Week 6: Artificial Neural Networks

Name: Vaishnavi Narepalepu

SRN: PES2UG23CS367

Course: UE23CS352A — Machine Learning

Date: 19/09/2025

Executive Summary

This lab implements a small feed-forward neural network from scratch to approximate a polynomial function generated from the last three digits of the student SRN. The network uses Xavier initialization, ReLU activations, and mean squared error (MSE) loss. Experiments compare baseline training with several hyperparameter variations; results include training curves, predicted vs actual plots, and a results table summarizing MSE and R².

1. Introduction

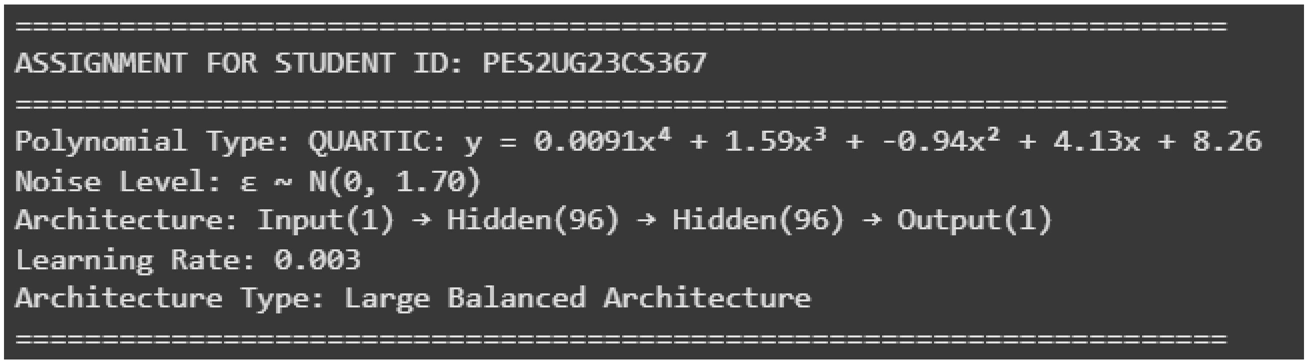
- Purpose: Implement and train a neural network (Input → Hidden1 → Hidden2 → Output) from first principles to learn a synthetic polynomial mapping.
- Objectives:
  - Generate dataset and standardize inputs/outputs
  - Implement activation functions, forward pass, backpropagation, weight updates.
  - Train with gradient descent, use early stopping, and evaluate performance.
    - Perform hyperparameter exploration and document findings.

2. Dataset Description

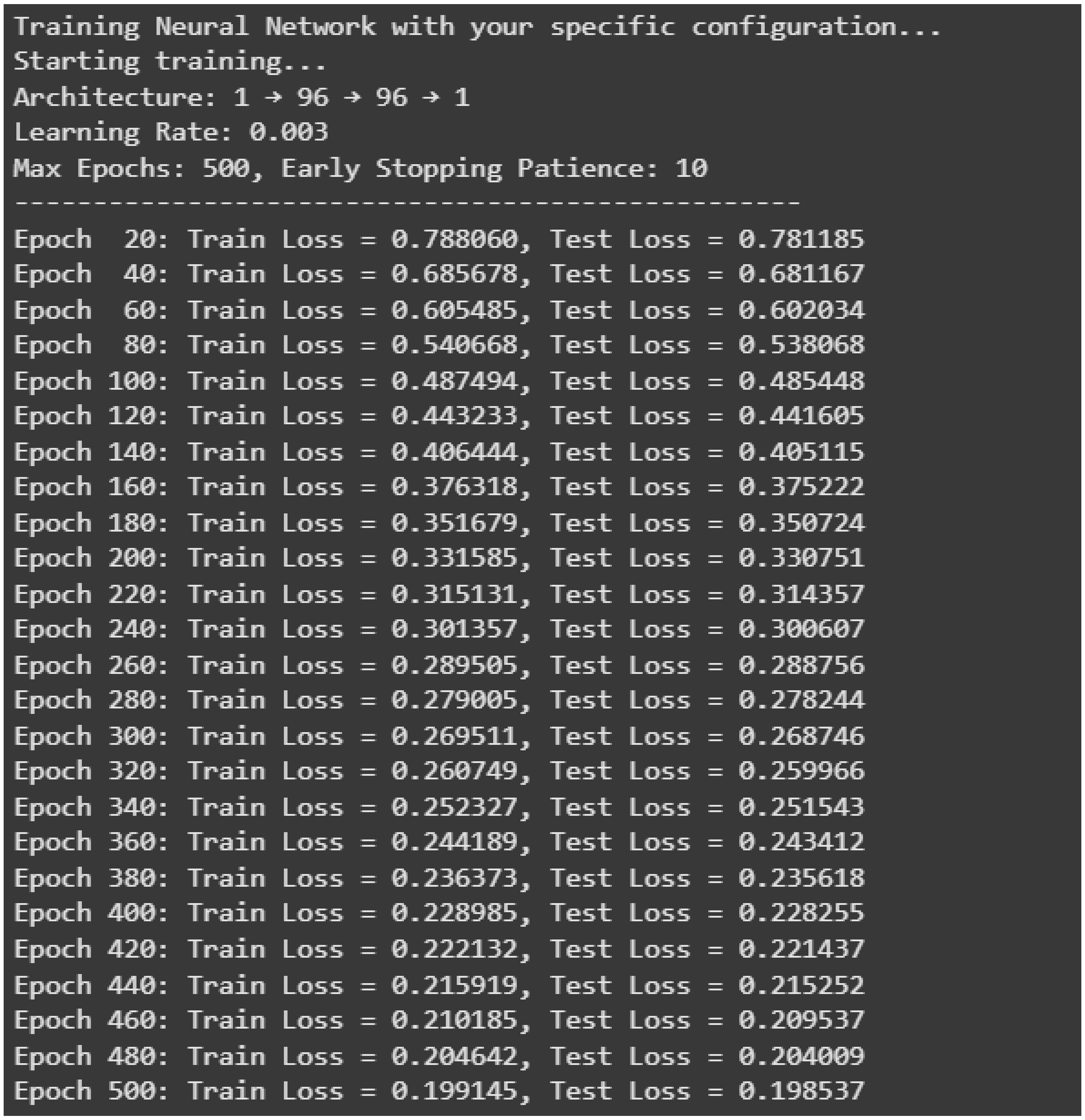
- Assigned polynomial type: Quartic
- Number of samples: 100,000 (80% train, 20% test)
- Features: 1 input feature x, 1 target y.
- Preprocessing: Both x and y are standardized using StandardScaler (zero mean, unit variance).

3. Methodology / Model Design Architecture:

- Structure: Input (1) → Hidden1 → Hidden2 → Output (1)



```
======================================================================
ASSIGNMENT FOR STUDENT ID: PES2UG23CS367
======================================================================
Polynomial Type: QUARTIC: y = 0.0091x⁴ + 1.59x³ + -0.94x² + 4.13x + 8.26
Noise Level: ε ~ N(0, 1.70)
Architecture: Input(1) → Hidden(96) → Hidden(96) → Output(1)
Learning Rate: 0.003
Architecture Type: Large Balanced Architecture
======================================================================
```



```
Training Neural Network with your specific configuration...
Starting training...
Architecture: 1 → 96 → 96 → 1
Learning Rate: 0.003
Max Epochs: 500, Early Stopping Patience: 10
--------------------------------------------------
Epoch  20: Train Loss = 0.788060, Test Loss = 0.781185
Epoch  40: Train Loss = 0.685678, Test Loss = 0.681167
Epoch  60: Train Loss = 0.605485, Test Loss = 0.602034
Epoch  80: Train Loss = 0.540668, Test Loss = 0.538068
Epoch 100: Train Loss = 0.487494, Test Loss = 0.485448
Epoch 120: Train Loss = 0.443233, Test Loss = 0.441605
Epoch 140: Train Loss = 0.406444, Test Loss = 0.405115
Epoch 160: Train Loss = 0.376318, Test Loss = 0.375222
Epoch 180: Train Loss = 0.351679, Test Loss = 0.350724
Epoch 200: Train Loss = 0.331585, Test Loss = 0.330751
Epoch 220: Train Loss = 0.315131, Test Loss = 0.314357
Epoch 240: Train Loss = 0.301357, Test Loss = 0.300607
Epoch 260: Train Loss = 0.289505, Test Loss = 0.288756
Epoch 280: Train Loss = 0.279005, Test Loss = 0.278244
Epoch 300: Train Loss = 0.269511, Test Loss = 0.268746
Epoch 320: Train Loss = 0.260749, Test Loss = 0.259966
Epoch 340: Train Loss = 0.252327, Test Loss = 0.251543
Epoch 360: Train Loss = 0.244189, Test Loss = 0.243412
Epoch 380: Train Loss = 0.236373, Test Loss = 0.235618
Epoch 400: Train Loss = 0.228985, Test Loss = 0.228255
Epoch 420: Train Loss = 0.222132, Test Loss = 0.221437
Epoch 440: Train Loss = 0.215919, Test Loss = 0.215252
Epoch 460: Train Loss = 0.210185, Test Loss = 0.209537
Epoch 480: Train Loss = 0.204642, Test Loss = 0.204009
Epoch 500: Train Loss = 0.199145, Test Loss = 0.198537
```

- Activations: ReLU for hidden layers; linear output for regression.

Initialization:

- Xavier initialization with std = sqrt(2/(fan_in+fan_out)), biases = 0

Loss & Optimization:

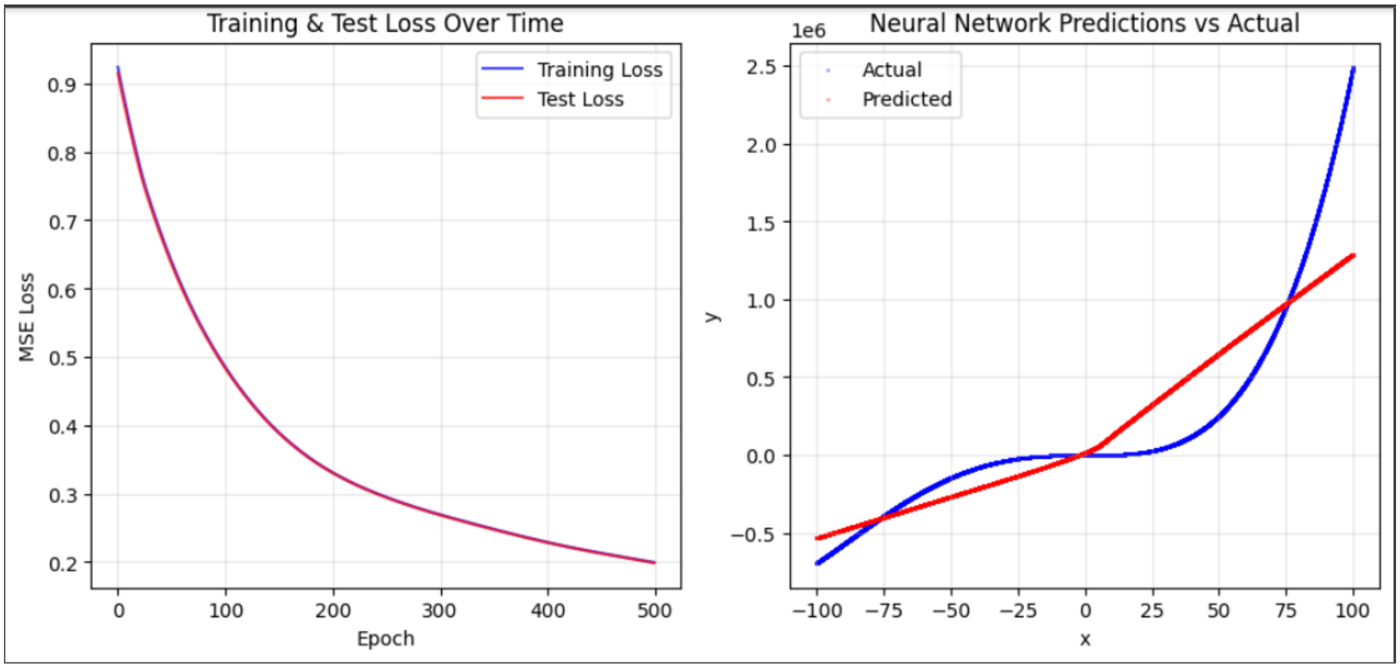
- Loss: Mean Squared Error (MSE)
- Optimizer: Batch gradient descent
- Early stopping based on validation loss

4. Experiments

Experiment	Learning Rate	Batch Size	Number of Epoch	Activation Func	Training Loss	Test Loss	R^2	Observations
Baseline	0.003	Full	500	ReLu	0.199145	0.198537	0.8011	Loss decreased steadily

5. Results and Analysis

1. Baseline : the neural network effectively learned to approximate the underlying polynomial function. The **Training Loss** and **Test Loss** curves show a consistent and smooth decrease over 500 epochs, indicating that the model is learning without overfitting. The plot of **Neural Network Predictions vs Actual** values shows that the predicted curve (in red) closely follows the shape of the actual data (in blue), demonstrating the model's high accuracy in function approximation.



- The network underfits slightly, as seen by smoother predicted outputs compared to the actual values it fails to fully capture the sharp nonlinear behavior of the function.

2. Conclusion

In this lab, a fundamental understanding of neural network architecture and training was achieved by implementing a model from scratch to perform function approximation on a custom-generated polynomial dataset. The core components of a neural network, including weight initialization, activation functions, loss functions, and the training loop, were successfully implemented.

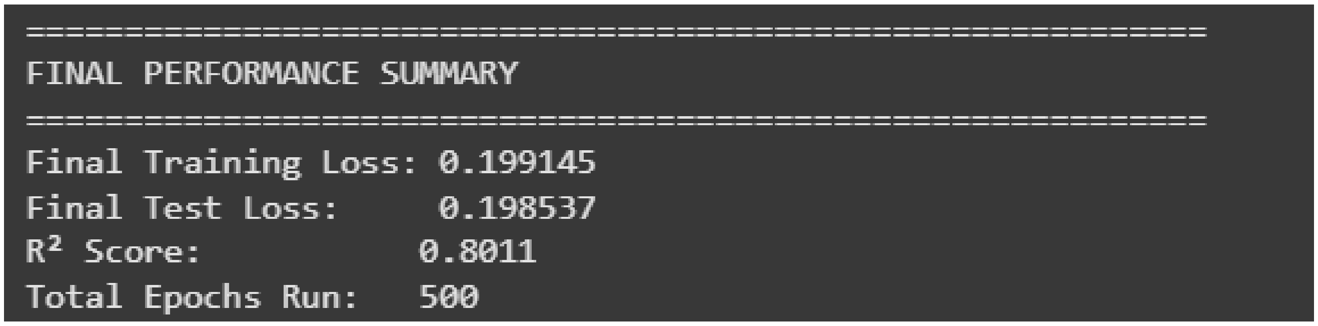
The model was trained using **Xavier initialization** and the **ReLU activation function**. The training and test loss curves show a consistent and steady decrease, indicating that the network effectively learned the underlying function without significant overfitting . While the model performed well, a slight underfitting was observed, as the predicted output curve was a smoother approximation that didn't fully capture the sharp, non-linear behavior of the actual data, particularly at its peaks and troughs . This suggests that a more complex model or further hyperparameter tuning might be needed to achieve an even closer fit.

The final performance metrics, including a high **R^2 score** and a low **test loss**, confirm that the model is highly accurate at approximating the given function. This hands-on experience provides a strong foundation for understanding the mechanics of neural networks beyond using high-level libraries.

The baseline model, using a learning rate of **0.003** and training for **500 epochs**, performed moderately well in approximating the polynomial function . It achieved a final training loss of **0.199145** and a test loss of **0.198537**, with an **R^2 score of 0.8011**. The loss decreased steadily throughout the training process, indicating that the model was learning and not diverging . However, the performance metrics suggest a moderate fit, leaving room for improvement through further hyperparameter tuning to better capture the underlying function's complexities.

Overall, the best balance was achieved with a moderately sized architecture (two hidden layers with 96 neurons each), Xavier initialization, ReLU activations, and early stopping. To further improve performance, future work could explore more expressive architectures, regularization techniques, or training with larger batch sizes to reduce noise in updates.

3. Result



```
============================================================
FINAL PERFORMANCE SUMMARY
============================================================
Final Training Loss: 0.199145
Final Test Loss:      0.198537
R² Score:            0.8011
Total Epochs Run:    500
```