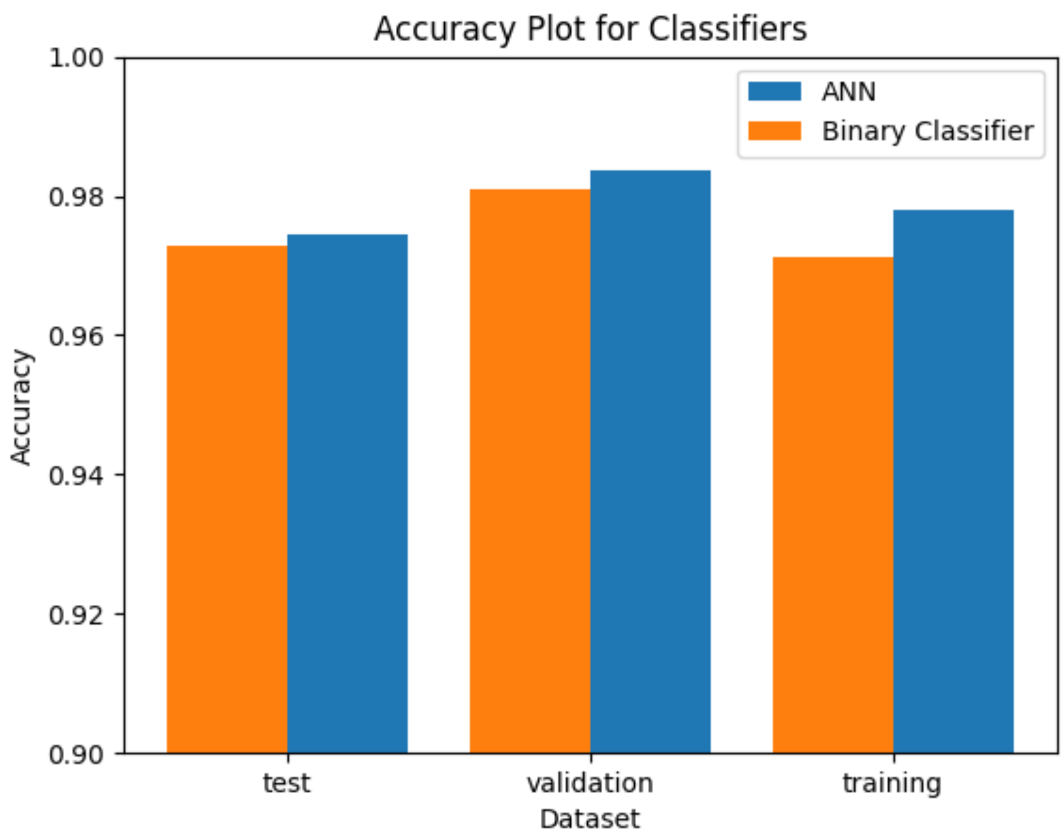
Feature vectors output from the base networks were used to calculate Euclidean distance between the outputs. If the Euclidean distance between feature vector was observed to be less than 1.0 then the classifier would give 1 as output, else 0.

Based on the above criterion the following results were obtained:

Metric	**Value**
Training Accuracy	0.9837
Validation Accuracy	0.981
Test Accuracy	0.972714285

3. Comparing results of both scenarios

For testing, we selected k random images from each class, forming 10*k pairs with each test sample.
The class of each testing sample was predicted based on the class it most frequently matches in the model's binary classification output.

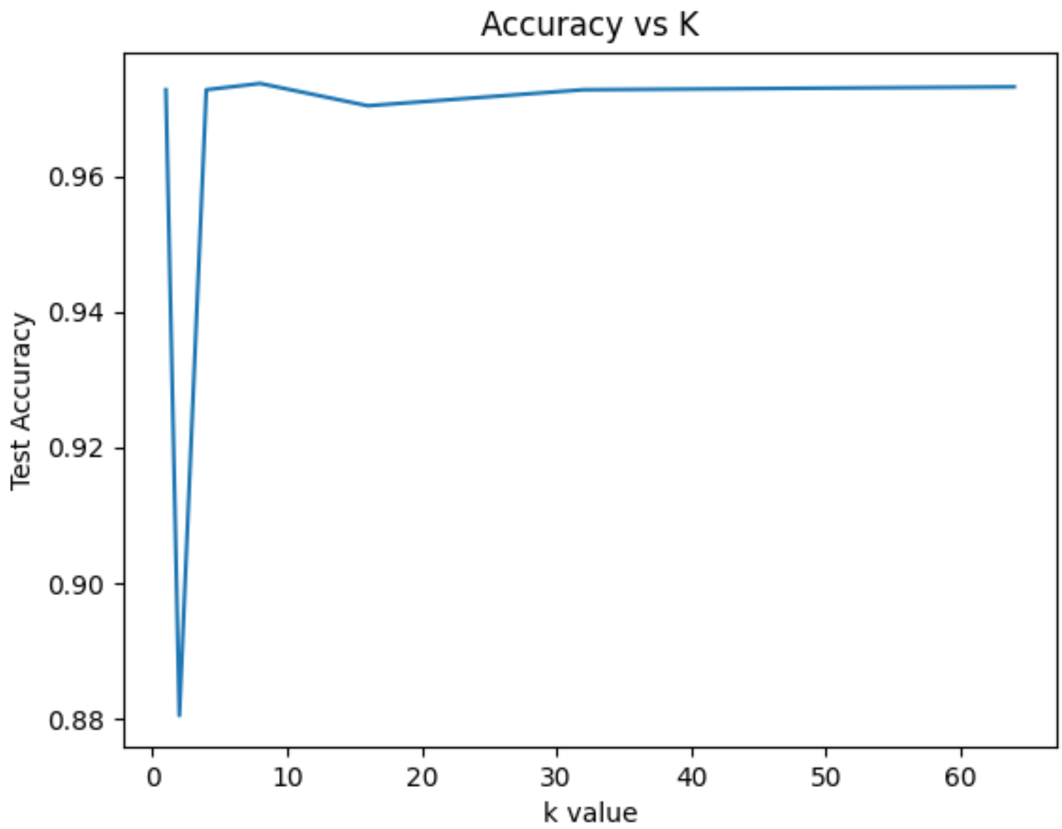


Accuracy Plots of classifiers

4. Fine Tuning k parameter

We used exponential steps to find a balanced k . Test values of k in the sequence 1,2,4,8,16,32 and 64.

We increased k exponentially and measured accuracy until accuracy gains are minimal



Accuracy vs K value

K value of 8 was found to be optimal for this particular task as increasing k value gave minimal increment in accuracy but very high increase in computation costs.

5. Findings and Insights

Performance:
The multiclass classification model generally performed better than binary classification even though the difference between the two classifiers was minimal

Speed:
Binary classifier approach was more computationally demanding and needed more than 2 times the time it took for multiclass classification as multiclass classification model was used 2 times in the binary classifier model.

Similarity focus:
Instead of directly learning digit patterns the binary classifier focuses on similarity which might generalize better for datasets where class labels are less distinct.