Neural Networks

Multi-Layer Perceptron (MLP) with Backpropagation(BP)

INTRODUCTION

The objective of this project is to implement a Multi-Layer Perceptron (MLP) using the backpropagation (BP) algorithm for training. The neural network will be designed to classify handwritten digits from the MNIST dataset. The dataset comprises grayscale images of handwritten digits, each represented as a column vector with values ranging from 0 to 1. The primary aim is to harness the potential of neural networks for the classification of handwritten digits sourced from the MNIST dataset.

The project is divided into three parts:

Part I: "1-Detector"

In this part, the goal is to create a network capable of detecting the digit "1." The architecture includes an input layer with 784 neurons, a hidden layer with a user-defined number of neurons (denoted as HP1), and one output neuron. The network is trained using the backpropagation algorithm, with two different learning rates (α). The results will be evaluated based on the mean squared error (MSE) for the training set. The chosen HP1 and learning rates will be reported.

Part II: "3-Detector"

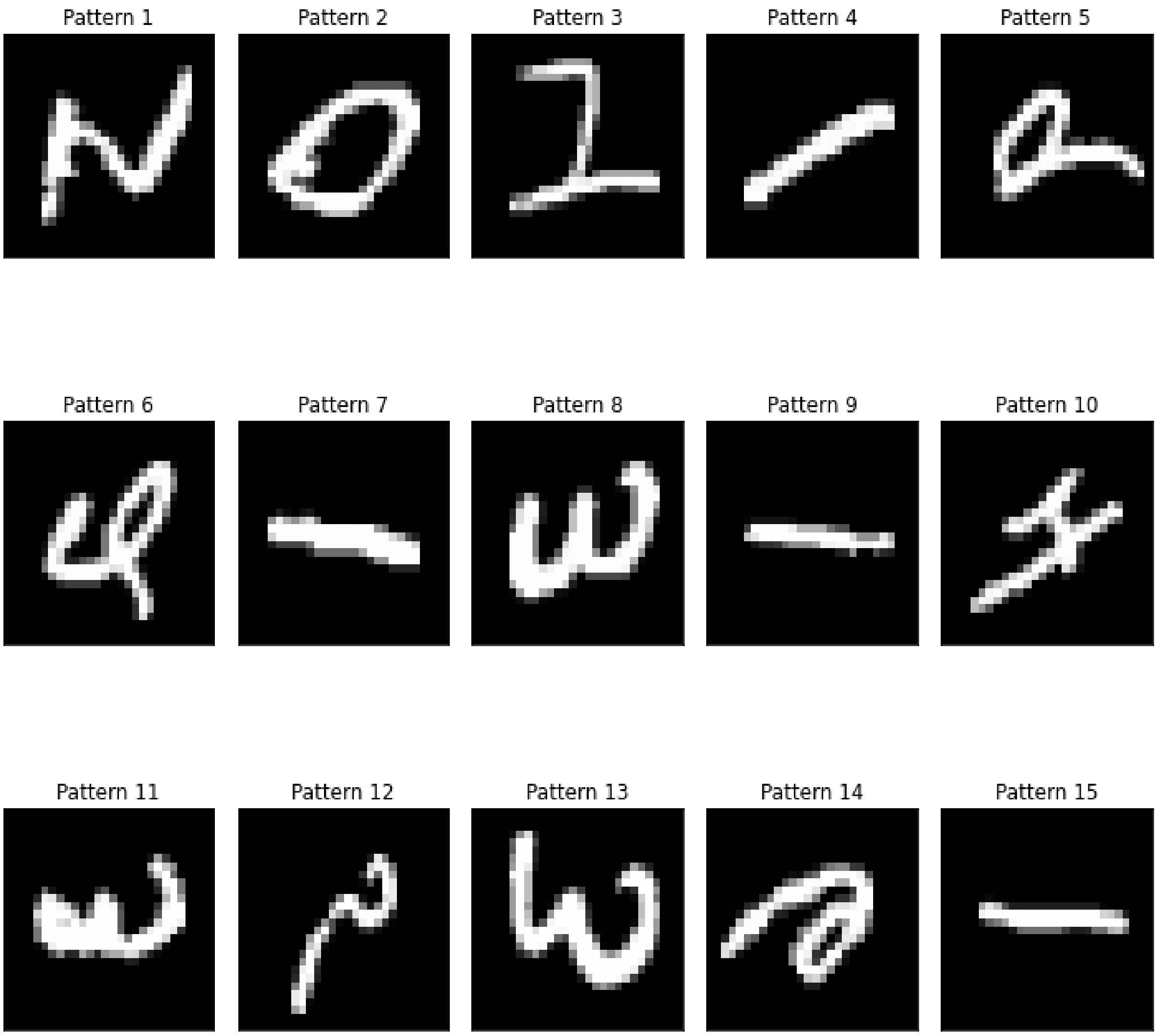
Similar to Part I, this section focuses on creating a network to detect the digit "3." Additionally, a validation set is introduced, and the training process involves monitoring the MSE on both the training and validation sets. The architecture includes an input layer, a hidden layer with a user-defined number of neurons (HP2), and one output neuron. The training will be conducted for two different learning rates (α), and the results, including MSE values, will be reported.

Part III: "Multi-Class Classification"

In this part, the objective is to design a network capable of identifying any digit from 0 to 9. The architecture consists of 784 input neurons, a hidden layer with a user-defined number of neurons (HP3), and 10 output neurons. The targets are 10 x 1 column vectors, with the correct digit set to +1 and others set to -1. The training process involves using both the training and validation sets. The performance will be evaluated based on the maximum activation of the output neurons during testing, indicating the identified digit.

The project emphasizes experimentation with different hyperparameters, including the number of hidden neurons and learning rates. Learning curves and weight evolution will be visualized to analyze the training process. Additionally, the final section of the report will include observations, conclusions, and explanations for the chosen stopping epochs in each part of the project. All the codes related to the project were implemented in the python programming language.

First 15 training patterns displayed below from the jupyter notebook:



This below MATLAB code is designed to load a numeric array from a file named 'TRN_MNIST.mat', which presumably contains training data related to the MNIST dataset. After loading the data, it converts the numeric array 'TRN10Y' into a CSV (Comma-Separated Values) file named 'TRN10Y.csv'.

```
% Load the numeric array from the provided file (replace 'your_file_name.mat' with the actual file name)

load('TRN_MNIST.mat');


% Specify the output CSV file name

outputFileName = 'TRN10Y.csv';


% Write the numeric array to a CSV file

% Make sure to replace 'your_numeric_array' with the actual variable name of your numeric array

writematrix(TRN10Y, outputFileName) ;
```

PART I : "1-detector"

Network Architecture

num_input_units = 784

num_hidden_units = 50

num_output_units = 1

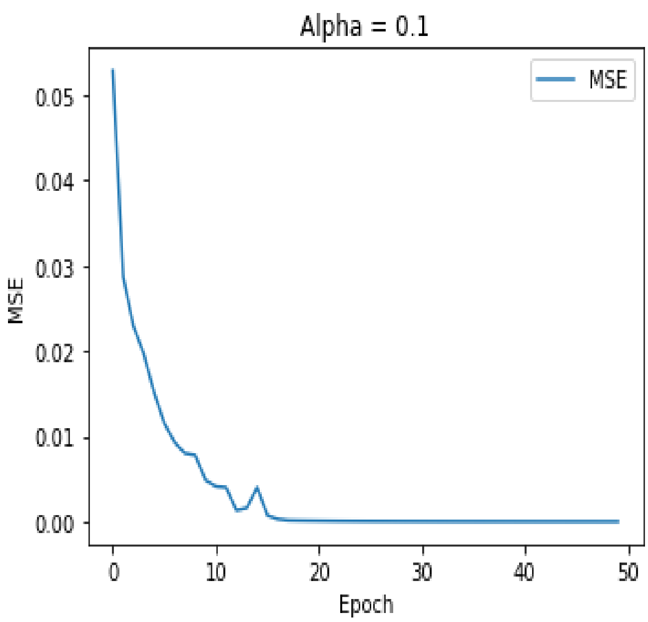
Results Obtained:

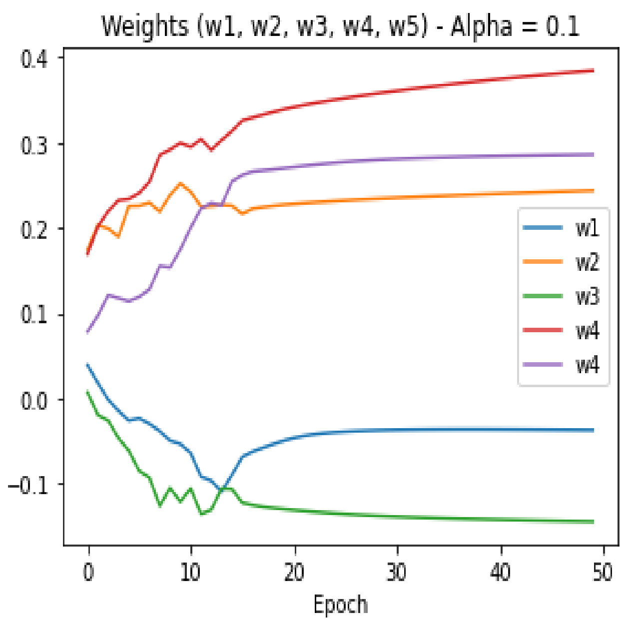
PART I: Using this number of Hidden Units, HP1 = **50**	
When using α = **0.1**	When using α = **0.001**
Trained for **50** epochs	Trained for **50** epochs
The MSE after last epoch of training was Final Training MSE = `8.50324487264841e-06`	The MSE after last epoch of training was Final Training MSE = `6.266157858466503e-06`

Below are the

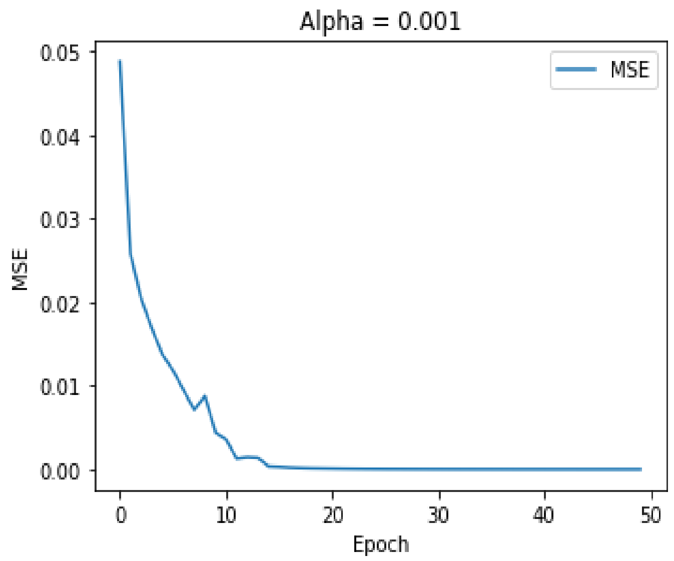
* Plots of the “Learning Curve” (i.e., MSE after each epoch of training vs. epoch number);
* Plot of the 5 weights from the network vs. epoch number. All 5 traces are appeared in the same figure.

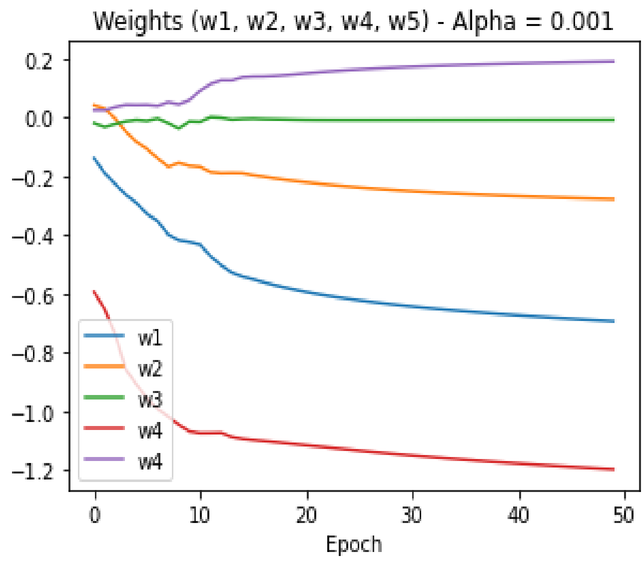
When alpha = 0.1





When alpha = 0.001





<u>Observations and Conclusions:</u>

Training with Different Learning Rates (α):

- Two training runs were conducted with the same number of hidden units (HP1 = 50) but different learning rates (α = 0.1 and α = 0.001).
- Both runs were trained for 50 epochs.

Final Training Mean Squared Error (MSE):

- The final training MSE for α = 0.1 is 8.50324487264841e-06.
- The final training MSE for α = 0.001 is 6.266157858466503e-06.

Comparison of MSE:

- The final training MSE for α = 0.001 is lower than the MSE for α = 0.1.
- A lower MSE indicates **better convergence** and accuracy of the neural network on the training set.

Choice of Best Learning Rate (α = 0.001):

- The decision to choose α = 0.001 as the best learning rate is based on achieving a **lower final training MSE**.

- A smaller learning rate often allows the optimization process to converge more accurately, leading to better generalization on the training set.

Observation on Learning Rates:

- The learning rate $\alpha = 0.001$ resulted in a **smoother convergence,** likely avoiding overshooting and converging to a more optimal solution.
- A learning rate that is too large (e.g., $\alpha = 0.1$) might lead to oscillations or **divergence** in the optimization process.

Conclusions:

- The choice of the learning rate is crucial in training neural networks, and it depends on the specific characteristics of the dataset and the network architecture.
- In this case, $\alpha = 0.001$ is preferred as it led to a lower MSE, indicating better training performance.
- The optimal learning rate is often found through experimentation and validation on different datasets.

<u>TEST SET RESULTS:</u>

For testing by using the final trained weights and biases from the alpha = 0.001,

I have achieved the below statistics:

```
Final Test MSE for the Best Model: [[0.020778596498603183]]
9937  (correct classifications)
63     (incorrect classifications)
0.9937  (Hit ratio)
Accuracy: 99.37%
```

Observation:

The final test Mean Squared Error (MSE) for the best model is approximately 0.020780.02078, indicating a relatively low error in predictions. Out of the 10,000 test samples, 9,937 were correctly classified, **resulting in an impressive hit ratio of 99.37**. This high accuracy suggests that the trained model performs well in identifying handwritten digits from the MNIST dataset.

PART II: “3-detector”

Network Architecture

num_input_units = 784

num_hidden_units = 50

num_output_units = 1

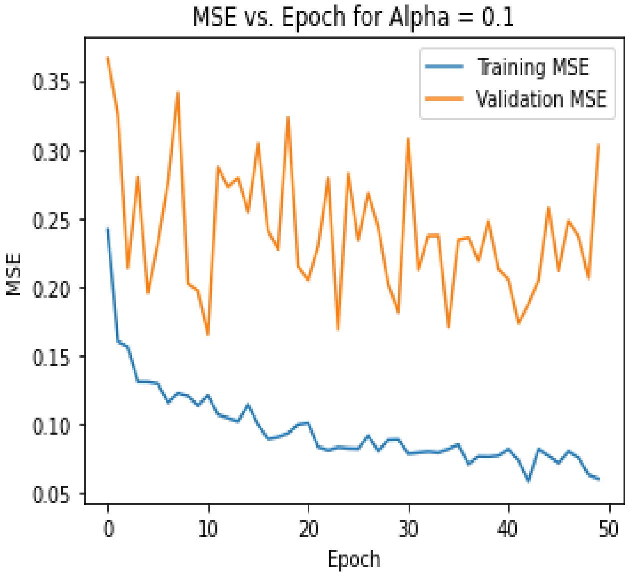
Results Obtained:

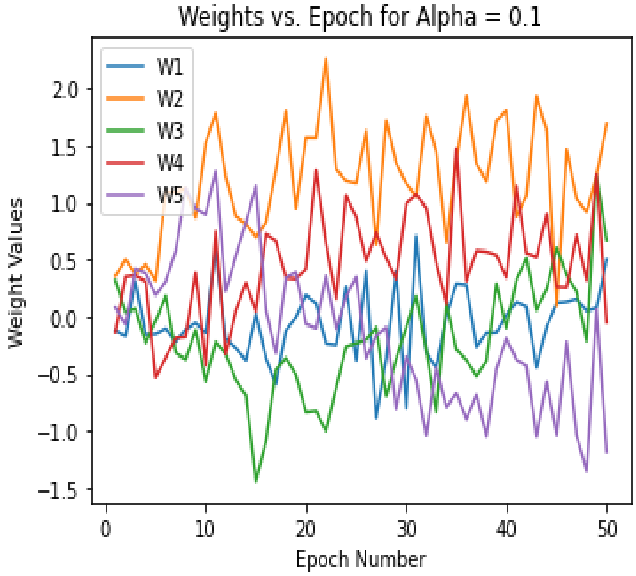
PART II: Using this number of Hidden Units, H_{P1} = **50**	
When using α = **0.1**	When using α = **0.001**
Trained for **50** epochs	Trained for **50** epochs
Final MSE *from the training set* = 0.060322833208443706	Final MSE *from the training set* = 0.007263365790496891
Final MSE *from the validation set* = 0.30263872693128463	Final MSE *from the validation set* = 0.08464098271772255

For each of the 2 alpha cases in the table, I have included the below plots:

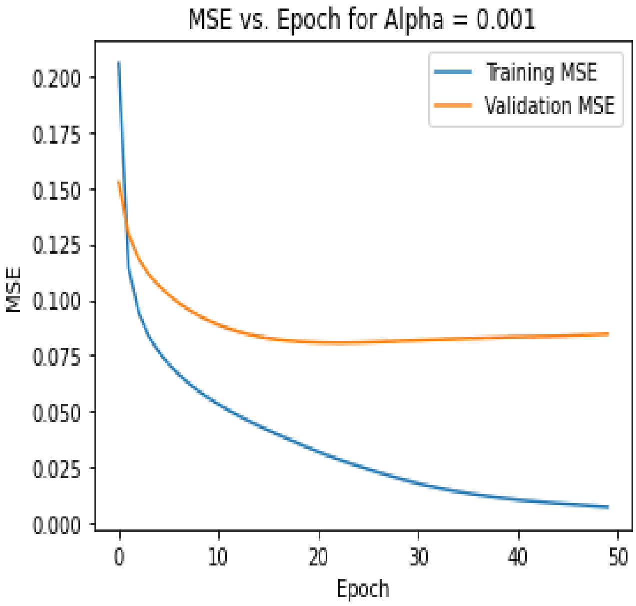
• The Plot of 5 weights from the network vs. epoch number.
• The Overlay Plot of the “Learning Curve” for Validation and training (i.e., MSE after each epoch of training vs. epoch number)

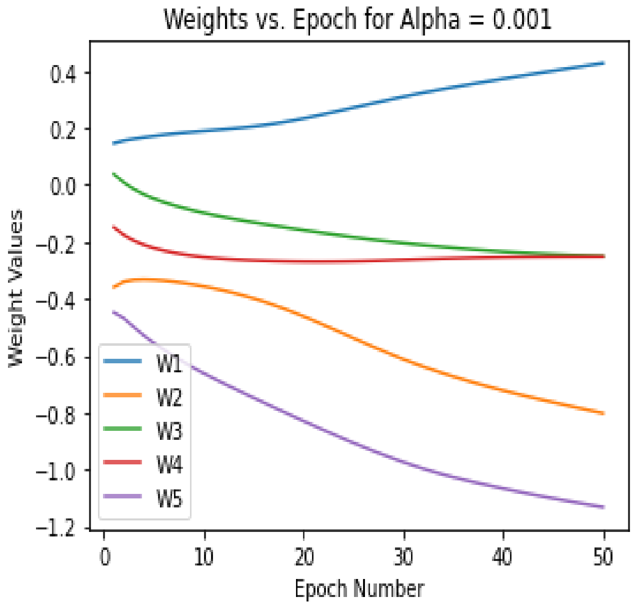
When alpha = 0.1





When = alpha 0.001





Observations and Conclusions for 3-Detector Training and Validation:

Training MSE:

- For $\alpha = 0.1$: Final MSE on the training set is 0.06032.
- For $\alpha = 0.001$: Final MSE on the training set is 0.00726.
- The lower MSE for $\alpha = 0.001$ indicates that the smaller learning rate allows for better convergence during training.

Validation MSE:

- For $\alpha = 0.1$: Final MSE on the validation set is 0.30264.
- For $\alpha = 0.001$: Final MSE on the validation set is 0.08464.
- The lower validation MSE for $\alpha = 0.001$ suggests **better generalization performance**, as the model performs well **on unseen data.**

Learning Rate Impact:

- A smaller learning rate (0.001) results in both lower training and validation MSE, indicating more stable and accurate weight updates during training.
- A larger learning rate (0.1) leads to higher MSE on both sets, suggesting that the model might overshoot the optimal weights during training.

Selection of $\alpha = 0.001$:

- The choice of $\alpha = 0.001$ is better because it results in lower final MSE values for both training and validation sets.
- This indicates that the model trained with $\alpha = 0.001$ performs well not only on the training data but also on new, unseen data represented by the validation set.

Generalization Performance:

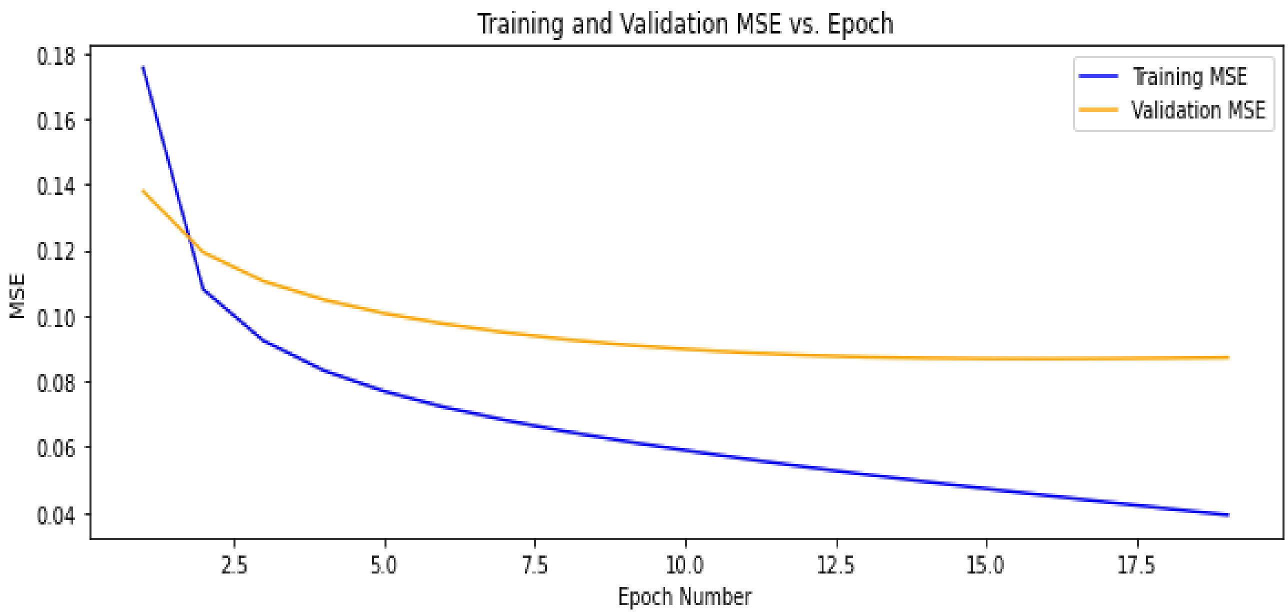
- The model with $\alpha = 0.001$ is expected to have **better generalization** performance on unseen data, which is crucial for the model's reliability in real-world scenarios.

In **conclusion,** a smaller learning rate ($\alpha = 0.001$) contributes to better convergence and generalization, as evidenced by lower final MSE values on both the training and validation sets. This choice reflects a more stable training process and a model that is **less likely to overfit** the training data.

Results and statistics obtained by using EARLY STOPPING with alpha = 0.001:

```
Epoch 1/50, Training MSE: [[0.1754744632071968]]
Epoch 1/50, Validation MSE: [[0.1379276653283358]]
Epoch 2/50, Training MSE: [[0.1079464249997841]]
Epoch 2/50, Validation MSE: [[0.1192940381929535]]
Epoch 3/50, Training MSE: [[0.09235734294266758]]
Epoch 3/50, Validation MSE: [[0.11046467151802417]]
Epoch 4/50, Training MSE: [[0.08327100728335789]]
Epoch 4/50, Validation MSE: [[0.10483007466823001]]
Epoch 5/50, Training MSE: [[0.07698419056557725]]
Epoch 5/50, Validation MSE: [[0.10070186577657474]]
Epoch 6/50, Training MSE: [[0.07213490025604846]]
Epoch 6/50, Validation MSE: [[0.09749392121557583]]
Epoch 7/50, Training MSE: [[0.06814919973080903]]
Epoch 7/50, Validation MSE: [[0.09491055041253191]]
Epoch 8/50, Training MSE: [[0.06473886735536792]]
Epoch 8/50, Validation MSE: [[0.09280830295746208]]
Epoch 9/50, Training MSE: [[0.06171857658049685]]
Epoch 9/50, Validation MSE: [[0.09110679362657202]]
Epoch 10/50, Training MSE: [[0.05895979292792998]]
Epoch 10/50, Validation MSE: [[0.0897447889040799]]
Epoch 11/50, Training MSE: [[0.056375071492971586]]
Epoch 11/50, Validation MSE: [[0.08868182496783725]]
Epoch 12/50, Training MSE: [[0.053919309660910415]]
Epoch 12/50, Validation MSE: [[0.08789662128102584]]
Epoch 13/50, Training MSE: [[0.051583019196607294]]
Epoch 13/50, Validation MSE: [[0.08735873348305594]]
Epoch 14/50, Training MSE: [[0.04936060254777975]]
Epoch 14/50, Validation MSE: [[0.08702180613399586]]
Epoch 15/50, Training MSE: [[0.04723433955392588]]
Epoch 15/50, Validation MSE: [[0.086847128799996]]
Epoch 16/50, Training MSE: [[0.04518065015157218]]
Epoch 16/50, Validation MSE: [[0.08680523150655978]]
Epoch 17/50, Training MSE: [[0.04317729862281505]]
Epoch 17/50, Validation MSE: [[0.08686888576360374]]
Epoch 18/50, Training MSE: [[0.04120914822623346]]
Epoch 18/50, Validation MSE: [[0.08701247129104626]]
Epoch 19/50, Training MSE: [[0.03927402147900722]]
Epoch 19/50, Validation MSE: [[0.08721830531241209]]
Early stopping at Epoch 19, Best Validation MSE: [[0.08680523150655978]
]
4855  (correct classification)
145    (incorrect classification)
0.971   (Hit ratio)
Accuracy: 97.10%
Final Training MSE for the Best Model: [[0.03927402147900722]]
Final Validation MSE for the Best Model: [[0.08680523150655978]]
```

<u>Overlay of Training & Validation Learning Curve FOR THIS 3rd AND FINAL RUN</u>



<u>Observations & Conclusions:</u>

Early Stopping Decision:

Early stopping is implemented to stop training when the validation MSE does not show significant improvement. The training stopped at Epoch 19 due to early stopping. This decision was made based on the validation MSE. The best validation MSE was achieved at Epoch 16, and the subsequent epochs did not show a substantial improvement, indicating that the model was not learning significantly from the training data.

Training and Validation MSE:

- Training MSE steadily decreases, indicating that the model is learning from the training data.
- Validation MSE decreases initially but starts to show signs of stabilization or slight increase after Epoch 16.

Accuracy and Hit Ratio:

- The model achieved a high accuracy of 97.10% on the validation set.
- The hit ratio (accuracy) is a good metric to evaluate the model's performance, indicating that 97.10% of the cases were correctly classified.

Early Stopping Importance:

- Early stopping is crucial to prevent overfitting and to ensure that the model generalizes well to unseen data.
- In this case, early stopping prevented the model from continuing to learn from the training data beyond Epoch 19, avoiding potential overfitting.

Validation MSE:

- Validation MSE is monitored to identify the point where the model's performance on unseen data starts to degrade.
- The increase in validation MSE after Epoch 16 suggests that the model may have started to overfit the training data.
- Validation MSE serves as a critical indicator for the model's ability to generalize to new data.

Why Early Stop at 19th epoch only:

Avoid Overfitting:

- Early stopping helps prevent overfitting, where the model becomes too specialized in the training data and performs poorly on new, unseen data.
- It provides a mechanism to halt training when further learning does not contribute significantly to better generalization.

Efficient Resource Usage:

- Stopping training early saves computational resources, time, and energy, as additional epochs may not yield substantial improvements.

Improved Generalization:

- By monitoring validation performance, early stopping ensures that the model is selected when it performs well on both training and validation data, leading to better generalization.

Reason behind selecting the model at epoch 16 as the best model:

- Overfitting occurs when a model learns the training data too well, including its noise and outliers, but fails to generalize to new, unseen data.
- Epoch 16 is a point where the model has learned from the training data but has not yet started to overfit the validation set.
- It strikes a balance between fitting the training data and generalizing to new data.

In **conclusion,** the goal is to halt training before overfitting occurs, ensuring better generalization to unseen data. Early stopping is a valuable technique to enhance model generalization, prevent overfitting, and optimize computational efficiency during the training process. It allows us to choose a model that strikes a good balance between learning from the training data and generalizing to new, unseen data.

TEST SET RESULTS:

For testing by using the final best trained weights and biases after early stopping with the alpha = 0.001,

I have achieved the below statistics:

```
Final Test MSE for the Best Model: [[0.05695269229460733]]
4913   (correct)
87    (incorrect)
0.9826 (Hit-Ratio)
Accuracy: 98.26%
```

Observations & Conclusions from the above obtained results:

The final test results after using the best model's weights and biases, which were determined through early stopping with alpha = 0.001, provide valuable insights

Test Mean Squared Error (MSE):

- The final test MSE for the best model is 0.05695269229460733.
- MSE measures the average squared difference between the predicted valu es and the actual values. A lower MSE indicates better model performanc e.
- Out of the total samples in the test set, 4913 were correctly classified, and 87 were incorrectly classified.

Generalization to Unseen Data:

- High Hit-Ratio and Accuracy values indicate strong model performance o n the test set.
- The high accuracy and Hit-Ratio on the test set suggest that the model ge neralizes well to previously unseen data.
- This generalization is crucial for the model to perform effectively in real-world scenarios where it encounters new, unobserved examples.

Consistency with Validation Performance:

- The test results are consistent with the model's performance on the validat ion set during training.
- Early stopping based on validation performance has likely contributed to t he model's ability to generalize to new data.

Model Robustness:

- The model's ability to achieve high accuracy on the test set indicates its ro bustness and reliability in making predictions on different data instances.

PART III: "MULTI-CLASS CLASSIFICATION"

Network Architecture:

num_input_units = 784

num_hidden_units = 60

num_output_units = 10

In this part the output layer has 10 processing elements (PEs). The unique aspect of this part lies in the testing phase, where the network is tasked with identifying which of the 10 digits (0 to 9) was presented at its input. To achieve this, the network is trained using 10 x 1 column vectors as targets, ensuring that ideally only one activation turns to "+1" while the activations of the remaining 9 output PEs are set to "-1" during the testing phase. This distinctive approach allows the neural network to explicitly classify the handwritten digit by activating the corresponding processing element, providing a clear and unambiguous identification of the presented digit based on the maximum activation in the output layer.

Results Obtained:

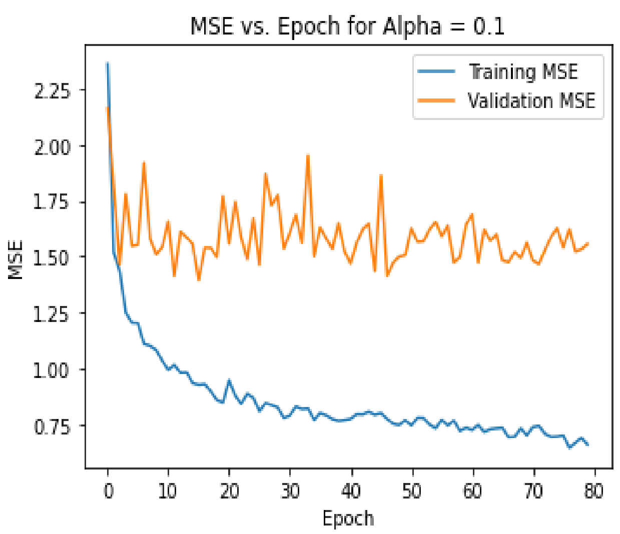
PART III: Using this number of Hidden Units, HP1 = **60**	
When using α = **0.1**	When using α = **0.001**
Trained for **80** epochs	Trained for **80** epochs
Final MSE *from the training set* = `0.6598082884345581`	Final MSE *from the training set* = `0.02374209153498613`
Final MSE *from the validation set* = `1.5553517980572074`	Final MSE *from the validation set* = `0.7557093381769275`

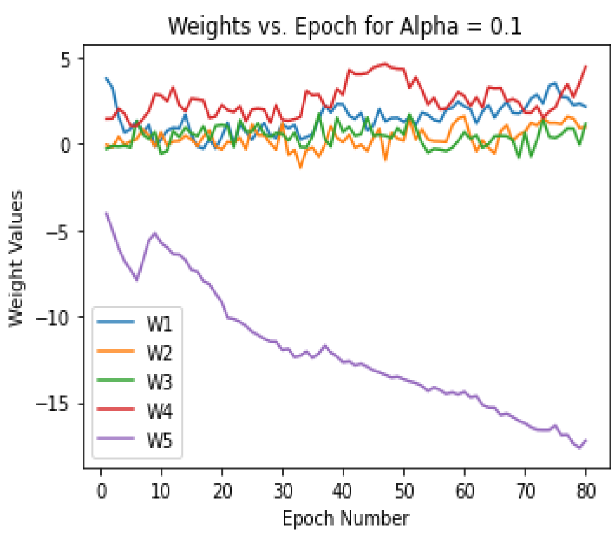
For each of the 2 alpha cases in the table, I have included the below plots:

• The Plot of 5 weights from the network vs. epoch number.

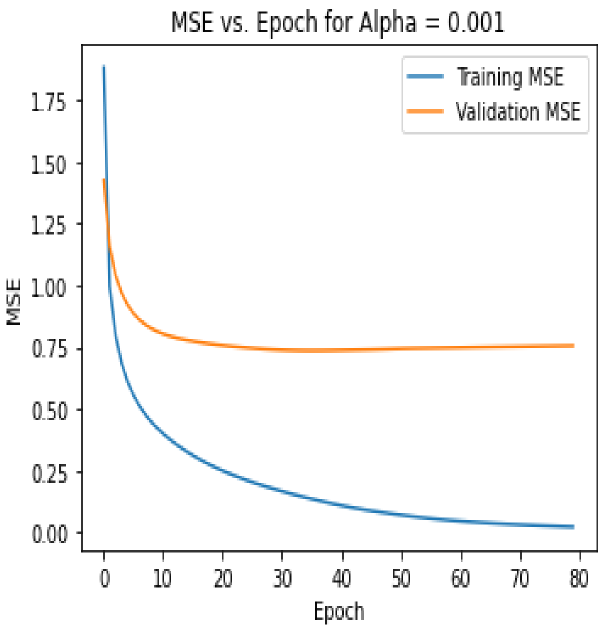
• The Overlay Plot of the "Learning Curve" for Validation and training (i.e., MSE after each epoch of training vs. epoch number)

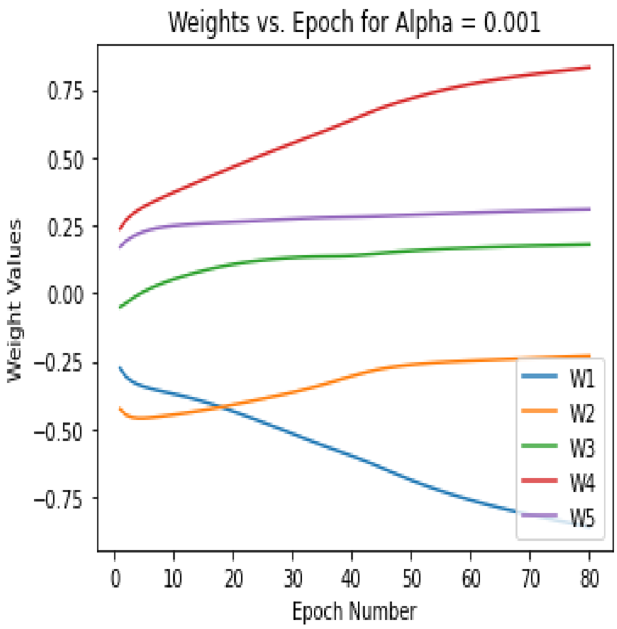
When Alpha = 0.1





When alpha = 0.001





Observations & Conclusions:

Training and Validation MSE: The final MSE values from the training set are significantly higher when using α = 0.1 (0.6598) compared to α = 0.001 (0.0237). This indicates that the lower learning rate (α = 0.001) results in a better overall fit to the training data.

Validation Set Performance: The validation set MSE is also notably higher with α = 0.1 (1.5554) compared to α = 0.001 (0.7557). This suggests that the lower learning rate generalizes better to unseen data, as reflected in the validation set.

Selection of α = 0.001: The choice of α = 0.001 as the best learning rate is supported by the lower MSE values for both the training and validation sets. A lower learning rate often helps the model converge more effectively, avoiding overshooting the minimum and providing a more stable learning process.

Training Epochs: Both training runs were conducted for 80 epochs, allowing the model to learn from the data over an extended period. This ensures that the weights and biases are adjusted to capture the underlying patterns in the data.

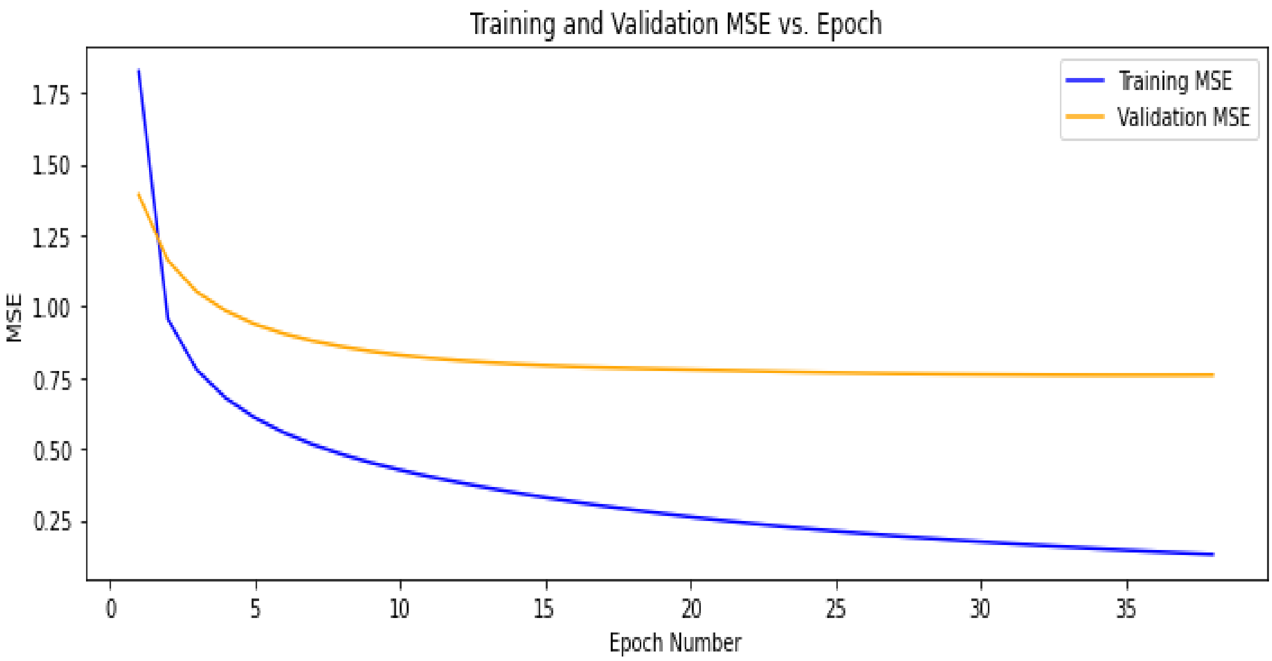
In conclusion, the selection of α = 0.001 is justified by its superior performance in terms of minimizing both training and validation set MSE, indicating better convergence and generalization. The lower learning rate helps prevent the model from learning too quickly, allowing for a more thorough understanding of the complex patterns in the data.

Results and statistics obtained by using <u>EARLY STOPPING</u> with alpha = 0.001:

```
Epoch 1/80, Training MSE: 1.8247478462590943
Epoch 1/80, Validation MSE: 1.3917136769314369
Epoch 2/80, Training MSE: 0.955711714399977
Epoch 2/80, Validation MSE: 1.1609504585038384
Epoch 3/80, Training MSE: 0.7766685811756272
Epoch 3/80, Validation MSE: 1.0509639741238874
Epoch 4/80, Training MSE: 0.6773569229733364
Epoch 4/80, Validation MSE: 0.9836429238115175
Epoch 5/80, Training MSE: 0.6085887289166828
Epoch 5/80, Validation MSE: 0.9372323951156967
Epoch 6/80, Training MSE: 0.5562995273436785
Epoch 6/80, Validation MSE: 0.9035313336895519
Epoch 7/80, Training MSE: 0.5146000550193216
Epoch 7/80, Validation MSE: 0.8780478852598866
Epoch 8/80, Training MSE: 0.48008547599257884
Epoch 8/80, Validation MSE: 0.8580834882975558
Epoch 9/80, Training MSE: 0.4506521964173253
Epoch 9/80, Validation MSE: 0.8420786212584859
Epoch 10/80, Training MSE: 0.4249777523155757
Epoch 10/80, Validation MSE: 0.8291094393364102
Epoch 11/80, Training MSE: 0.4021629821418656
Epoch 11/80, Validation MSE: 0.8185518283155037
Epoch 12/80, Training MSE: 0.38153710795223467
Epoch 12/80, Validation MSE: 0.8099409461152782
Epoch 13/80, Training MSE: 0.36258518614527446
Epoch 13/80, Validation MSE: 0.8028979685310934
Epoch 14/80, Training MSE: 0.3449681634906112
Epoch 14/80, Validation MSE: 0.7971189816705326
Epoch 15/80, Training MSE: 0.328525676020817
Epoch 15/80, Validation MSE: 0.7923638949480166
Epoch 16/80, Training MSE: 0.3131637793812369
Epoch 16/80, Validation MSE: 0.7884455458258609
Epoch 17/80, Training MSE: 0.29876208144744437
Epoch 17/80, Validation MSE: 0.7851957475210196
Epoch 18/80, Training MSE: 0.28520167436560223
Epoch 18/80, Validation MSE: 0.7824429072549064
Epoch 19/80, Training MSE: 0.2723995493076428
Epoch 19/80, Validation MSE: 0.7800049457113506
Epoch 20/80, Training MSE: 0.2603131520263849
Epoch 20/80, Validation MSE: 0.7776869751431199
Epoch 21/80, Training MSE: 0.24893691308504168
Epoch 21/80, Validation MSE: 0.7753398436504626
Epoch 22/80, Training MSE: 0.23826570285060894
Epoch 22/80, Validation MSE: 0.7729682783739978
Epoch 23/80, Training MSE: 0.22826450513644567
```

```
Epoch 23/80, Validation MSE: 0.7706926285033279
Epoch 24/80, Training MSE: 0.21888955692052578
Epoch 24/80, Validation MSE: 0.7685961291448926
Epoch 25/80, Training MSE: 0.21010659536673115
Epoch 25/80, Validation MSE: 0.7666976589803216
Epoch 26/80, Training MSE: 0.2018723730459099
Epoch 26/80, Validation MSE: 0.7650029425885029
Epoch 27/80, Training MSE: 0.19412324198606018
Epoch 27/80, Validation MSE: 0.7635271303621577
Epoch 28/80, Training MSE: 0.18678464736666486
Epoch 28/80, Validation MSE: 0.7622842507515838
Epoch 29/80, Training MSE: 0.17978649572896344
Epoch 29/80, Validation MSE: 0.7612571735952525
Epoch 30/80, Training MSE: 0.17307958425519754
Epoch 30/80, Validation MSE: 0.7603747650714278
Epoch 31/80, Training MSE: 0.16664734127229414
Epoch 31/80, Validation MSE: 0.7595597104111663
Epoch 32/80, Training MSE: 0.16049420672885018
Epoch 32/80, Validation MSE: 0.7588352033035061
Epoch 33/80, Training MSE: 0.1546218679863478
Epoch 33/80, Validation MSE: 0.7582834398112533
Epoch 34/80, Training MSE: 0.14901816241757412
Epoch 34/80, Validation MSE: 0.7579365560070334
Epoch 35/80, Training MSE: 0.14365550736468194
Epoch 35/80, Validation MSE: 0.7577804375577628
Epoch 36/80, Training MSE: 0.1385050337082294
Epoch 36/80, Validation MSE: 0.7577897733987533
Epoch 37/80, Training MSE: 0.13354120300464586
Epoch 37/80, Validation MSE: 0.7579469702972982
Epoch 38/80, Training MSE: 0.12873788606027498
Epoch 38/80, Validation MSE: 0.758245797397033
Early stopping at Epoch 38, Best Validation MSE: 0.7577804375577628
4454
546
0.8908
Accuracy: 89.08%
Final Training MSE for the Best Model: 0.12873788606027498
Final Validation MSE for the Best Model: 0.7577804375577628
```

Overlay of Training & Validation Learning Curve FOR THIS 3rd AND FINAL RUN:



Observations & Conclusions:

The training for the third run was stopped at Epoch 38. The decision to stop training is based on the early stopping method, which monitors the validation Mean Squared Error (MSE) over epochs. The specific reasons for stopping at Epoch 38 and the observations is explained below:

Early Stopping Criteria: Early stopping is employed to prevent overfitting and to choose a model that generalizes well to unseen data. The stopping criterion is often based on observing the trend in the validation MSE. If the validation MSE starts to increase or remains stagnant for a certain number of consecutive epochs, it suggests that the model's performance on new data may not improve further, or it may even degrade.

Validation MSE Trend: In the provided output, the validation MSE decreases consistently up to a certain point and then starts to plateau or slightly increase. This indicates that the model is learning well from the training data initially, but further training may not significantly improve its performance on the validation set.

Accuracy and Generalization: The final accuracy achieved on the validation set is 89.08%. The model has learned to generalize well to unseen data, as evidenced by the relatively low final validation MSE.

Avoiding Overfitting: Stopping the training at Epoch 38 prevents overfitting, as continuing to train beyond this point might lead to the model fitting noise in the training data and result in poorer performance on new data.

Why Early Stop at 38th EPOCH Method:

Resource Efficiency: Early stopping helps save computational resources by avoiding unnecessary training epochs that do not contribute significantly to model improvement.

Generalization: It ensures that the model generalizes well to new data and does not become too specialized in the training set.

Preventing Overfitting: Early stopping is a regularization technique that prevents the model from fitting the training data too closely, which can lead to poor performance on unseen data.

In **conclusion**, stopping training at Epoch 38 is a strategic decision based on monitoring the validation MSE, and it ensures that the model is selected at a point where it balances good fit to the training data and generalization to new data.

TEST SET RESULTS:

For testing by using the final best trained weights and biases after early stopping with the alpha = 0.001,

I have achieved the below statistics:

```
Final Test MSE for the Best Model: 0.4179845156242826
4711 (correct)
289   (incorrect)
0.9422
Accuracy: 94.22%
```

Observations and Conclusions from the Test Set results for the Multi-Class Detector:

Test Mean Squared Error (MSE):

The final test MSE for the best model is 0.4179845156242826. This metric represents the average squared difference between the predicted and actual values for the test set. A lower MSE indicates better model performance.

Accuracy:

The model achieved an accuracy of 94.22% on the test set. This is calculated as the ratio of correctly classified instances to the total number of instances in the test set.

Correct and Incorrect Classifications:

Out of 5000 instances in the test set, 4711 were correctly classified, and 289 were incorrectly classified.

Generalization Performance:

The high accuracy and relatively low test MSE suggest that the model has successfully generalized its learning from the training and validation sets to unseen data. This is a positive indication of the model's ability to recognize and classify different digits.

Comparative Performance:

The test set performance aligns with the model's performance on the training and validation sets, reflecting a consistent ability to generalize across different datasets.

Model Reliability:

The high accuracy on the test set reinforces the reliability and effectiveness of the trained model. It demonstrates the model's capability to make accurate predictions on new and unseen data.

Further Investigation:

While the accuracy is high, further investigation into the types of errors made (e.g., which digits are commonly misclassified) could provide insights into potential areas for model improvement.

In **conclusion**, the test set results indicate that the multi-class detector, trained with a hidden layer of 60 units and a learning rate of 0.001, performs well in accurately identifying and classifying handwritten digits. The model exhibits strong generalization capabilities and is reliable in making predictions on new and unseen data.

CONCLUSIONS

- For both Part 2 & Part 3, the use of validation performance for **early stopping** has likely contributed to a model that strikes a good balance between fitting the training data and generalizing to new, unseen data.
- The **success** of all the models in correctly classifying the majority of test samples reflects the quality of the training process and the **generalization** capabilities of the neural network.
- The choice of alpha = 0.001 as the best learning rate for all three models is **may be the indicative of a well-suited hyperparameter** that offers stability, generalization, and consistent performance across the above architectures to classify handwritten digits from the MNIST Dataset.
- The observation that the TEST MSE is **lower** than the VALIDATION MSE in both Part II and Part III when employing **early stopping** indicates **positive aspects** of the model's **generalization** and the effectiveness of the early stopping strategy.
- The **choice** of the stopping epoch is critical, and the **early stopping** method identifies a point where the model has learned **relevant patterns** without **memorizing noise**. The lower TEST MSE suggests that stopping at this point leads to a model that performs well on new, unseen examples.
- The validation set plays a **crucial role** in determining when to stop training. The fact that the early stopping strategy, informed by validation performance, results in a model with **lower TEST MSE** indicates the utility of using a validation set to **guide** the training process.
- The **final test results** obtained for PART 1, PART II & PART III are 99.37%, 98.26%, 94.22%.