Lab1 – Backpropagation

314552036 周子翔

I. Introduction

This report presents the implementation and evaluation of a simple neural network with two hidden layers trained using the backpropagation algorithm. The network is designed to classify data generated by synthetic datasets. The goal is to understand the role of activation function, weight, biases, and hyperparameters in training a neural network.

II. Implementation Details

A. Network Architecture

The neural network consists of:

- Input Layer: Accepts 2-dimensional input data.
- Hidden Layers: Two hidden layers with configurable numbers of neurons.
- Output Layer: A single neuron for binary classification.
- Weights and Biases: Initialized using He initialization to improve convergence.

B. Activation Functions

i. Sigmoid:

1. Sigmoid: $\sigma(x) = \frac{1}{1+e^{-x}}$
2. Derivative: $\sigma' = \sigma(x) * (1 - \sigma(x))$

ii. ReLU

1. ReLU: $f(x) = max(0,\ x)$
2. Derivative: $f'(x) = \begin{cases} 0\ for\ x < 0 \\ 1\ for\ x \geq 0 \end{cases}$

C. Backpropagation

The backpropagation algorithm computes gradients of the loss function

with respect to weights and biases:

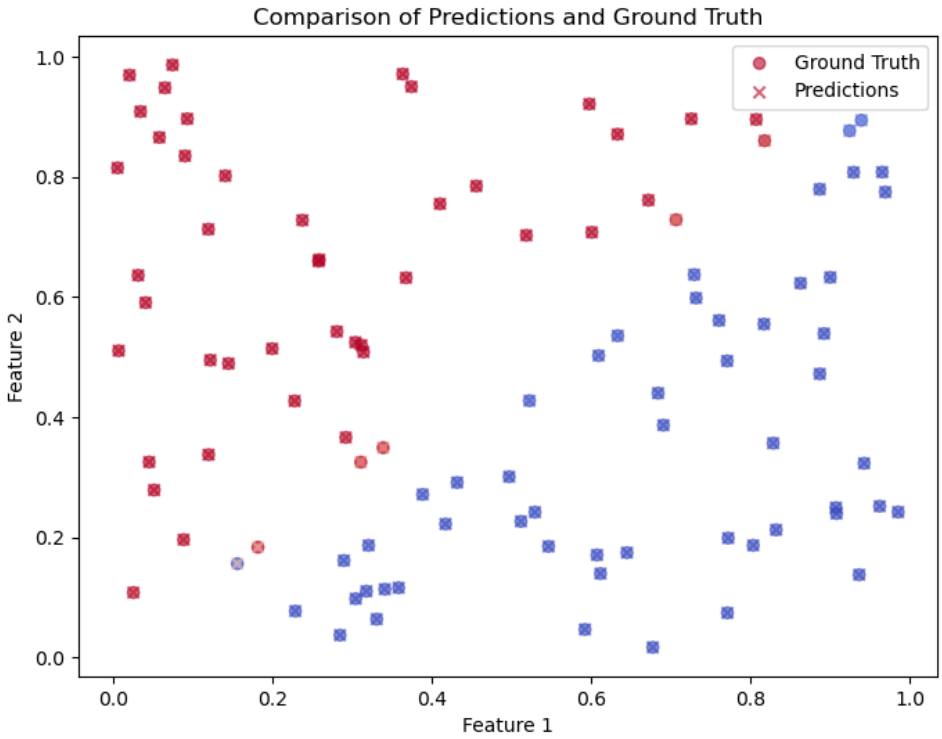
$$dW_i = \frac{1}{m} A_{i-1}^T \delta_i$$

$$db_i = \frac{1}{m} \sum_{k=1}^{m} \delta_{i,k}$$

$$\delta_i = (\delta_{i+1} W_{i+1}^T) \cdot \sigma'(Z_2)$$

III. Experimental Results

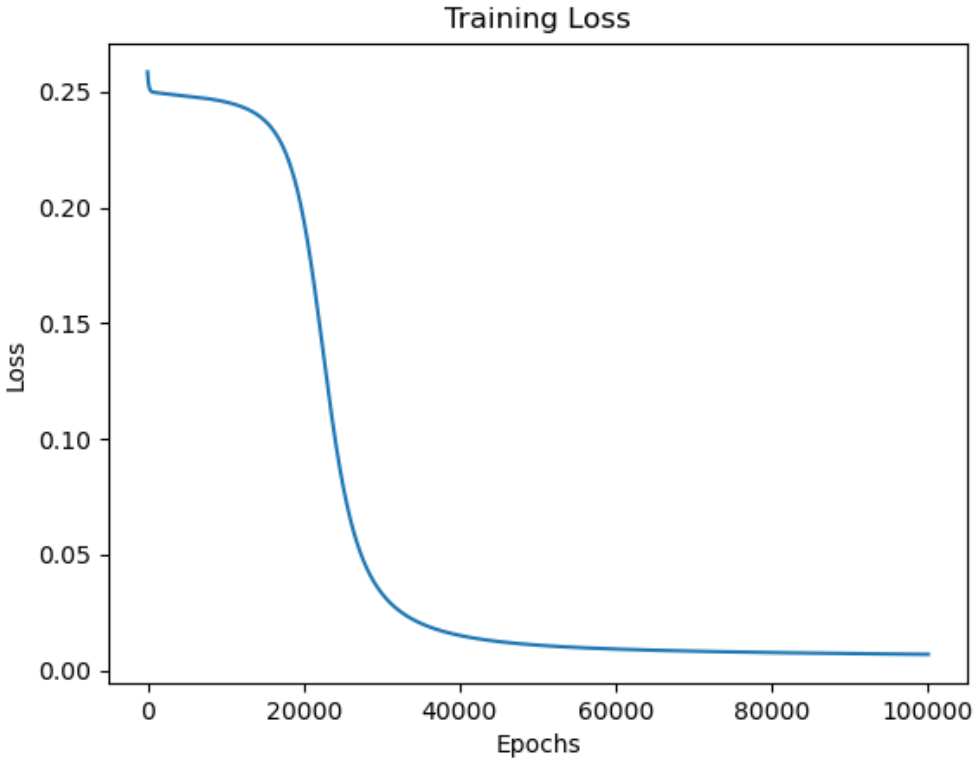
A. Screenshot and comparison figure



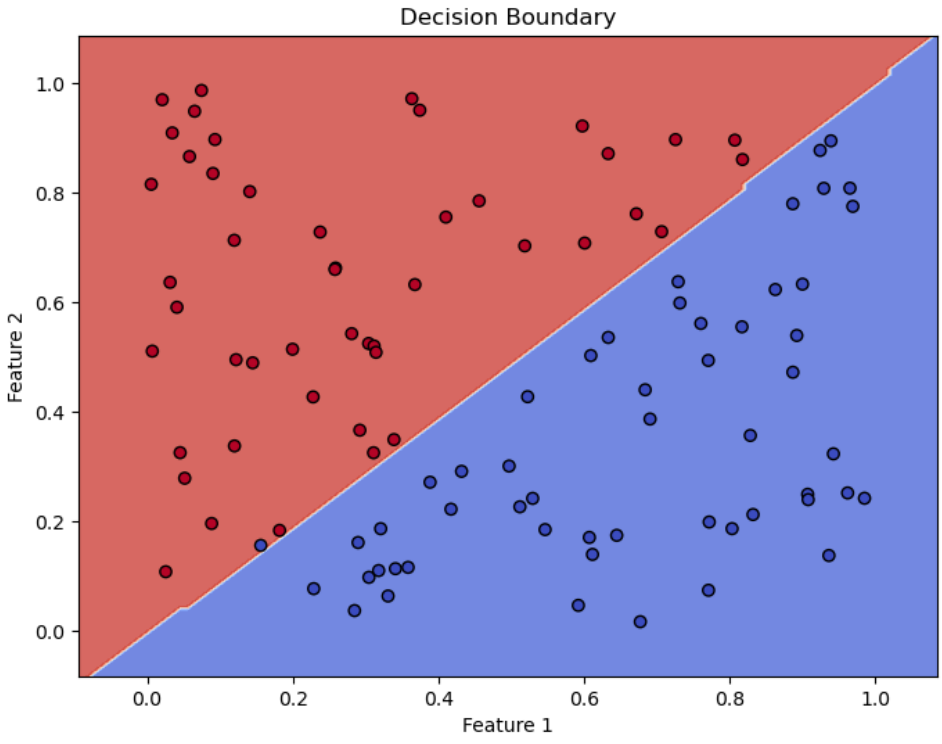
B. Show the accuracy of your prediction

Accuracy: 99.00%

C. Learning curve (loss-epoch curve)



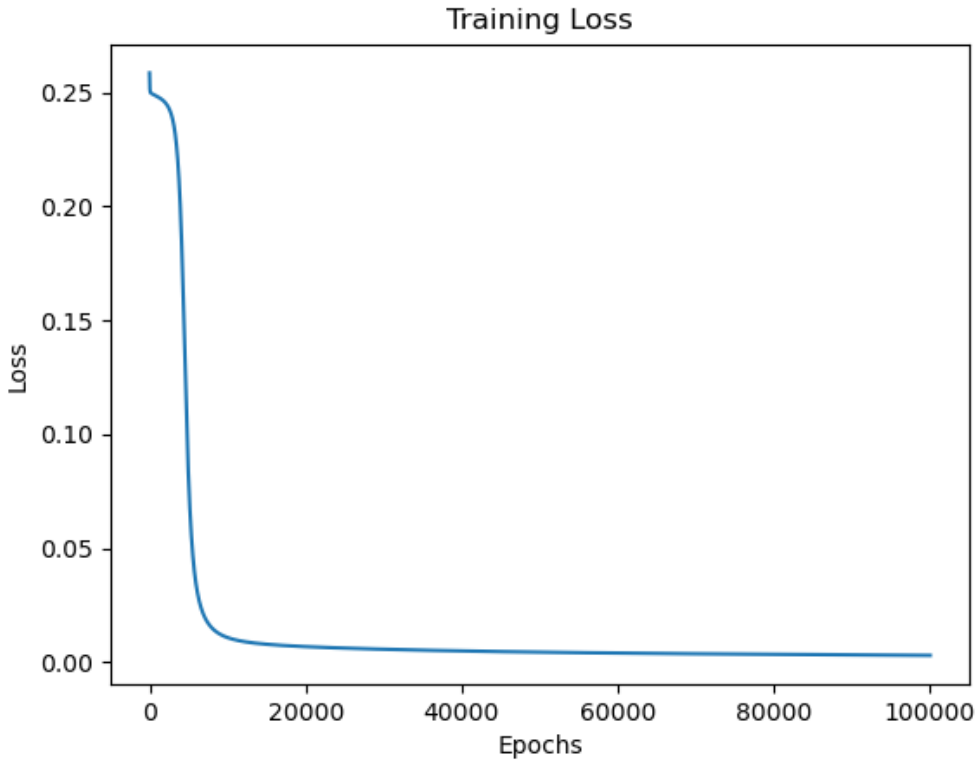
D. Anything you want to present



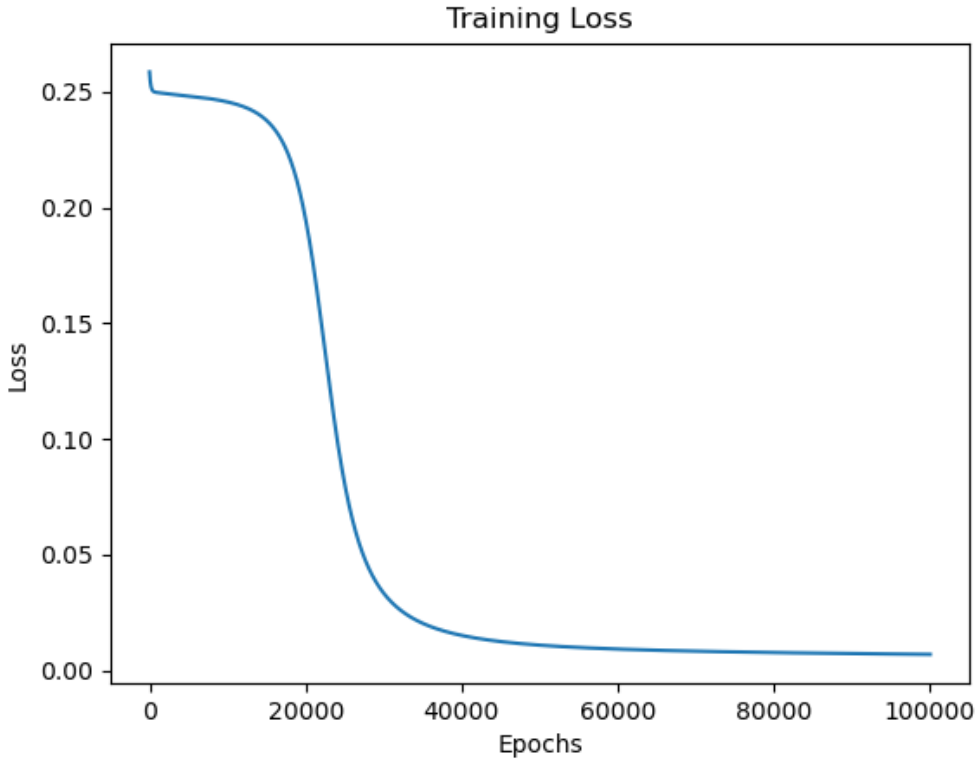
IV. Discussions

A. Try different learning rates

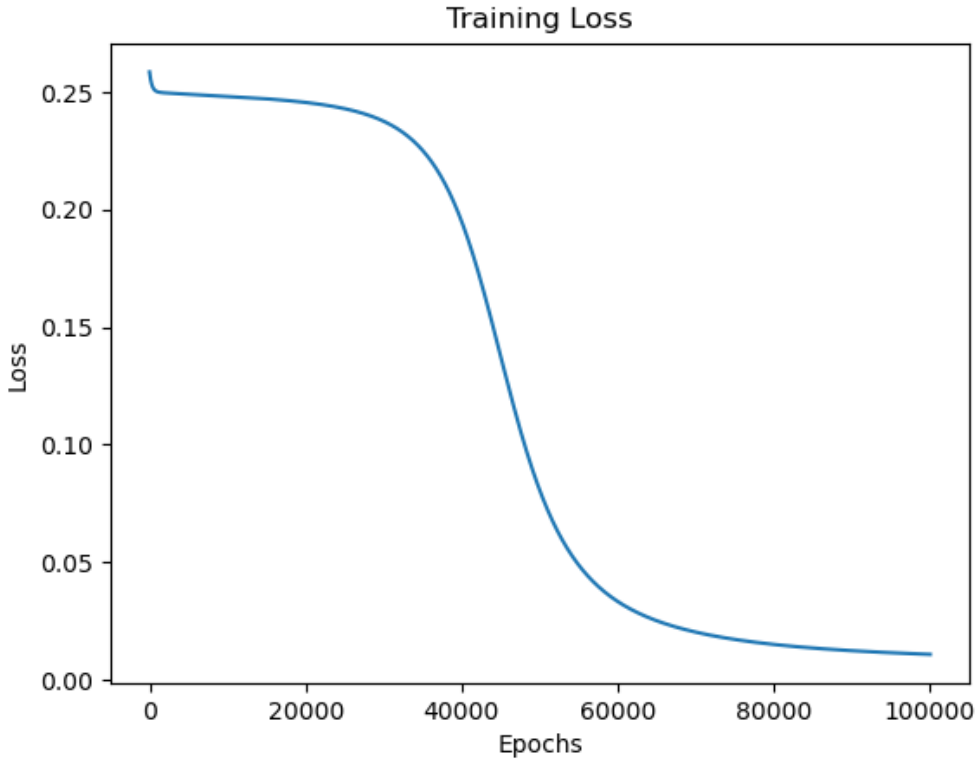
i. Learning rate = 0.05



ii. Learning rate = 0.01

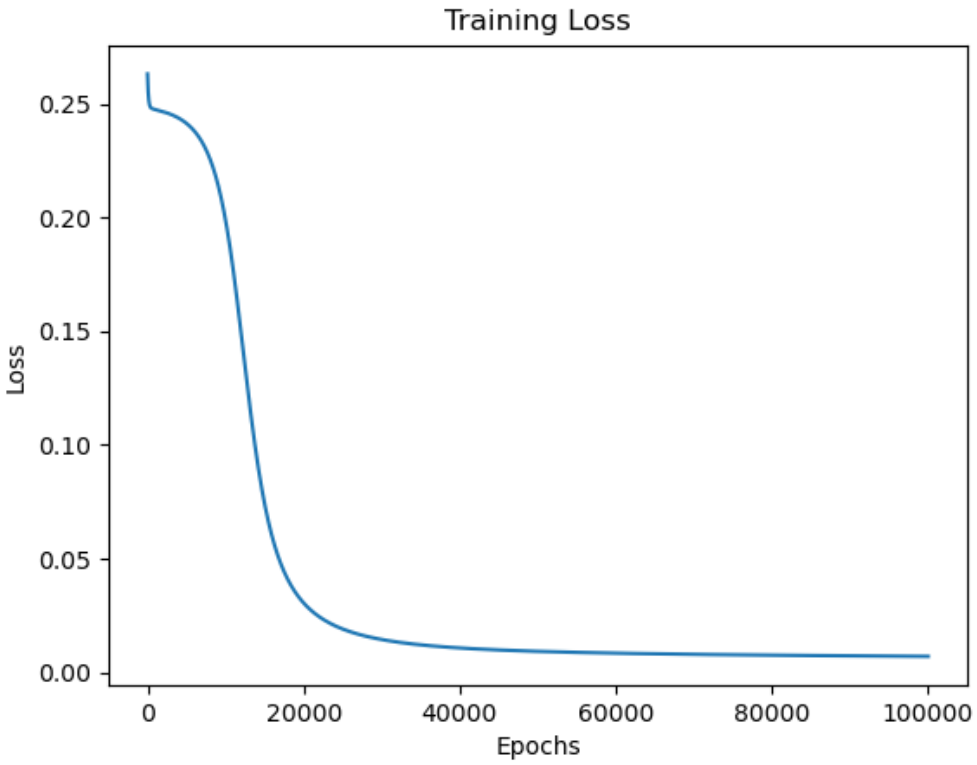


iii. Learning rate = 0.005

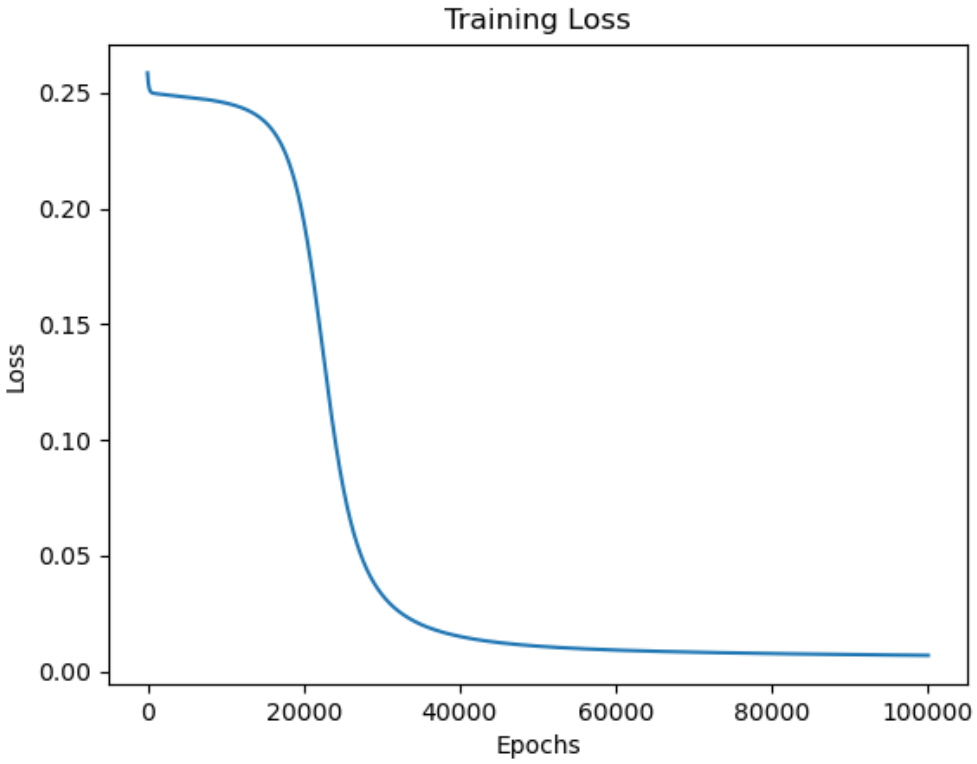


B. Try different numbers of hidden units

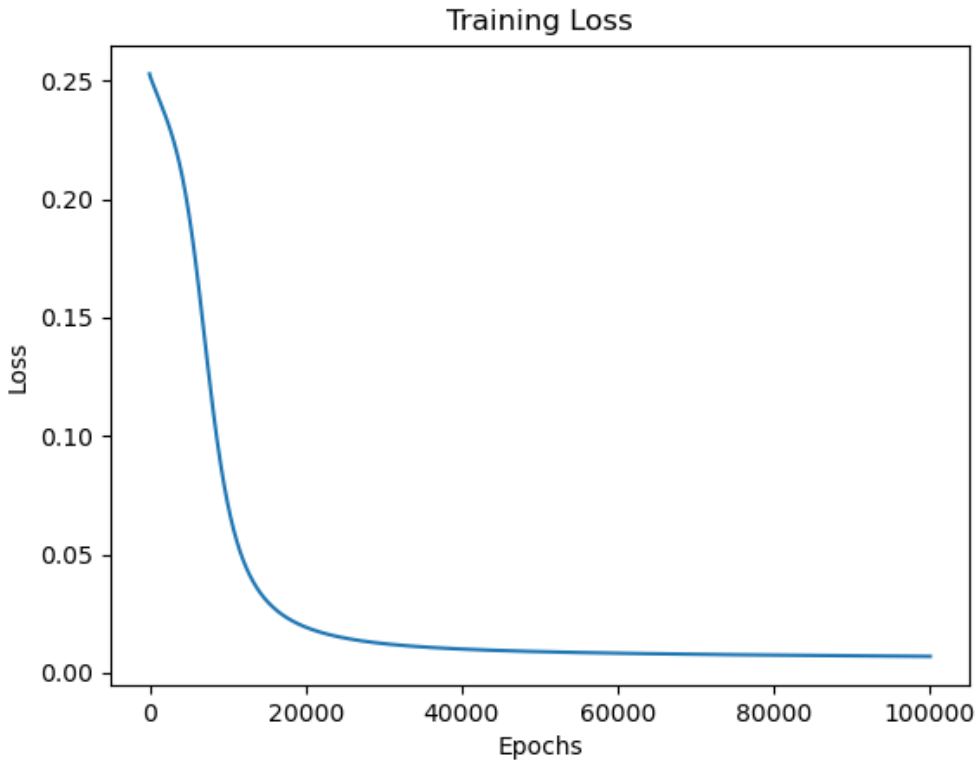
i. Hidden units = (4,3)



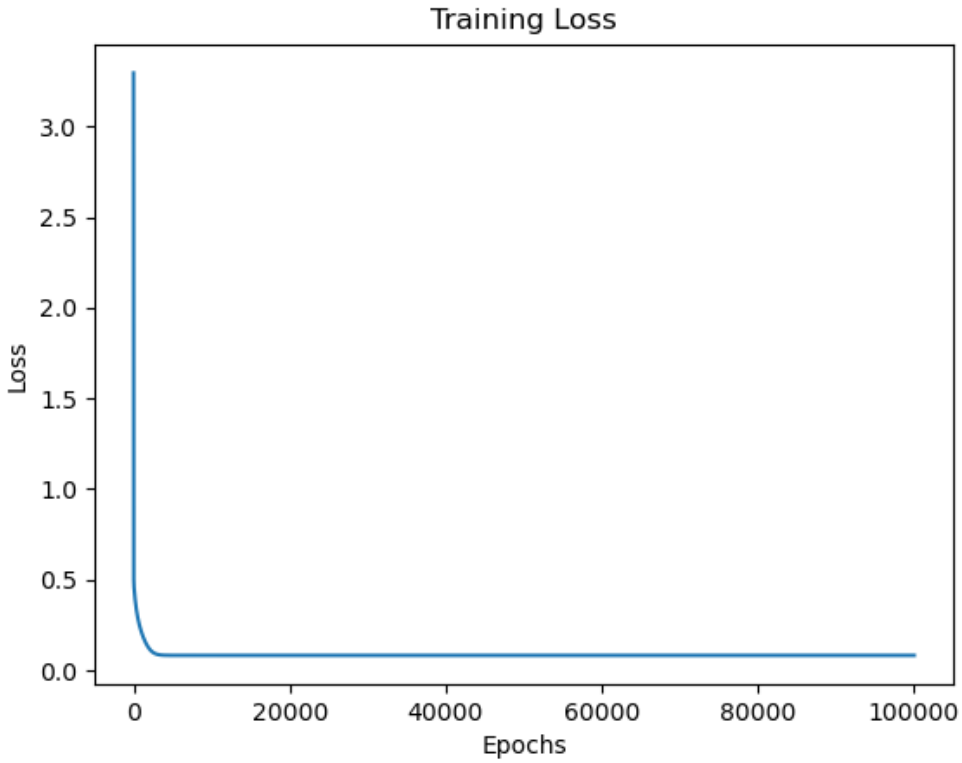
ii. Hidden units = (6,5)



iii. Hidden units = (8,7)



C. Try without activation functions



D. Extra Implementation Discussions

V. Questions

A. What are the purposes of activation functions?

Activation functions introduce non-linearity into the network, enabling it to model complex relationships in the data. Without them, the network would behave like a linear model regardless of its depth.

B. What if the learning rate is too large or too small?

Too Large: The network may fail to converge or oscillate around the minimum.

Too Small: The network converges very slowly, requiring more epochs to reach an optimal solution.

C. What are the purposes of weights and biases in a neural network?

Weights: Determine the importance of each input feature and connection in the network.

Biases: Allow the network to shift the activation function, enabling it to fit the data better.