Assignment 6

ATHARVA DATE - B22AI045

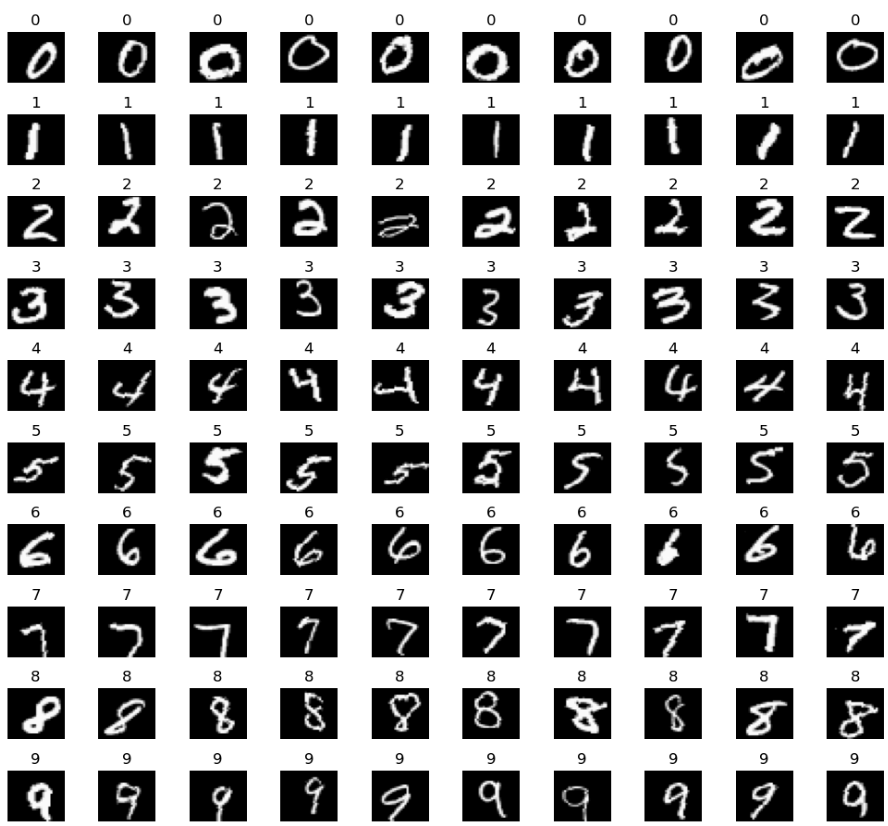
Dataset Description:

The MNIST dataset is a collection of handwritten digits ranging from 0 to 9. It consists of 60,000 training images and 10,000 testing images, each of size 28x28 pixels. The task is to classify these images into their respective digit classes.

Tasks:

Task 1: Data Visualization

Images from each class in the training dataset were plotted to visualize the dataset's contents. Each class is represented by five sample images.



Task 2: Multi-Layer Perceptron (MLP) Model

A 3-layer MLP model was implemented using PyTorch. The model consists of two hidden layers with ReLU activation functions.

```
model = MLP(input_size = 784, hidden_size1= 1024, hidden_size2= 512, output_size = 10)
```

Task 3: Model Training

The MLP model was trained for 5 epochs using the Adam optimizer and CrossEntropyLoss as the loss function. The training loop includes evaluation on the validation set after each epoch.

Only Linear layers without any activation function:

```
Epoch 1/5, Train Loss: 8.3577, Train Acc: 0.6046, Val Loss: 1.2101, Val Acc: 0.6198
Epoch 2/5, Train Loss: 1.0533, Train Acc: 0.6732, Val Loss: 0.9654, Val Acc: 0.6923
Epoch 3/5, Train Loss: 1.0554, Train Acc: 0.6681, Val Loss: 1.0675, Val Acc: 0.6586
Epoch 4/5, Train Loss: 1.0926, Train Acc: 0.6556, Val Loss: 1.0557, Val Acc: 0.6640
Epoch 5/5, Train Loss: 386.6724, Train Acc: 0.5640, Val Loss: 14.2534, Val Acc: 0.5897
Training is completed!
```

With ReLU:

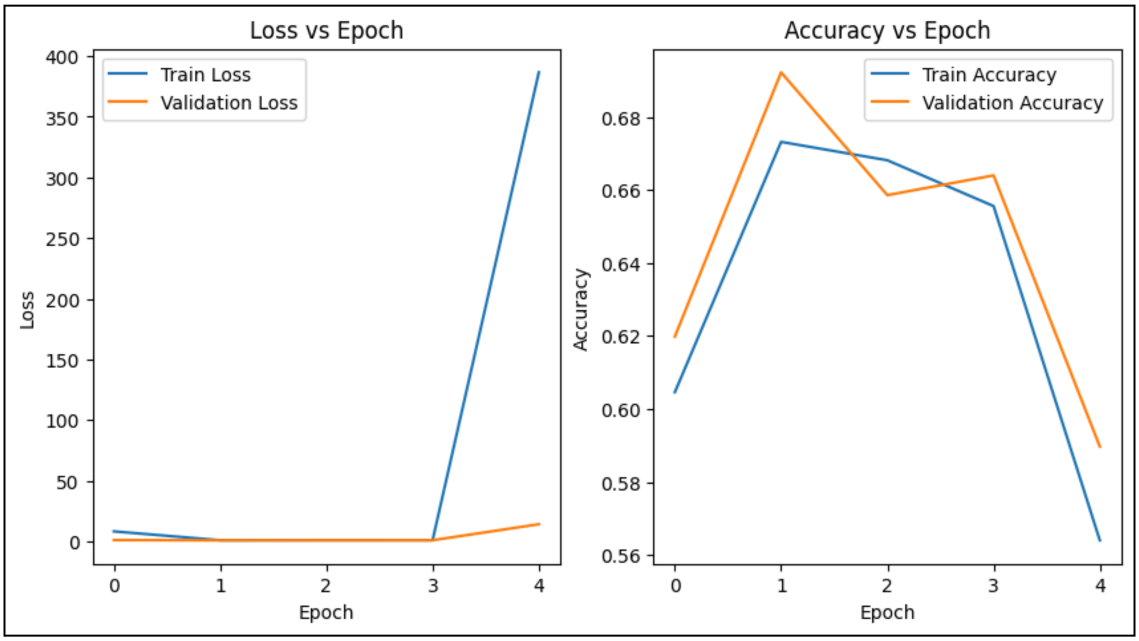
```
Epoch 1/5, Train Loss: 0.9035, Train Acc: 0.7259, Val Loss: 0.6719, Val Acc: 0.7940
Epoch 2/5, Train Loss: 0.5740, Train Acc: 0.8249, Val Loss: 0.5933, Val Acc: 0.8179
Epoch 3/5, Train Loss: 0.5309, Train Acc: 0.8396, Val Loss: 0.5315, Val Acc: 0.8438
Epoch 4/5, Train Loss: 0.5114, Train Acc: 0.8476, Val Loss: 0.5717, Val Acc: 0.8276
Epoch 5/5, Train Loss: 0.4939, Train Acc: 0.8548, Val Loss: 0.5303, Val Acc: 0.8468
Training is completed!
```

The accuracy of a neural network with only linear layers and no ReLU activation may decrease while the loss increases over epochs for several reasons. First, without non-linearity, the network struggles to grasp complex patterns in the data, resulting in poorer performance. Additionally, the absence of activation functions can hinder effective weight updates during backpropagation, leading to the vanishing gradient problem where gradients become extremely small as they move through the layers. Moreover, the simplicity of linear transformations limits the model's understanding of the data's complexity. Overfitting can also occur when the model memorizes the training data too well without considering nonlinear regularization, impacting its ability to generalize to unseen data. Overall, the lack of non-linear activation functions can impede the network's learning process and predictive accuracy over time.

Task 4: Visualization of Results

- Loss-Epoch and Accuracy-Epoch graphs were plotted for both training and validation sets to visualize the model's performance over epochs.
- Correct and incorrect predictions were visualized by displaying the images along with their predicted labels.

Only linear layers without any activation function:



With Relu:

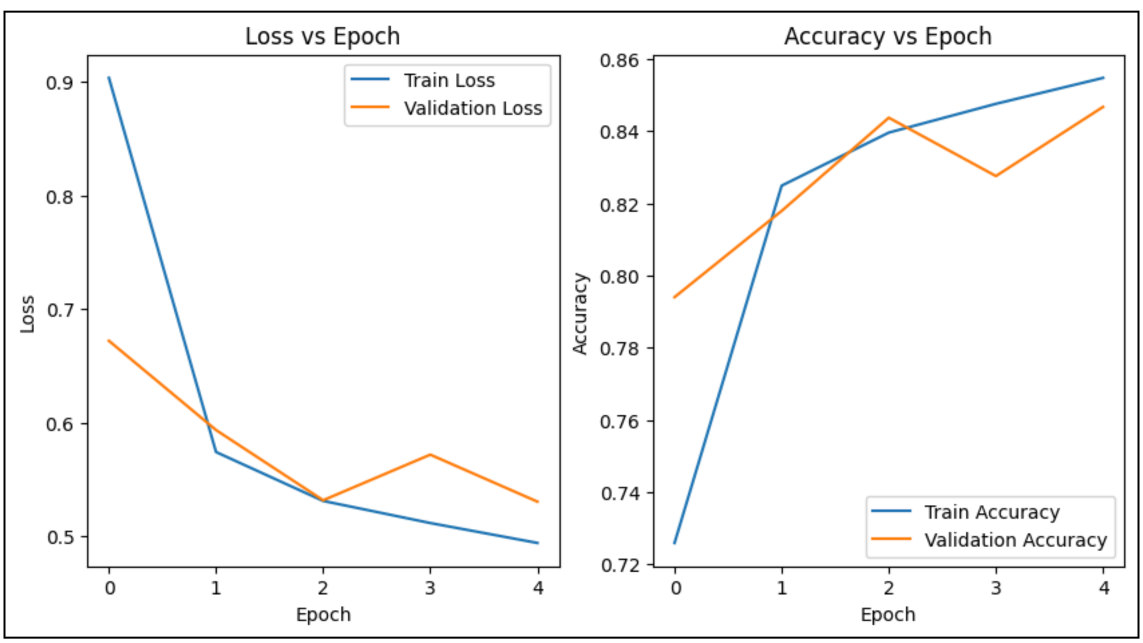
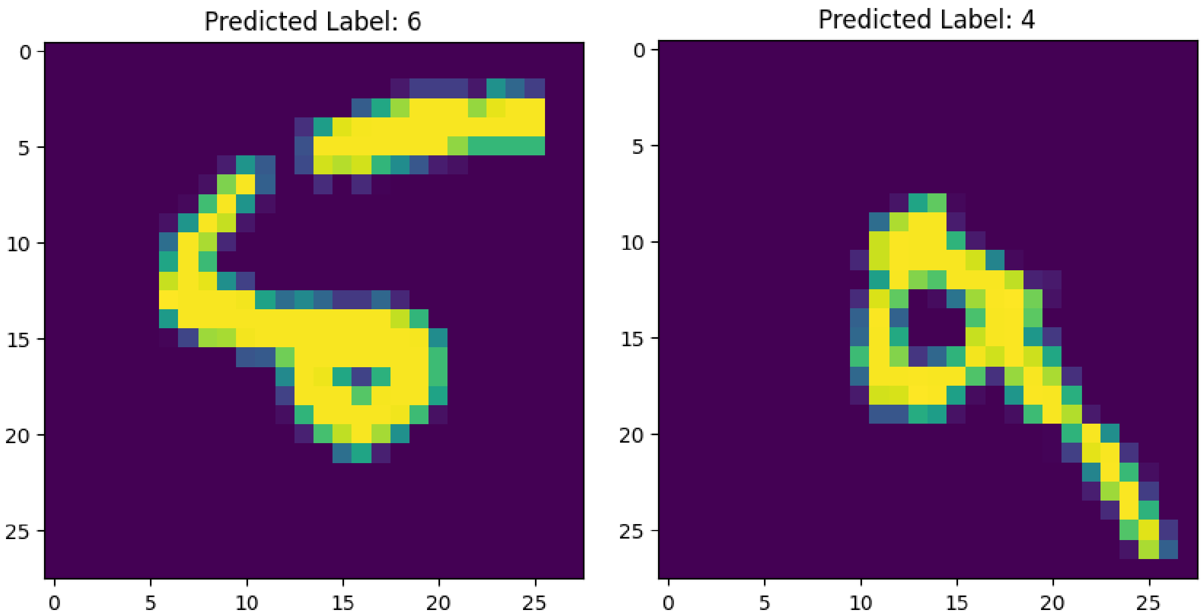


Image which was classified correct(left) and incorrectly(right):



Conclusion:

When non-linear activation functions like ReLU were used, the MLP model's performance—which was trained and assessed on the MNIST dataset—rose. ReLU activation functions allowed the model to obtain a far higher accuracy on the validation data—roughly 84%—than it could have with only linear layers. This important improvement emphasizes the need of incorporating nonlinearity into neural network topologies so that the model may recognize more complex patterns and connections in the input. ReLU activation has shown to be useful in improving neural network performance in tasks like photo classification, most especially in the case of MNIST digit recognition, overall, by the higher accuracy obtained.