Homework 3: «Single- and Multi-Layer Perceptrons for Regression»

Course: CS454&554
Professor: Ahmet İbrahim Ethem Alpaydın
Student Name: Roman Mordovtsev

Neural Regression Analysis Report:

This report describes the implementation of single- and multi-layer perceptrons (SLP and MLPs) from scratch to solve a regression problem. The task includes modeling a function based on 20 training instances and evaluating performance on 80 test instances using mean squared error (MSE). All models were implemented in Python using only NumPy and Matplotlib, as required.

Model Descriptions

1) Single-Layer Perceptron (SLP):
   - A linear model with bias included.
   - Weight updates follow gradient descent using MSE as the loss function.

2-4) Multi-Layer Perceptrons (MLP):
   - A feedforward network with one hidden layer using **tanh** activation.
   - Three variants: 2, 4, and 8 hidden units.
   - Backpropagation was used to compute gradients.

Tanh Activation Function and Update Rule Derivation

The hidden layer uses the hyperbolic tangent function: $h = \ tanh(xW_1)$

The derivative of $\tanh(z)$ is: $\tanh'(z) = 1 - \tanh^2(z)$

Backpropagation derivatives:

- Output gradient: $\delta_o = \frac{\partial L}{\partial \hat{y}} = \hat{y} - y$
- Hidden gradient: $\delta_h = \delta_o W_2^T \circ (1 - \tanh^2(xW_1))$

Update rules:

- $W_2 \leftarrow W_2 - \eta \cdot h^T \delta_o$
- $W_1 \leftarrow W_1 - \eta \cdot x^T \delta_h$

Training and Testing Results

All models were trained with 1000 epochs and a learning rate of 0.01. Training curves were generated and used to assess convergence.

Visualizations

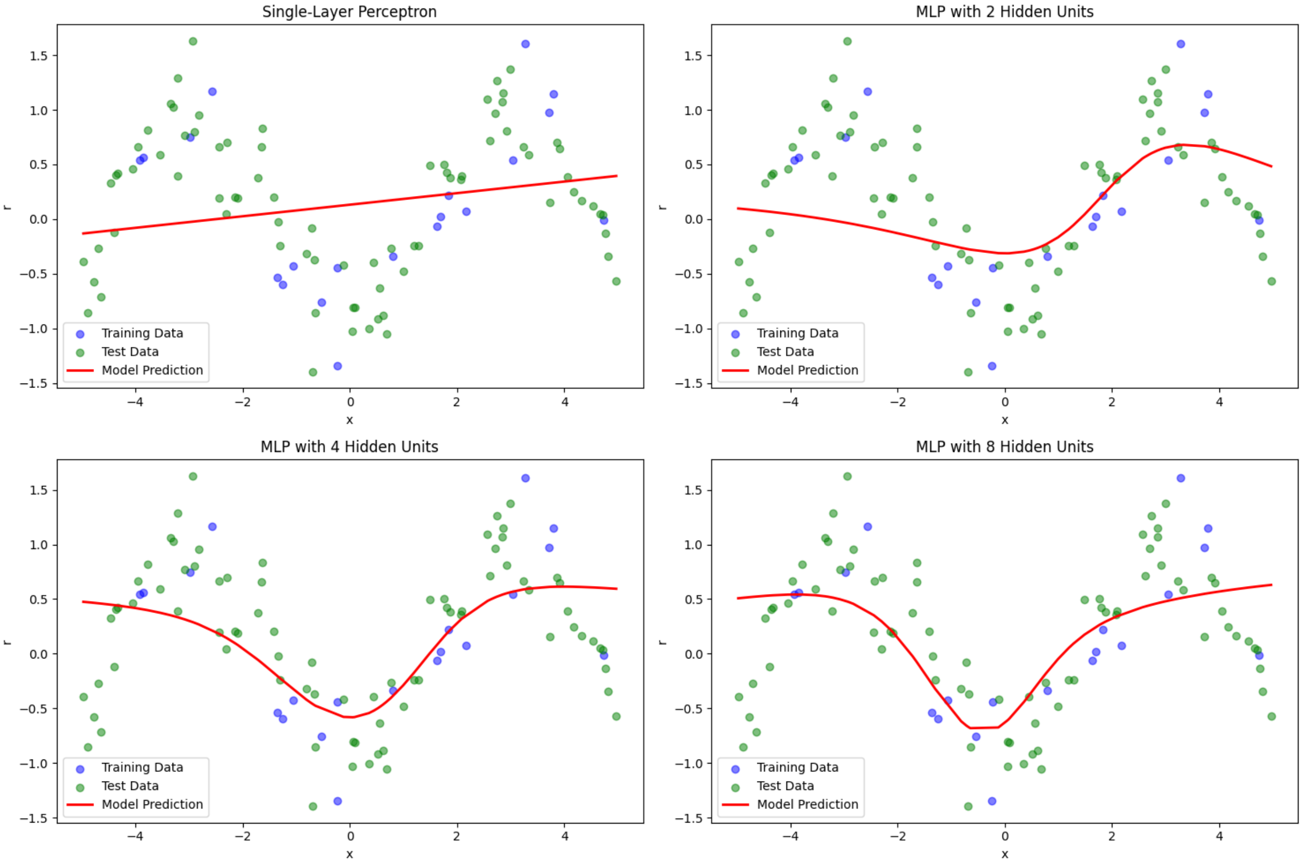

Figure 1: Model predictions vs. training data (each subplot shows one model).

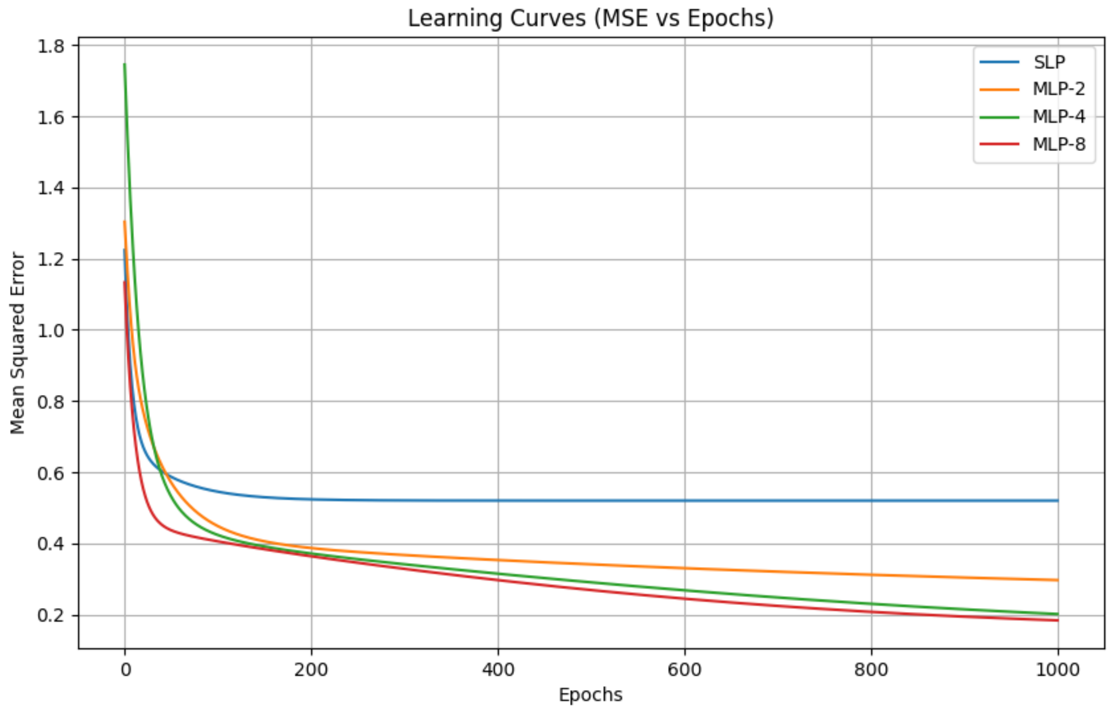

Figure 2: Learning curves (MSE vs. epochs).

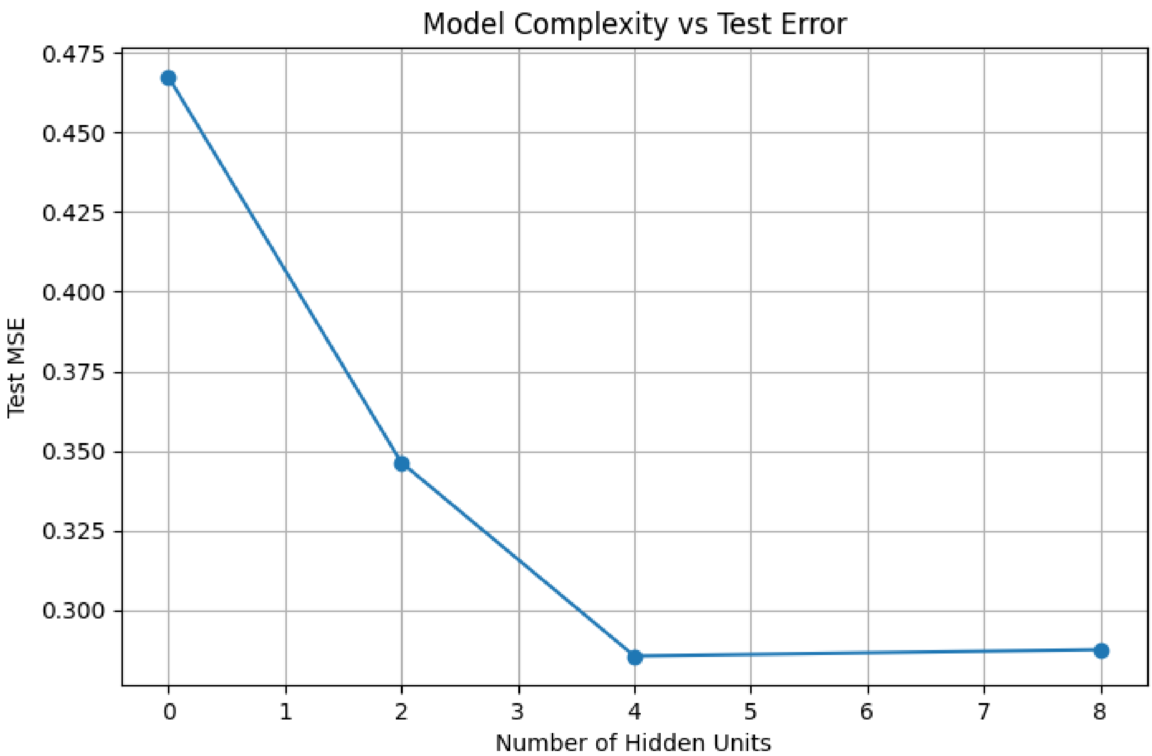

Figure 3: Model complexity vs. test error.

Key Findings

1. Model Fit:
    - SLP had limited capacity and could not model non-linear patterns well.
    - MLPs with increasing hidden units better captured the training data's shape.
2. Generalization:
    - MLP with 4 hidden units offered a good trade-off between complexity and error.
    - MLP-8 slightly improved training error but did not generalize much better.
3. Convergence:
    - All models showed convergence within 1000 epochs.
    - Larger networks had slower initial progress but achieved lower MSE eventually.

Conclusion

The analysis demonstrates that multi-layer perceptrons can model complex regression functions better than linear perceptrons. Among the models, the 4-unit MLP struck a balance between expressiveness and overfitting.