AIP701 Machine Learning Assignment 4

Kashish Srivastava (2024AIB2289)

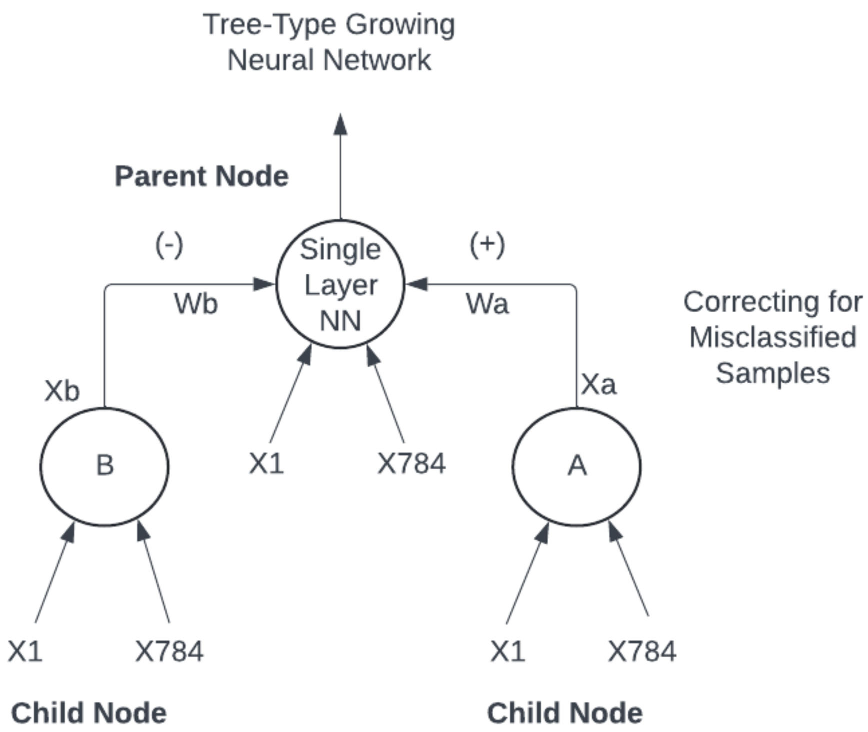
6th November 2024

1. **Using Growing Tree 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).

   a. **Multiclass Classification**:

      The structure of network grows incrementally, with new child node being added to address misclassified sample by parent node. The network as a whole is used to classify samples as 0 or not 0 (i.e. 0 vs all classifier)



Growing Tree Neural Network Architecture

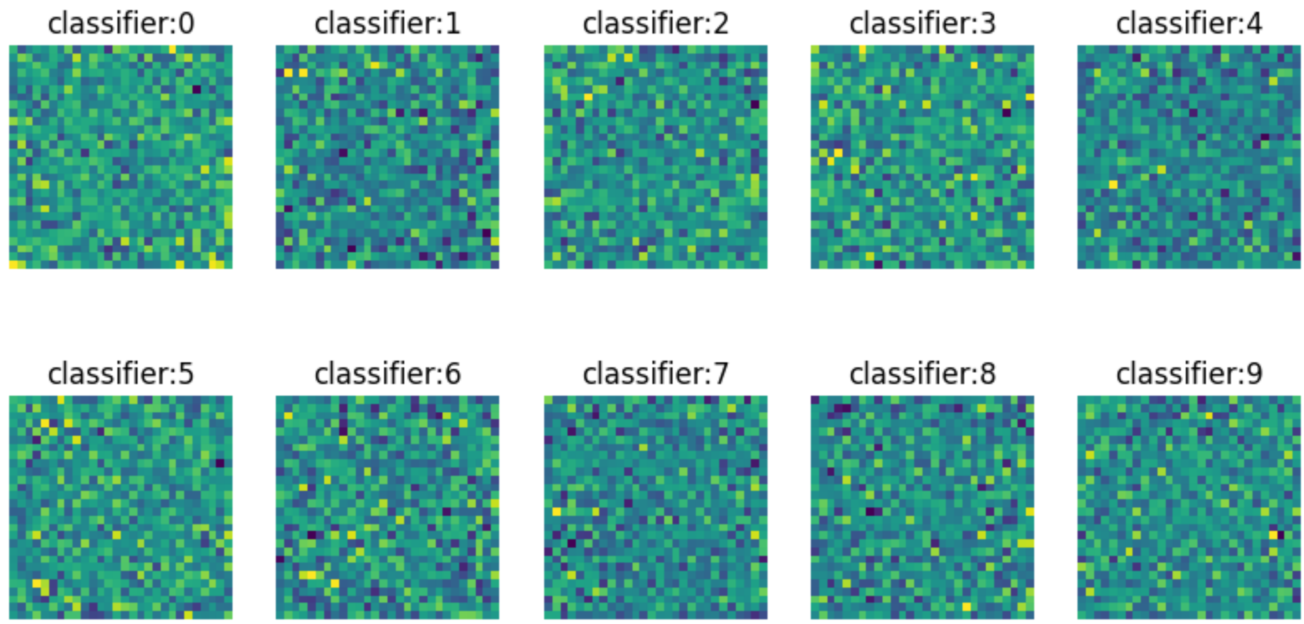
The Growing Tree NN was trained on the dataset for **20 epochs** with l**earning rate = 0.08.**

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

One vs All classifier testing accuracy values observed are as follows:

One Vs All Classifier (class)	Testing Accuracy	Testing Loss
0	0.9818571428571429	0.016873172240245243
1	0.9825	0.01661679576040701
2	0.9473571428571429	0.05114363895819381
3	0.8228571428571428	0.17239904157954486
4	0.9722142857142857	0.027064803236303756
5	0.8672142857142857	0.1299379313578792
6	0.9643571428571428	0.03454770202840275
7	0.9521428571428572	0.04599217537619044
8	0.8092857142857143	0.18749018839895087
9	0.8925714285714286	0.10446397284983545

Now, for multiclass classification we will pass the sample data to be used for prediction is passed through each one vs all classifier . The classifier with the most activation (argmax is used) will be the label or the prediction class of the sample.



Denormalized Weights of root note of each one vs all classifier

Overall Multiclass Classifier

Metric	Value
Training Accuracy	0.814411
Validation Accuracy	0.8136
Test Accuracy	0.8083571

Multiclass Growing Tree Neural Network observations

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.

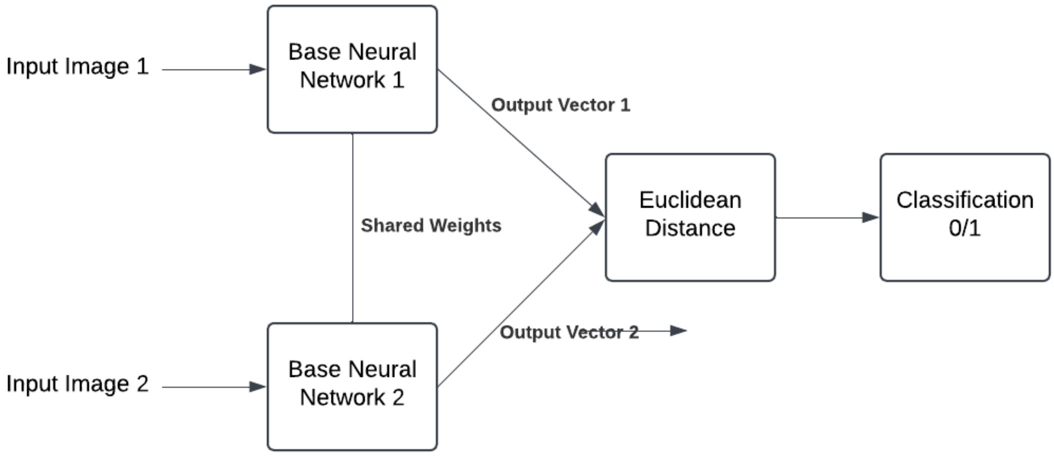
   b. Base Neural Networks:
      Base Neural Network refers to the Growing Tree Neural Network we trained in the 1st part of the assignment. We use the same network to obtain the output vector which are used in the next part of the classifier network.

   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 be 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.

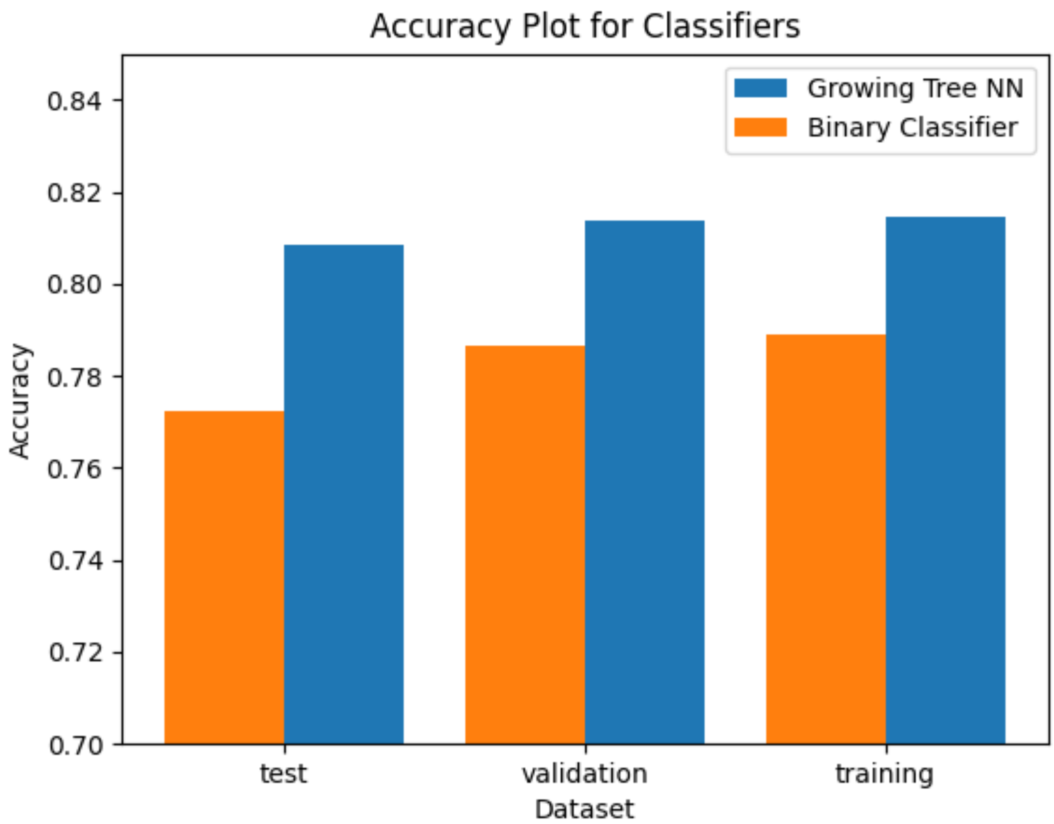
Feature vectors output from the base networks were used to calculate Euclidean distance between the outputs. If the Euclidean distance between feature vector was observed to be less than 1.0 then the classifier would give 1 as output, else 0.

Based on the above criterion the following results were obtained:

Metric	**Value**
Training Accuracy	0.7890196
Validation Accuracy	0.7866
Test Accuracy	0.7723571

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.

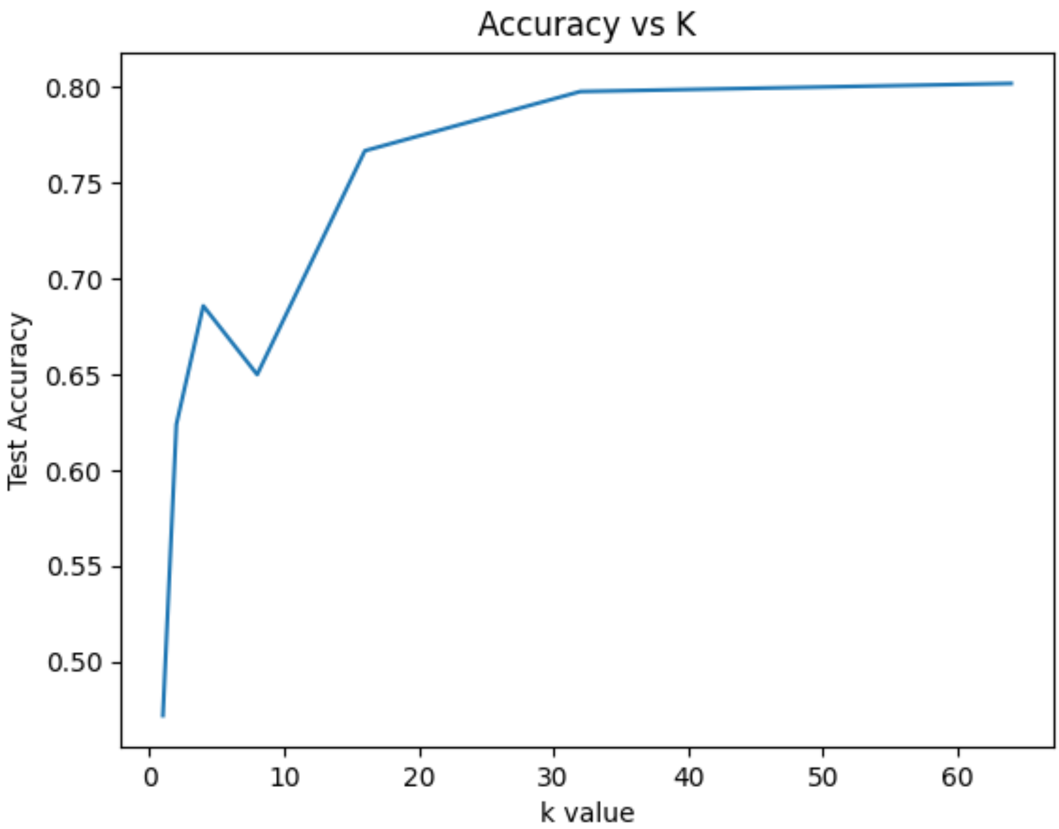


Accuracy Plots of classifiers

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 20 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.