CSE 343: Machine Learning

Assignment 3: Report

Aditya Aggarwal **2022028**

SECTION B

DATA PREPROCESSING

- The provided dataset is MNIST digit classification. It contains 60,000 images of digits.
- We loaded the dataset on the local machine and flattened the 28 X 28-pixel images to a matrix of the order (60000, 784) representing the number of samples and number of features. We also obtained the corresponding labels in an array.
- We shuffled the data using random seed 42 and then split the data into train, test, and val in an 8:1:1 ratio.
- We applied min-max scaling on the dataset.

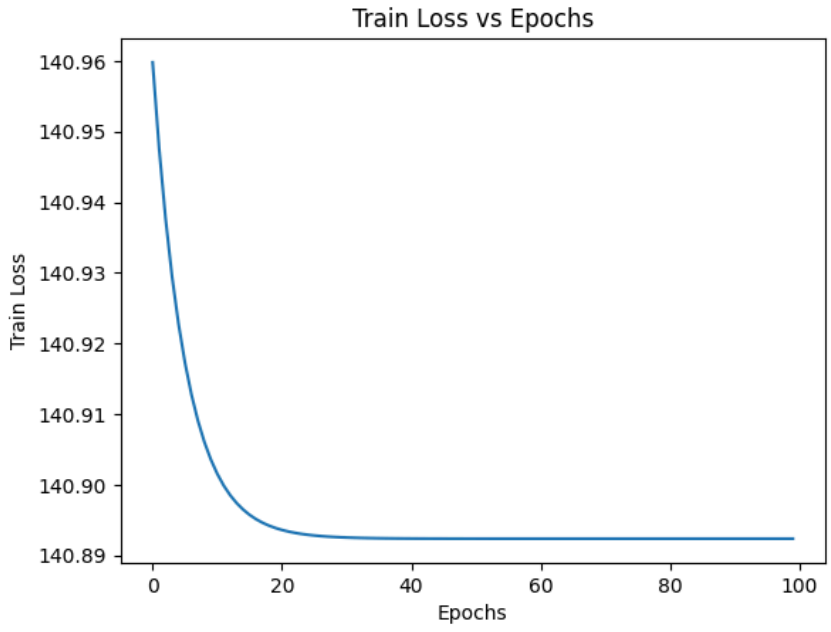
RESULTS

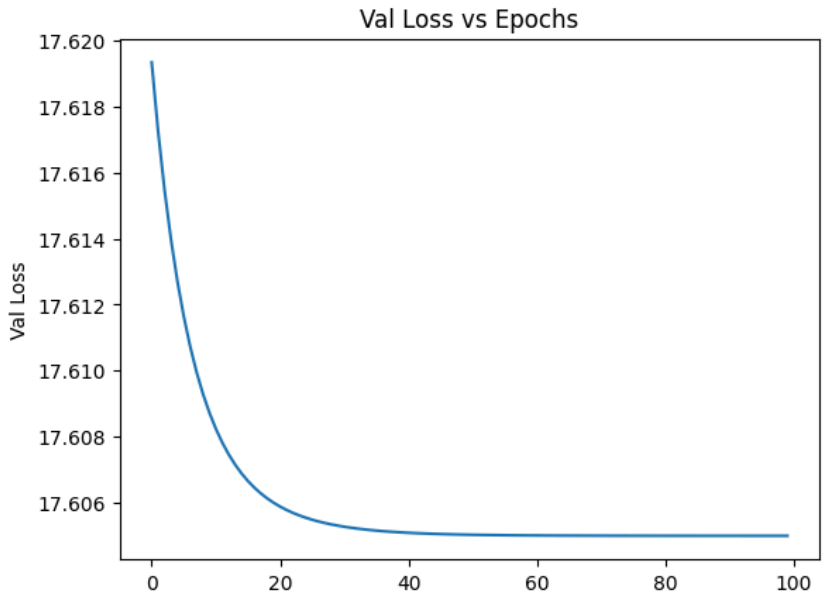
Activation: Sigmoid

Weight Initialization: Zero Initialization

Accuracy: 11.23%

Loss:



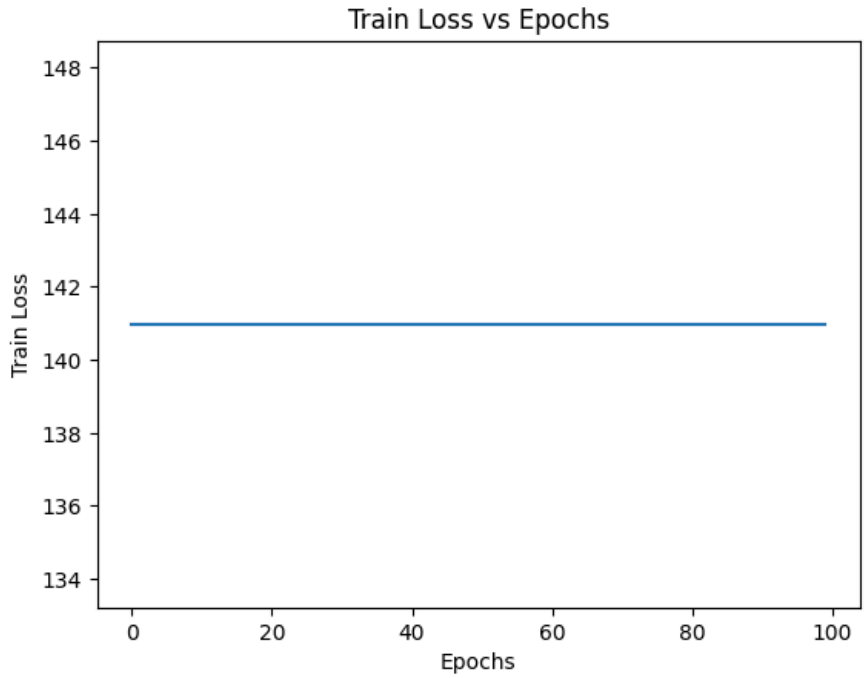


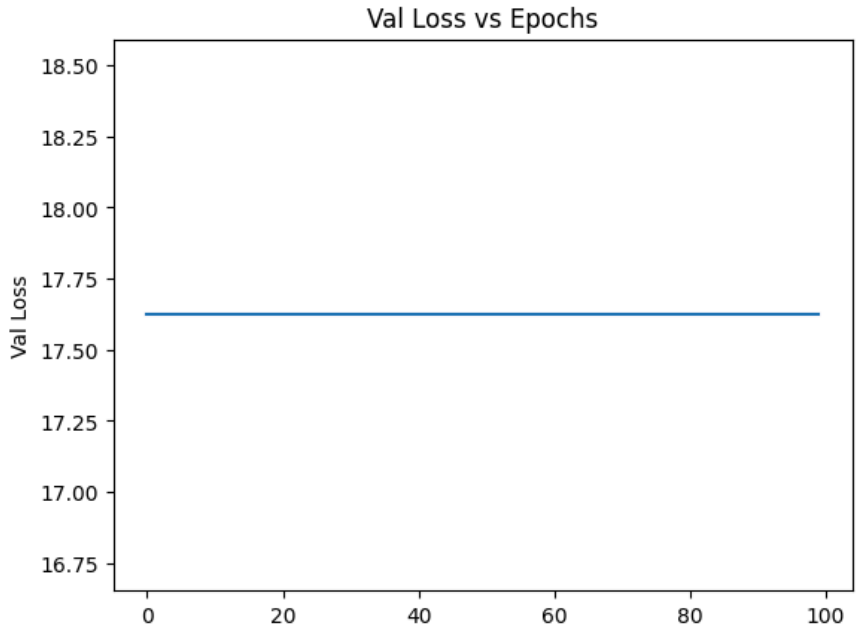
Activation: tanh

Weight Initialization: Zero Initialization

Accuracy: 9.82%

Train Loss:



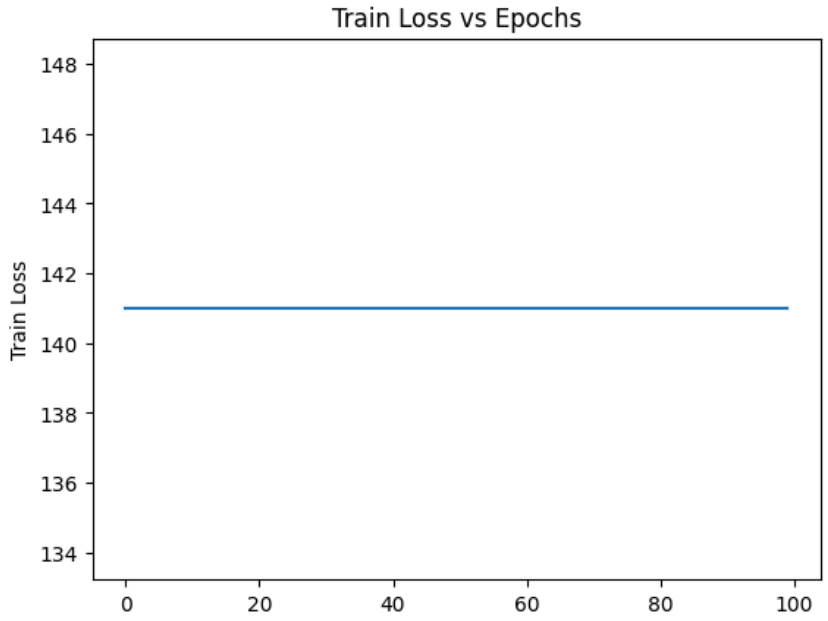


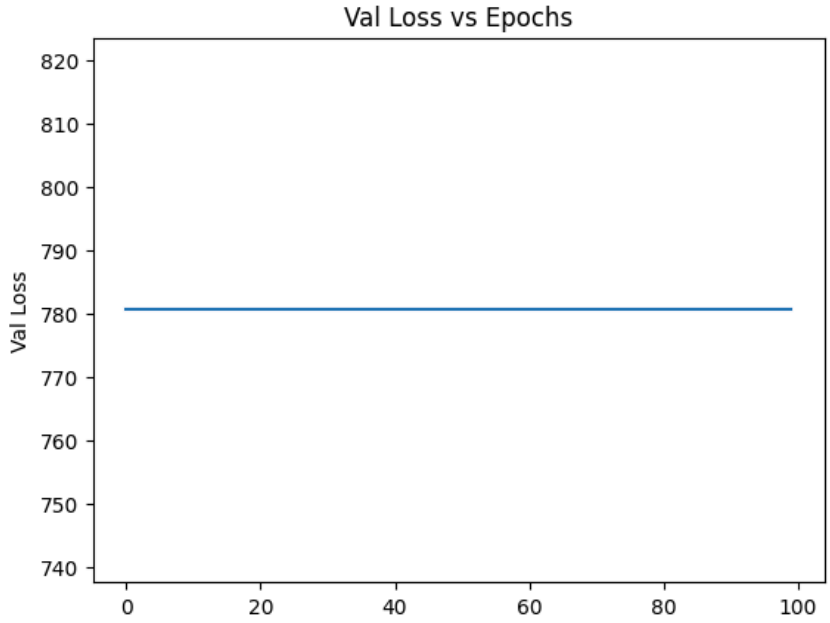
Activation: ReLU

Weight Initialization: Zero Initialization

Accuracy: 9.82%

Train Loss:



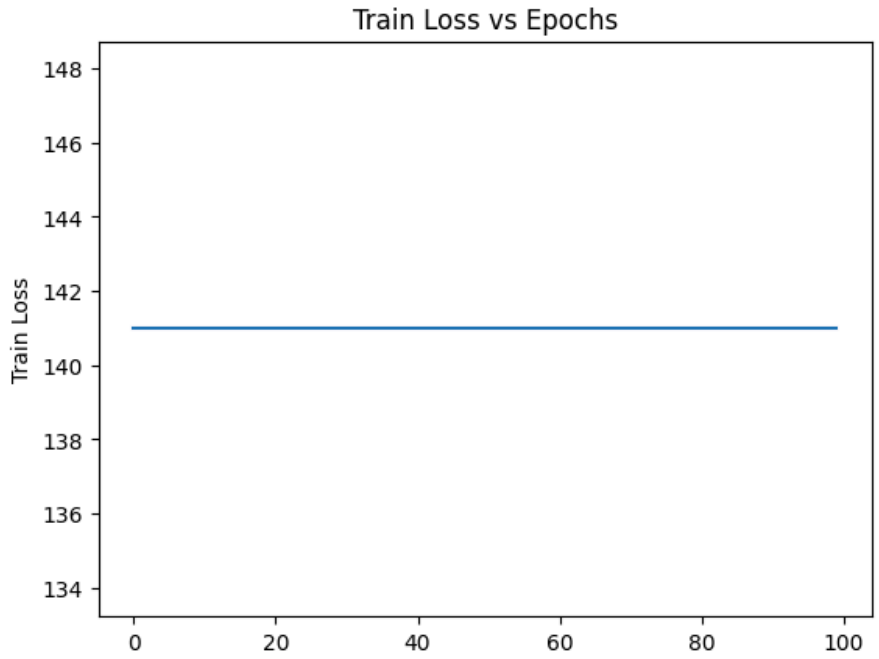


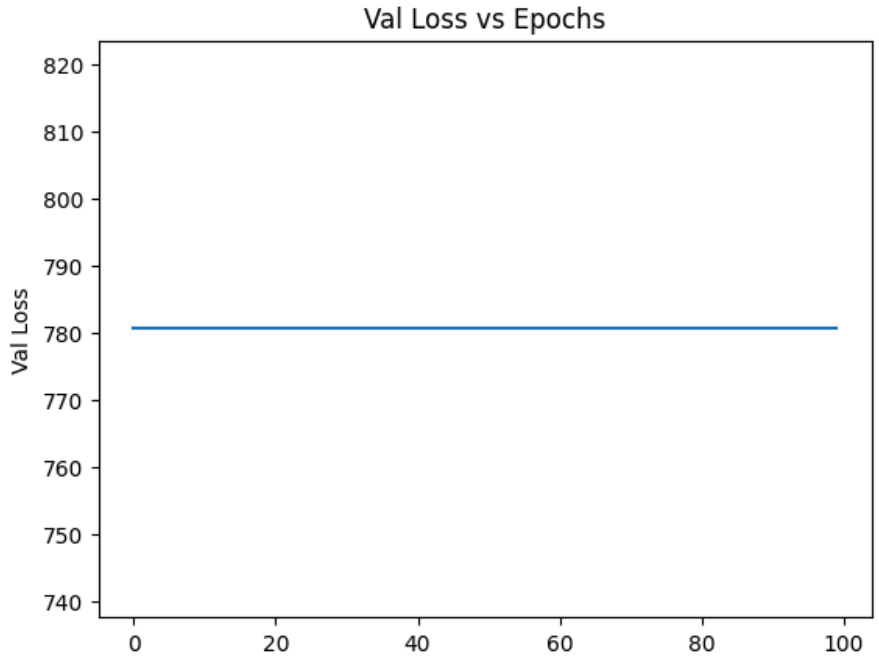
Activation: Leaky ReLU

Weight Initialization: Zero Initialization

Accuracy: 9.82%

Train Loss:



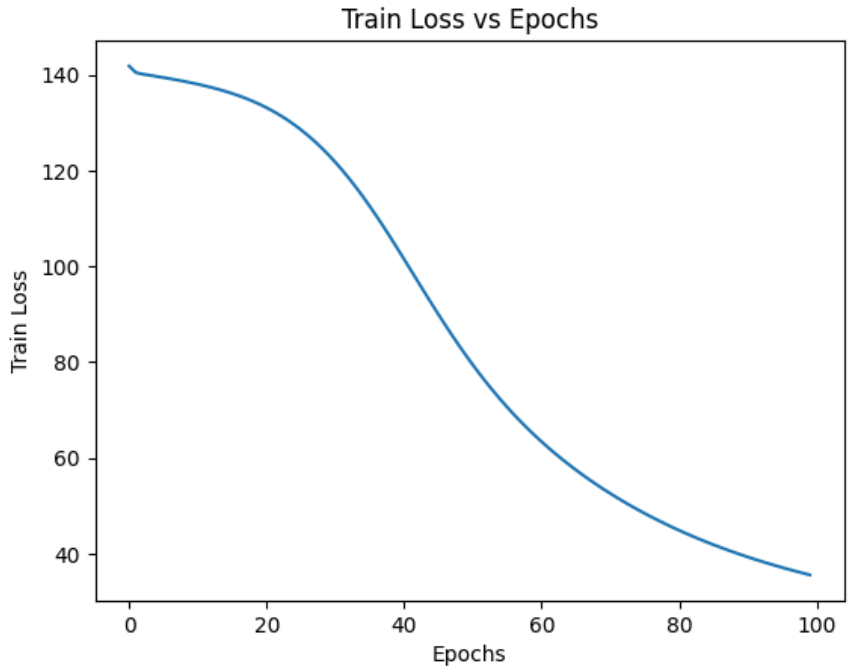


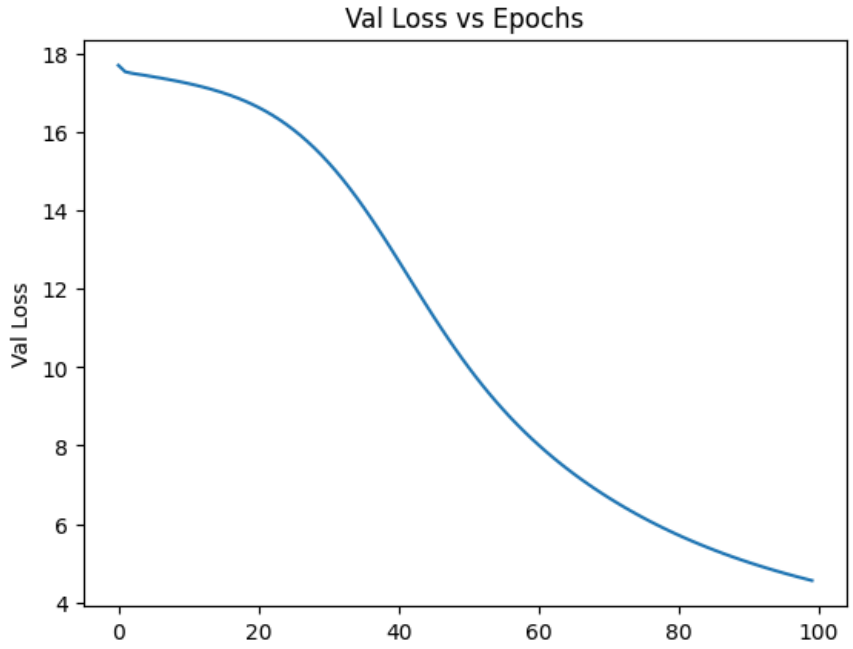
Activation: Sigmoid

Weight Initialization: Random Initialization

Accuracy: 85.43%

Loss:



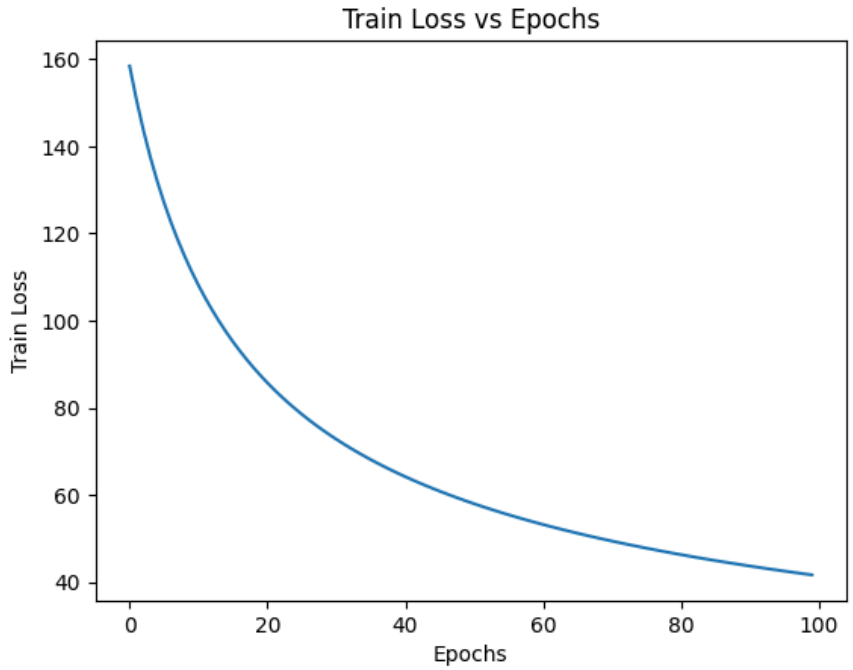


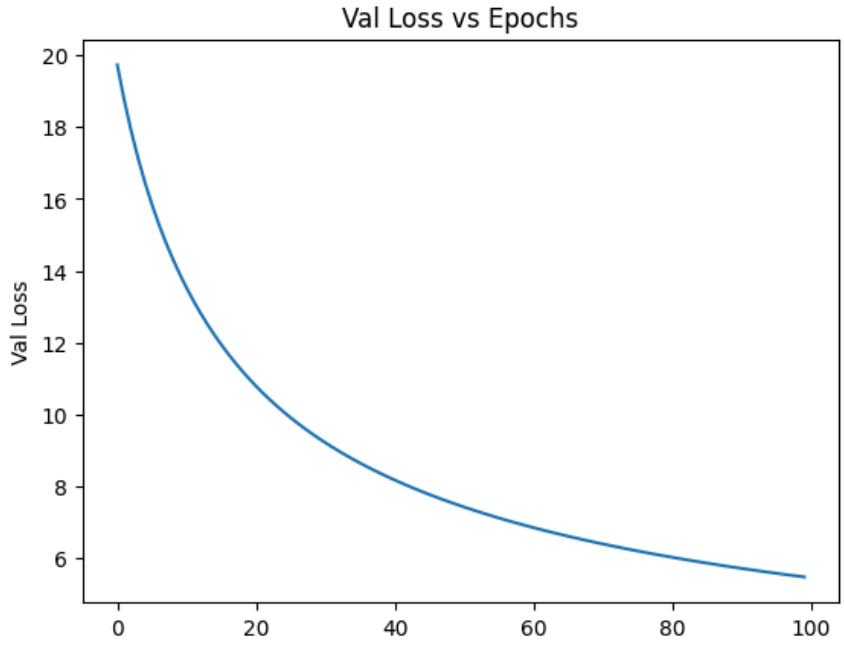
Activation: tanh

Weight Initialization: Random Initialization

Accuracy: 79.22%

Loss:



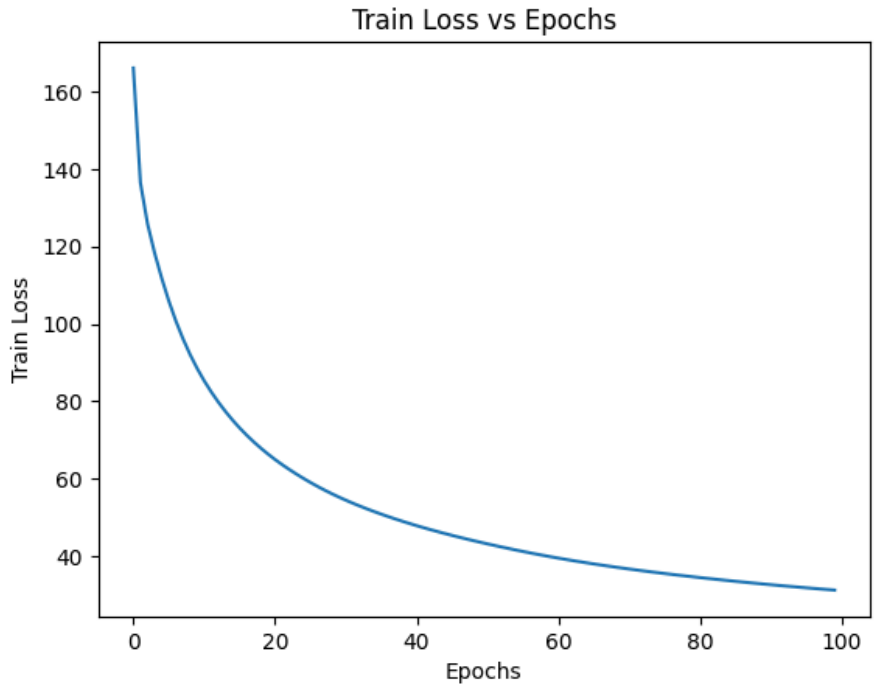


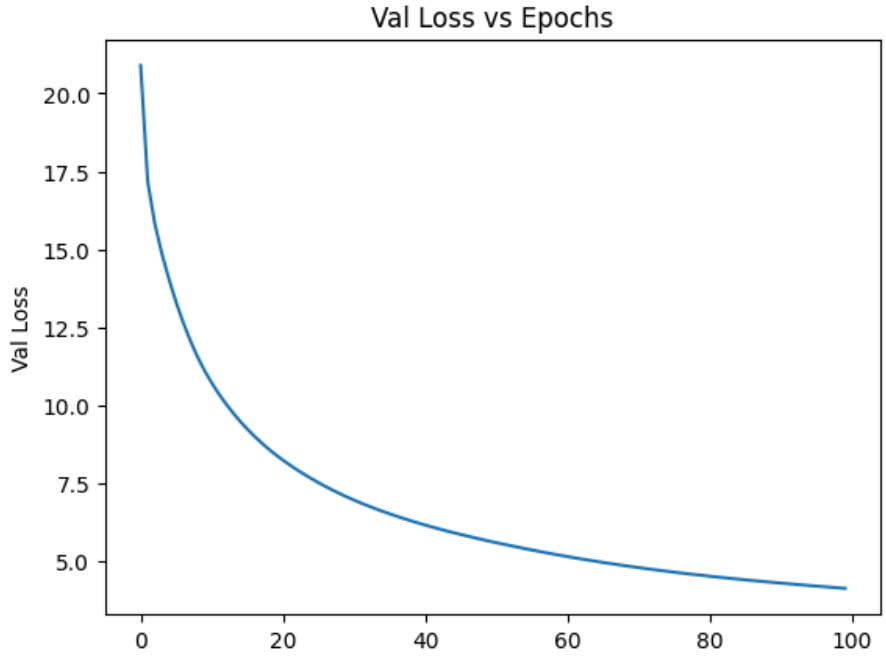
Activation: ReLU

Weight Initialization: Random Initialization

Accuracy: 83.28%

Loss:



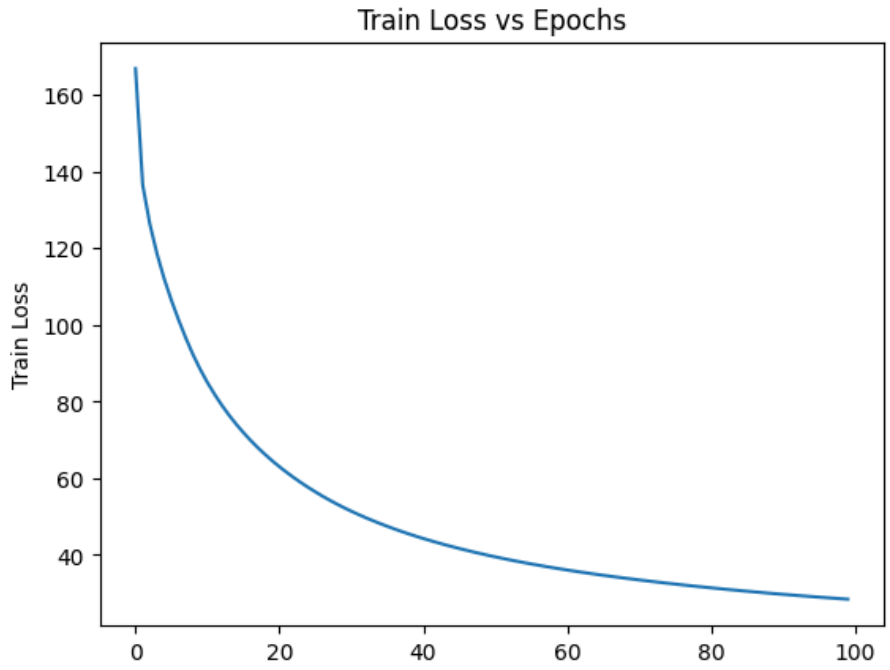


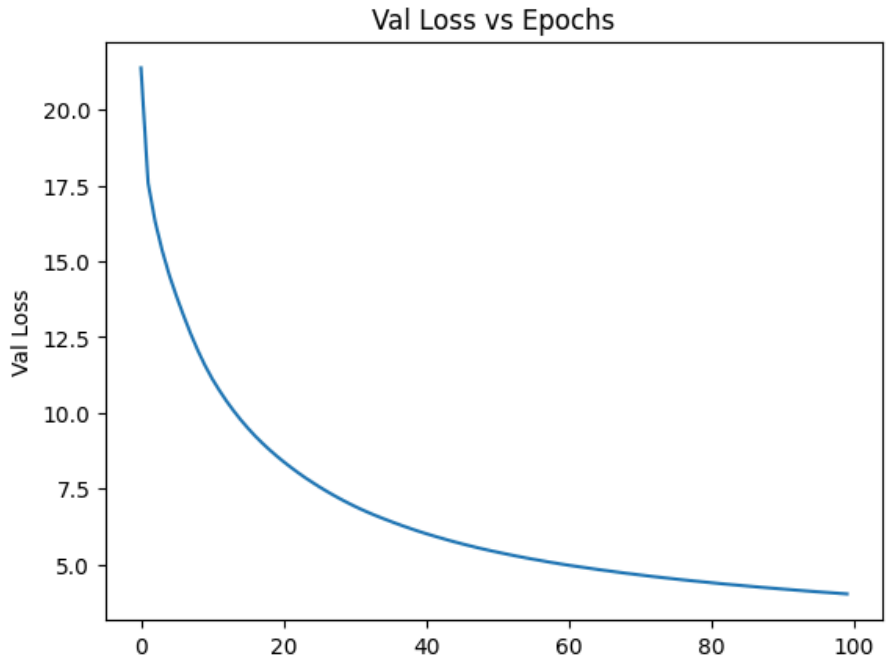
Activation: Leaky ReLU

Weight Initialization: Random Initialization

Accuracy: 86.26%

Loss:



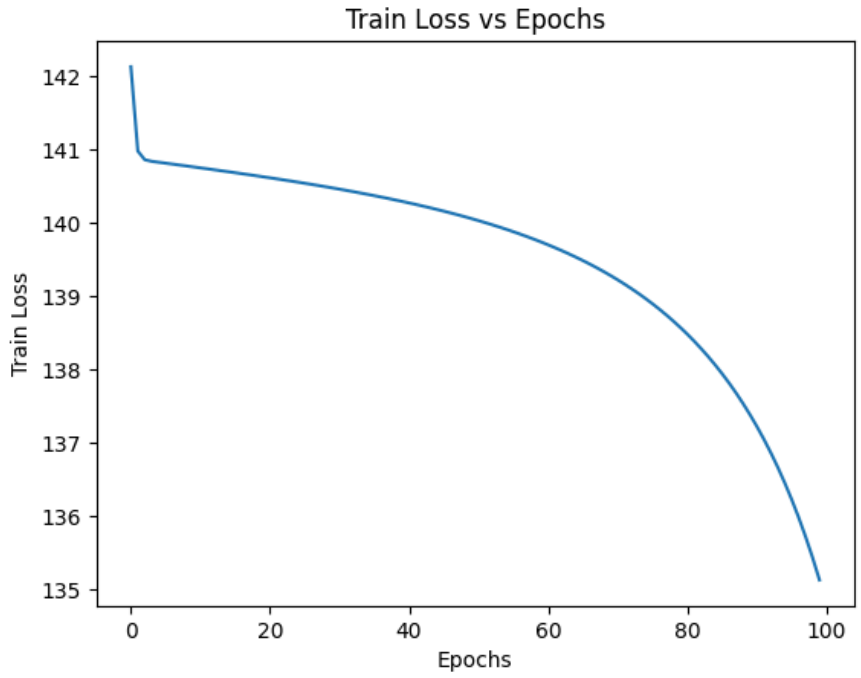


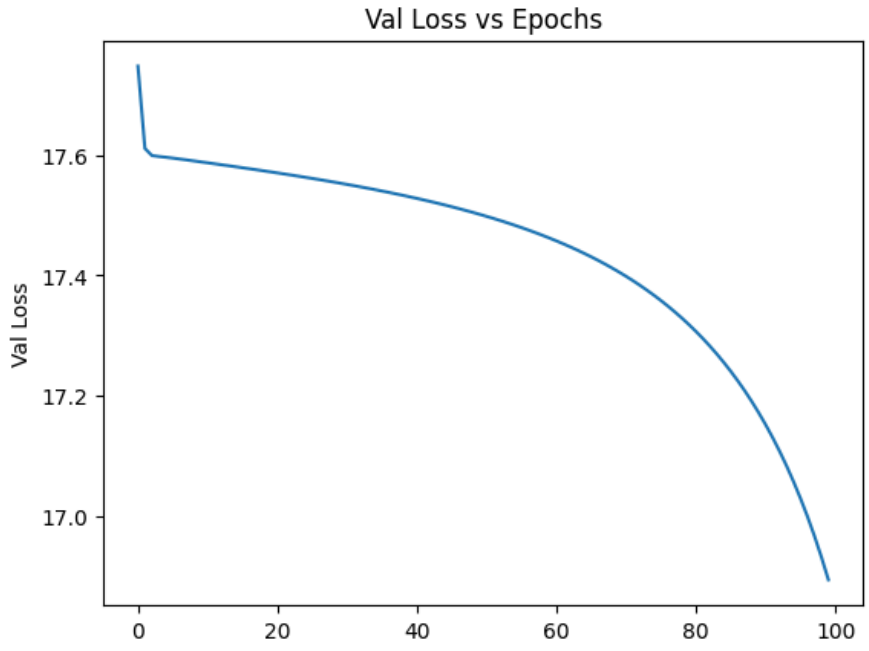
Activation: Sigmoid

Weight Initialization: Normal Initialization

Accuracy: 43.42%

Loss:



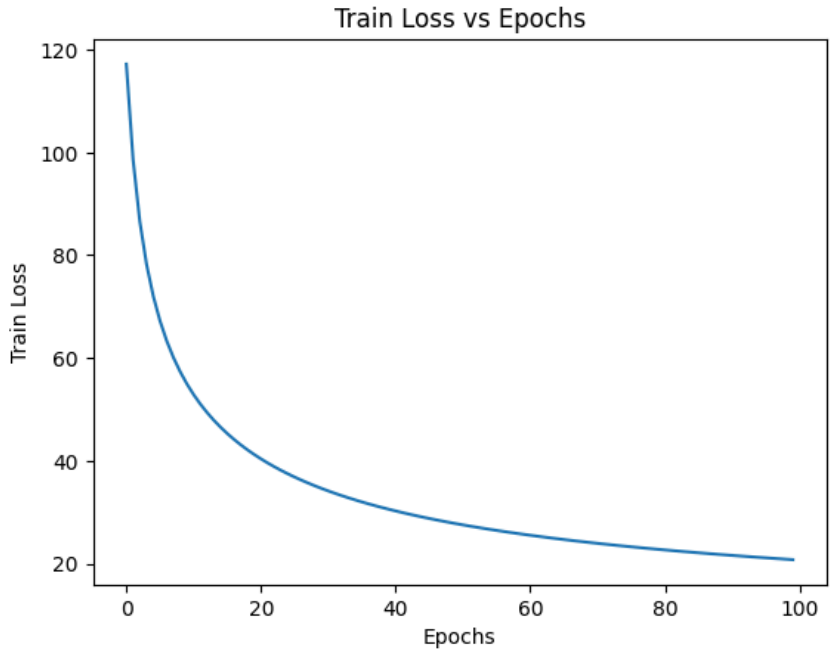


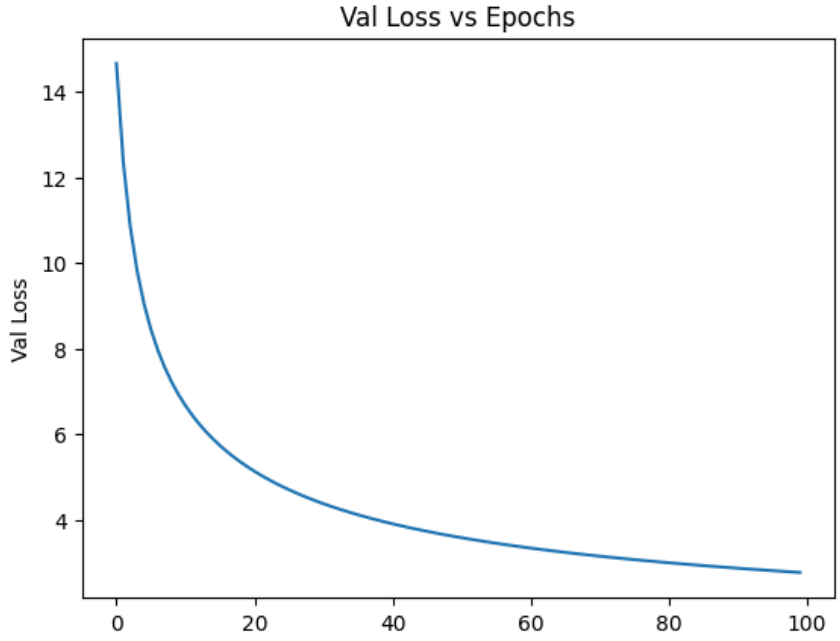
Activation: tanh

Weight Initialization: Normal Initialization

Accuracy: 91.27%

Train Loss:



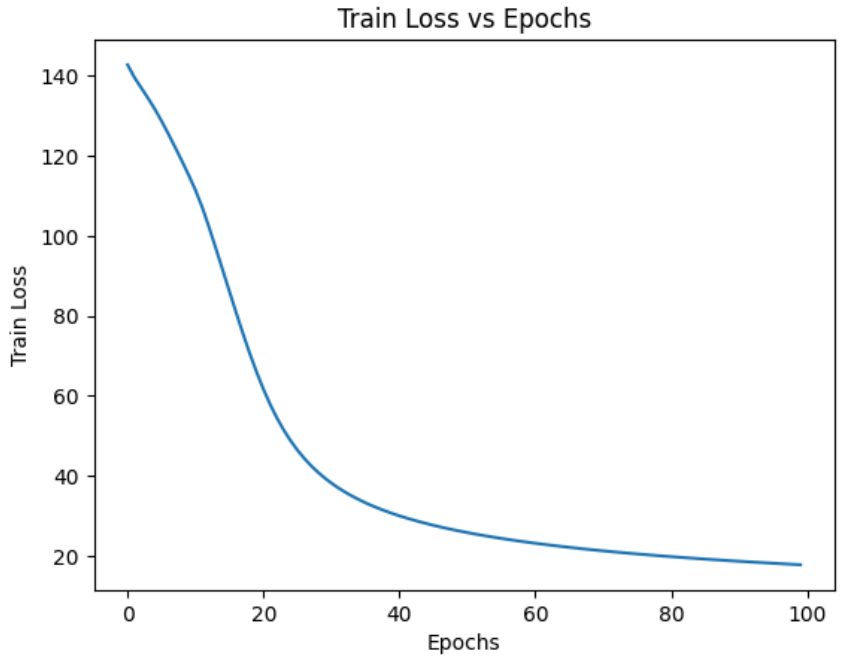


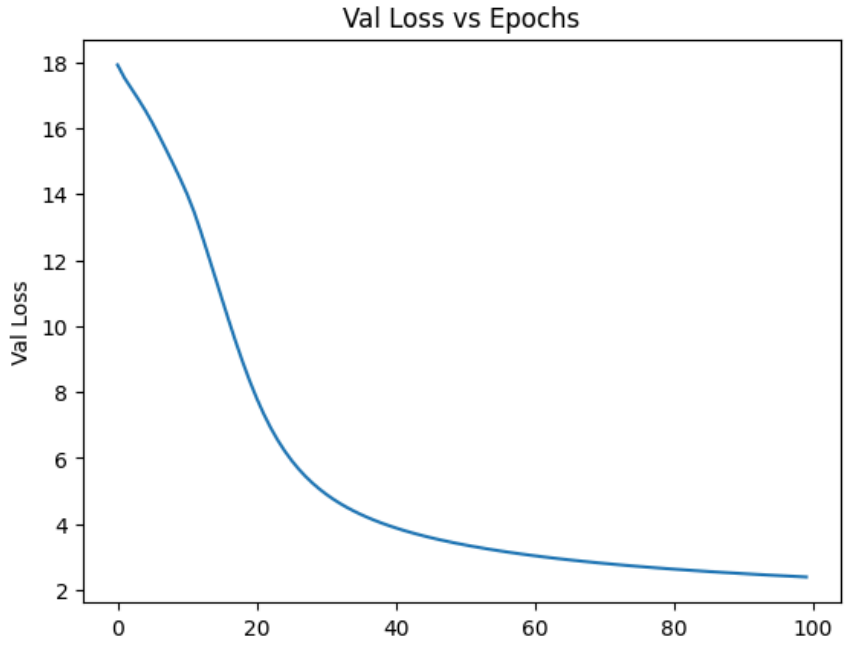
Activation: ReLU

Weight Initialization: Normal Initialization

Accuracy: 91.52%

Loss:



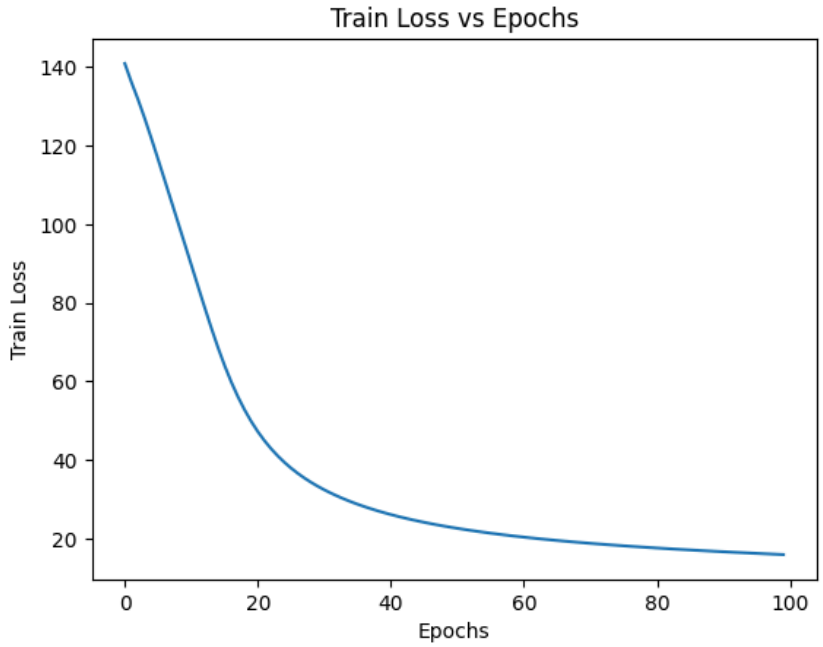


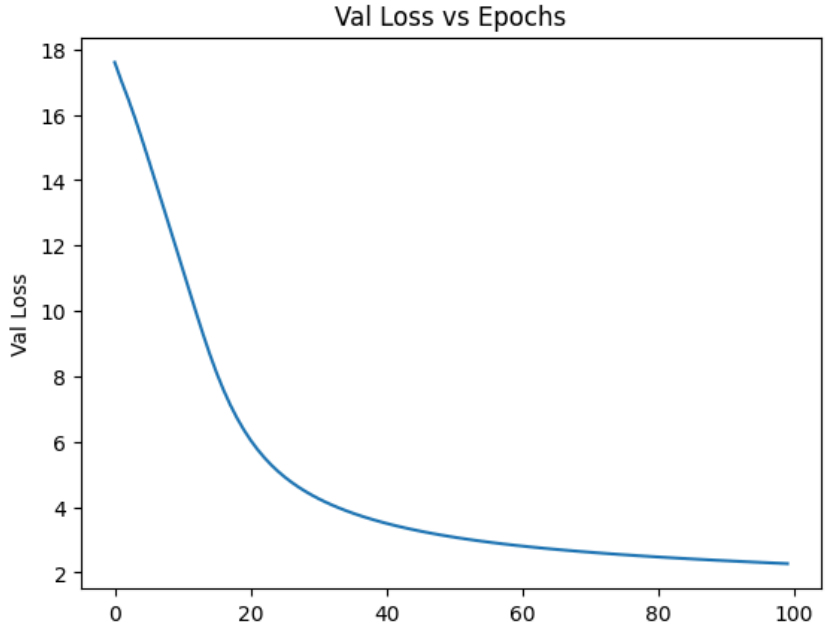
Activation: Leaky ReLU

Weight Initialization: Normal Initialization

Accuracy: 85.43%

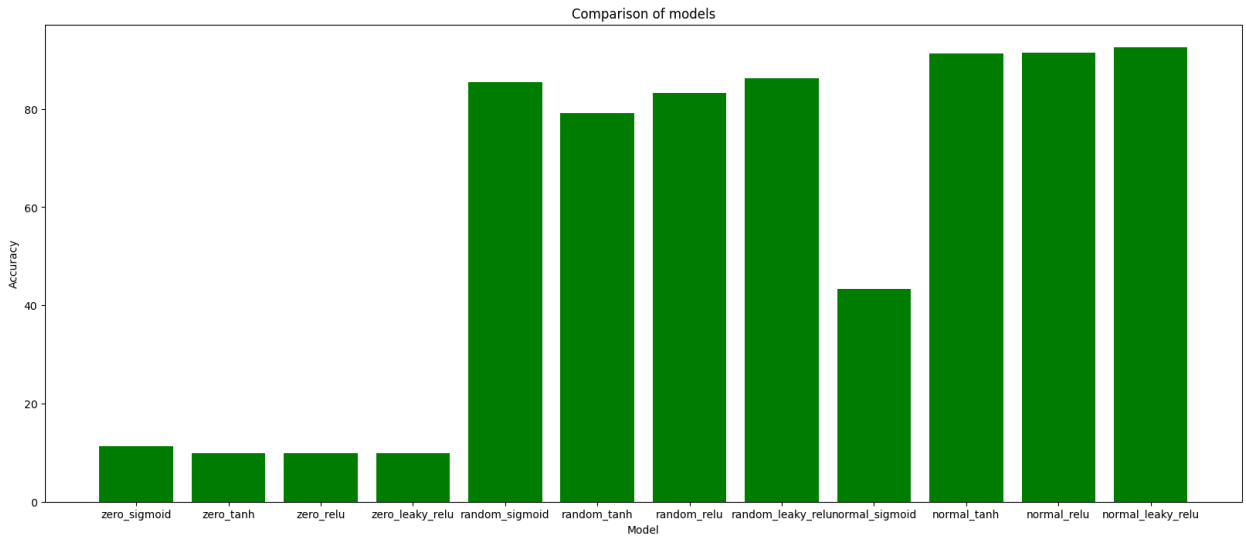
Loss:





ANALYSIS

- Leaky ReLU with normalized (He) initialization performs best on the test set with an accuracy of 92.6%.
- All models with weights initialized to zero perform poorly on the test set, with an approximate accuracy of 10%, which is the same as a random guess for the 10-way digit classification. This is because the neural network consists of only dead neurons and becomes symmetric when weights are zero.
- The sigmoid function requires a larger learning rate to compete with the performance of other models. This might be because a very small learning rate results in the problem of vanishing gradients. Another significant problem is that the number of epochs is very low for the given learning rate.
- Normal initialization of weights uses He-initialization, which scales the weights with the distribution W~N(0, 2/n), where n is the number of neurons in the previous layer. The use of this method especially boosts the performance in ReLU and leaky ReLU activations.
- Even though the val loss follows the same trend as the training loss (due to the low learning rate and a low number of epochs), it can be observed that this pattern deviates from a very large learning rate. This is due to overfitting, where the val loss shoots up after a certain number of epochs due to a lack of generalization. This is demonstrated in the code where $lr = 0.1$.
- The sigmoid function is the only activation function that performs better when the weights are initialized randomly compared to when they are initialized normally using He-scaling. This is because He-scaling was specifically designed for ReLU and ReLU-like functions that are unbounded on the positive side and have a mean close to zero after activation.



Analysis of models based on their accuracies