CS 615 - Deep Learning

Assignment 5 - MLPs
Winter 2022
By: Willie Hood

03/07/22

1 Theory

1. (10pts) In class and the lecture notes, we provided the gradient of the *tanh* function without actually walking through the derivation. For this assignment's only theory question, show the work on how the partial derivative of the tanh function, $g(z) = \frac{e^z - e^{-z}}{e^z + e^{-z}}$, is $\frac{\partial g(z)}{\partial z} = (1 - g^2(z))$

   $g(z) = \frac{e^z - e^{-z}}{e^z + e^{-z}} = \frac{(e^z + e^{-z})(e^z + e^{-z}) - (e^z - e^{-z})(e^z - e^{-z})}{(e^z + e^{-z})^2} = 1 - \frac{(e^z - e^{-z})^2}{(e^z + e^{-z})^2} = 1\text{-}\tanh^2(z) = (1 - g^2(z))$

2 Multi-Class Logistic Regression

The architecture is:

$$\text{Input} \rightarrow \text{Fully-Connected} \rightarrow \text{Sigmoid Activation} \rightarrow \text{Log Loss Objective}$$

For this architecture I used ADAM I had 100 epochs and a learning rate of 0.001. I had 10 outputs from my fully connected layer. I also used one hot encoding on my targets.

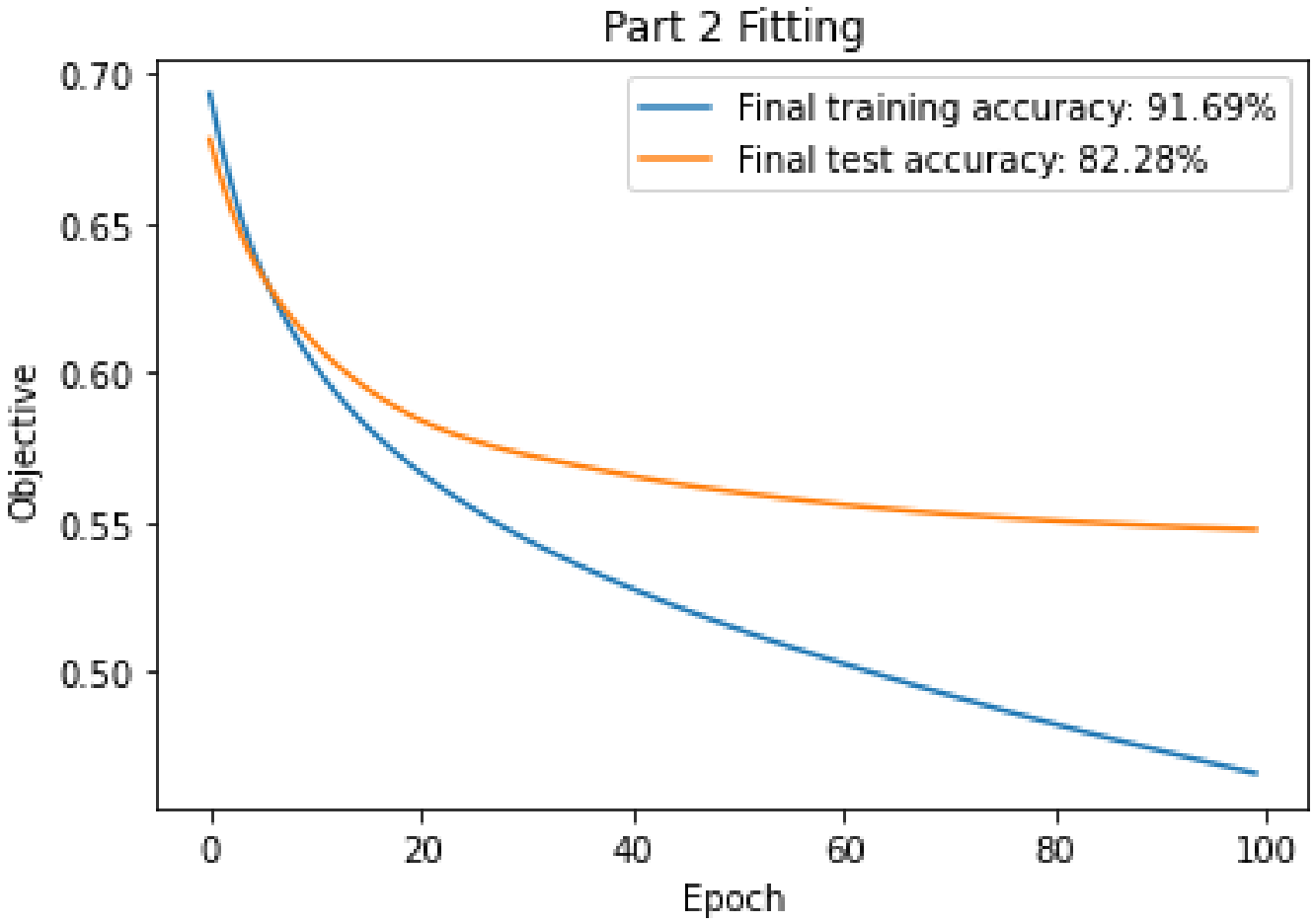

Figure 1: Architecture

3 Artificial Neural Networks

Architecture	Training Accuracy	Validation Accuracy
First Architecture		
Input:	99.5%	90.29%
FC(Output: 450):		
Relu:		
FC(Output: 10):		
Sigmoid		
Logloss		
Second Architecture		
Input:	100%	90.49%
FC(Output: 450)		
Relu		
FC(Output: 10)		
Softmax		
Logloss		
Third Architecture		
Input	81.68%	74.17%
FC(Output: 300)		
Linear		
FC(Output: 10)		
Sigmoid		
Cross Entropy		

Table 1: All Configurations used ADAM Architecture 1: 120 epochs and a learning rate of 0.001; Architecture 2: 100 epochs 0.001 learning rate; Architecture 3: 100 epochs 0.1 learning rate. (Note: it took hours for this to run my architecture is very slow)

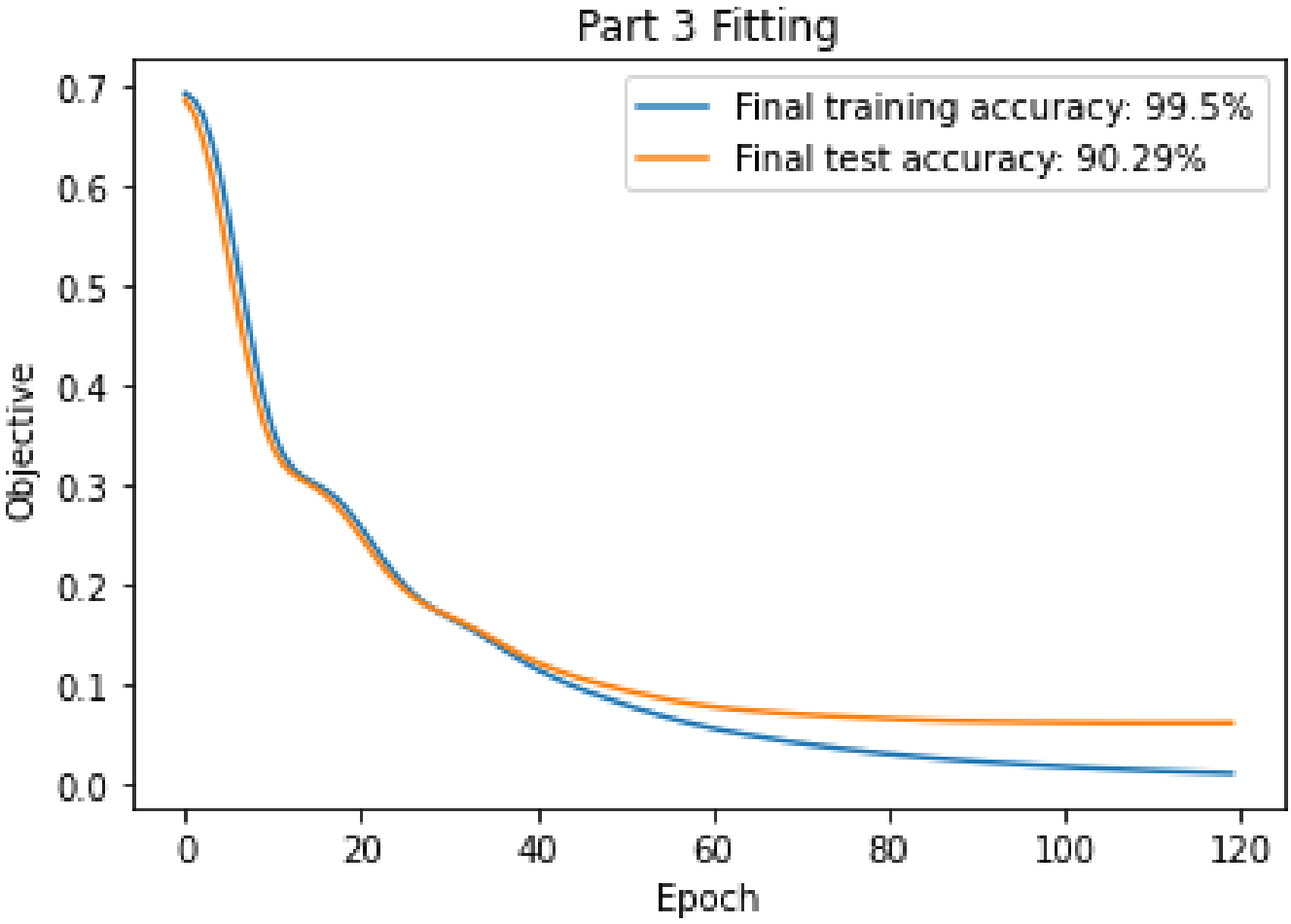

Figure 2: Architecture 1

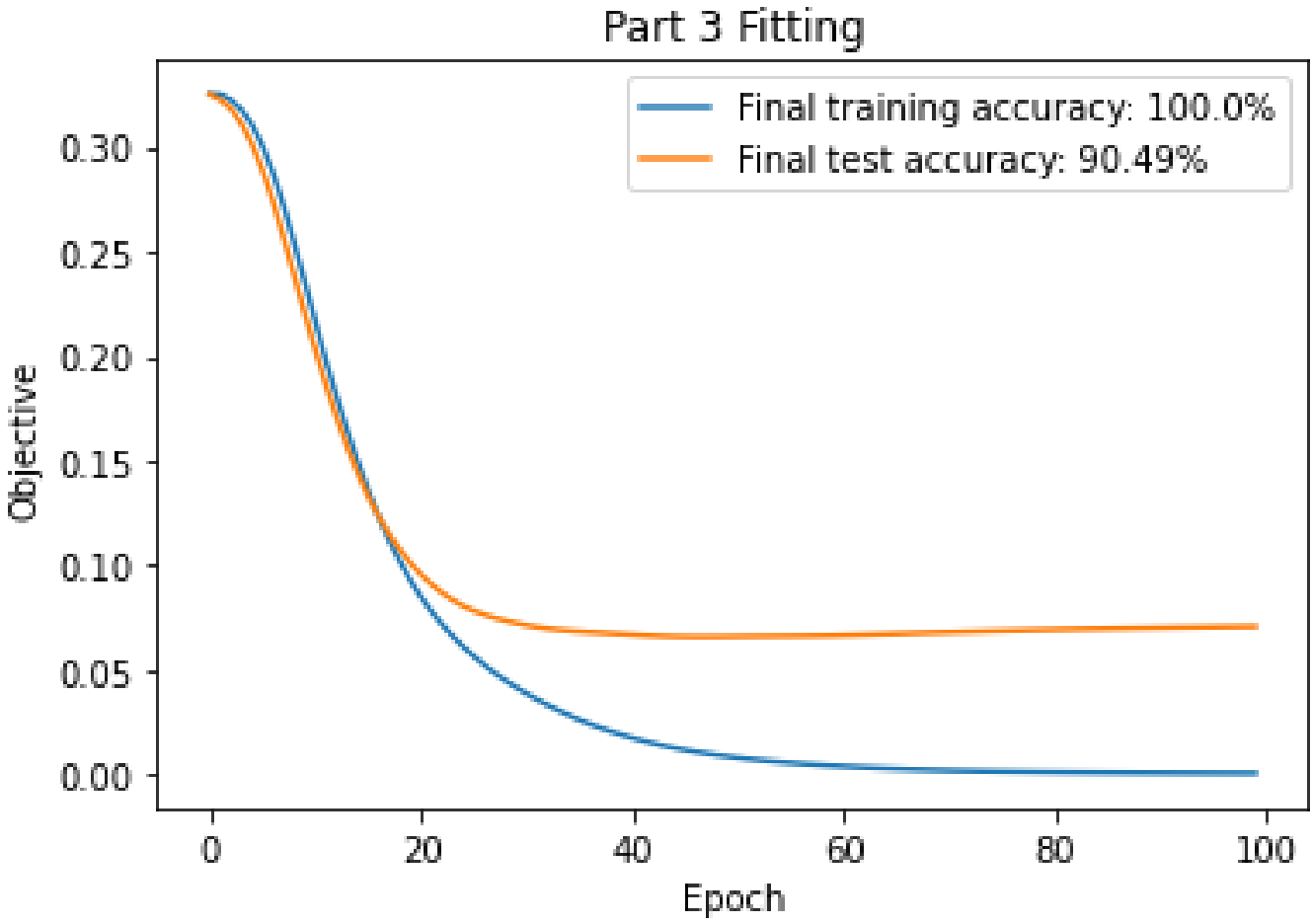

Figure 3: Architecture 2

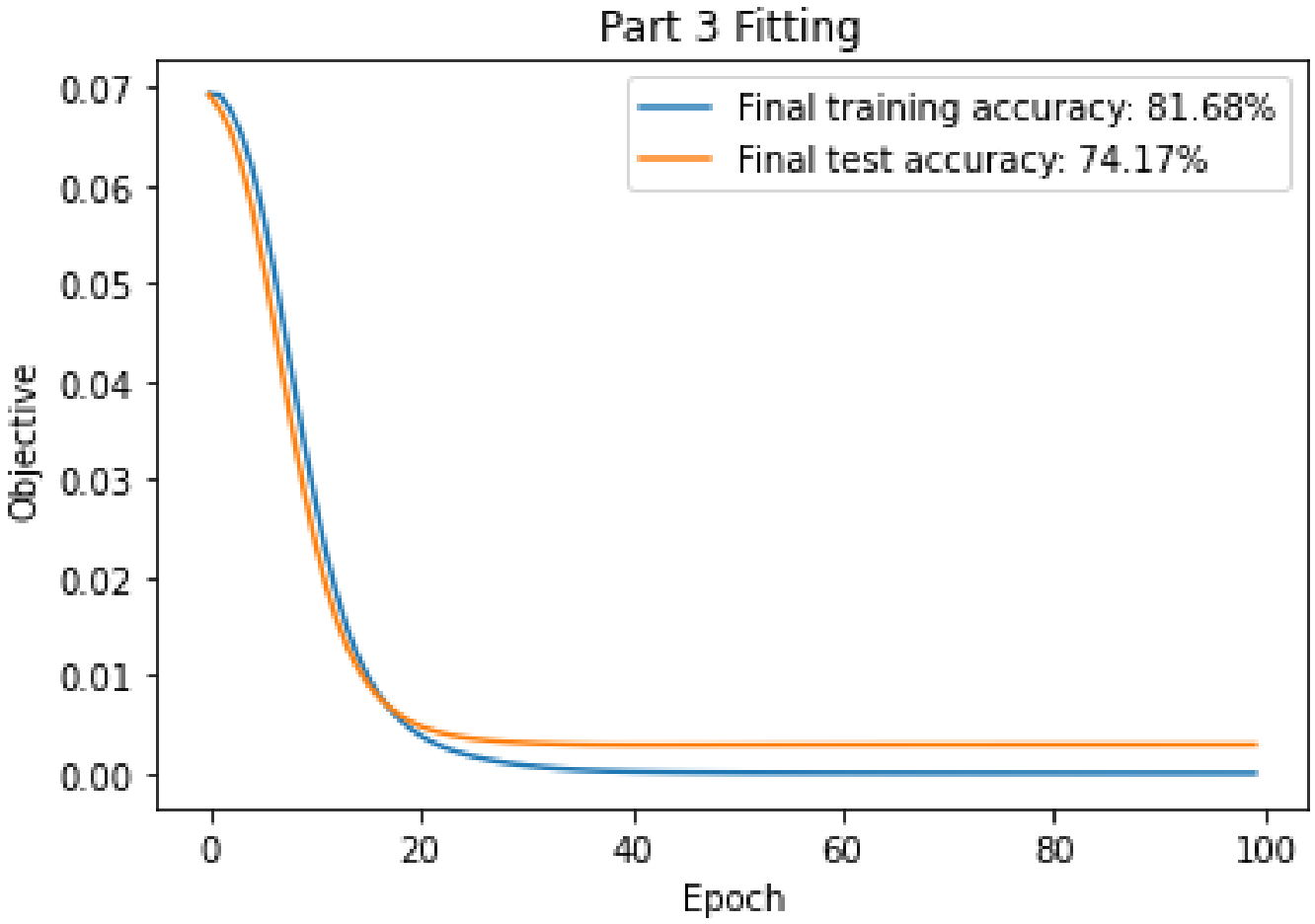

Figure 4: Architecture 3

4 Multi-Layer Perceptron

Architecture	Training Accuracy	Validation Accuracy
First Architecture		
Input:	100%	83.88%
FC(Output: 400):		
Relu:		
FC(Output: 250):		
Tan:		
FC(Output: 10):		
Softmax:		
Logloss		
Second Architecture		
Input:	98.60%	78.48%
FC(Output: 400)		
Tan		
FC(Output: 250)		
Sigmoid		
FC(Output: 10)		
Softmax		
Cross Entropy		
Third Architecture		
Input	100%	81.28%
FC(Output: 400)		
Tan		
FC(Output: 250)		
Relu		
FC(Output: 10)		
Sigmoid		
Logloss		

Table 2: All Configurations used ADAM Architecture 1: 2500 epochs and a learning rate of 0.001; Architecture 2: 1000 epochs 0.1 learning rate; Architecture 3: 150 epochs 0.1 learning rate. (Note: it took days for this to run my architecture is very slow)

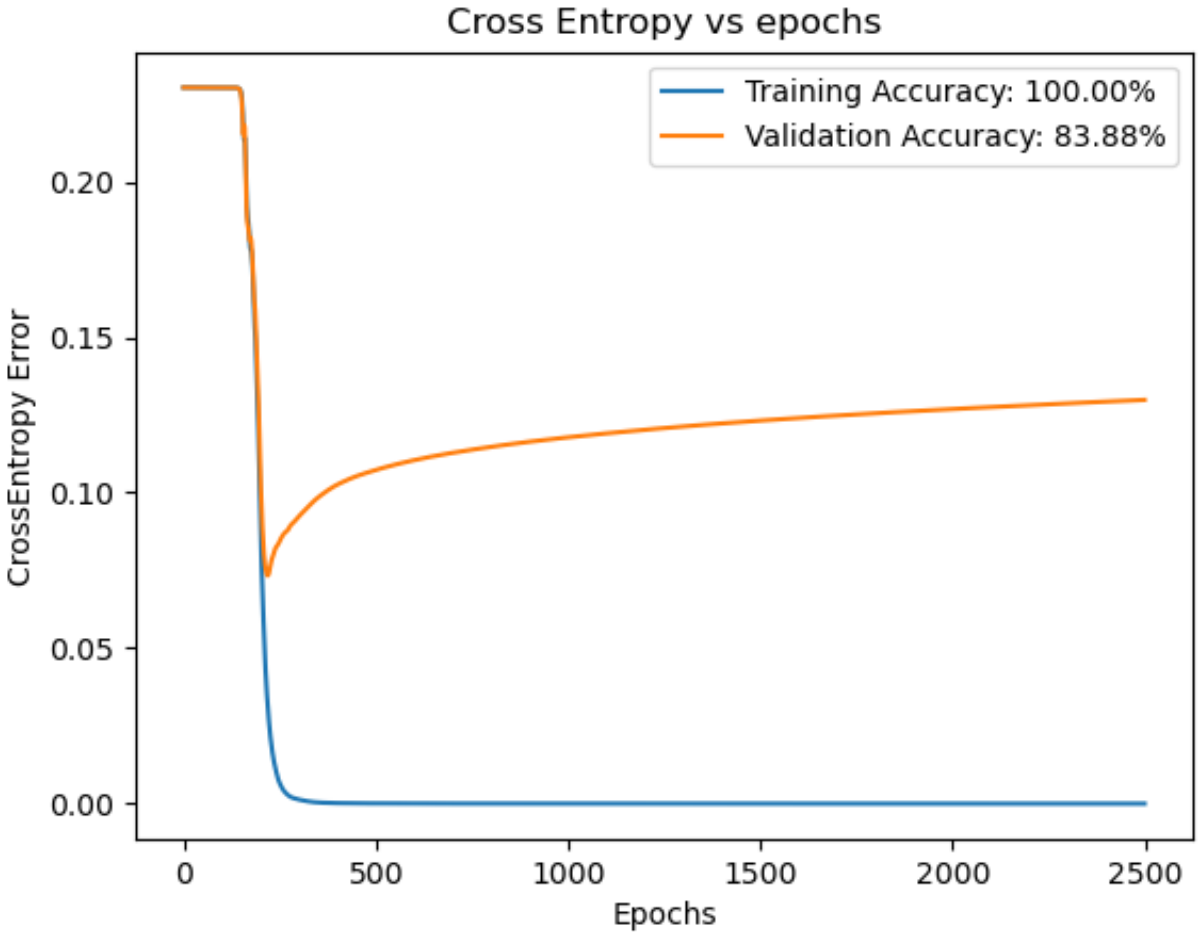

Figure 5: Architecture 1

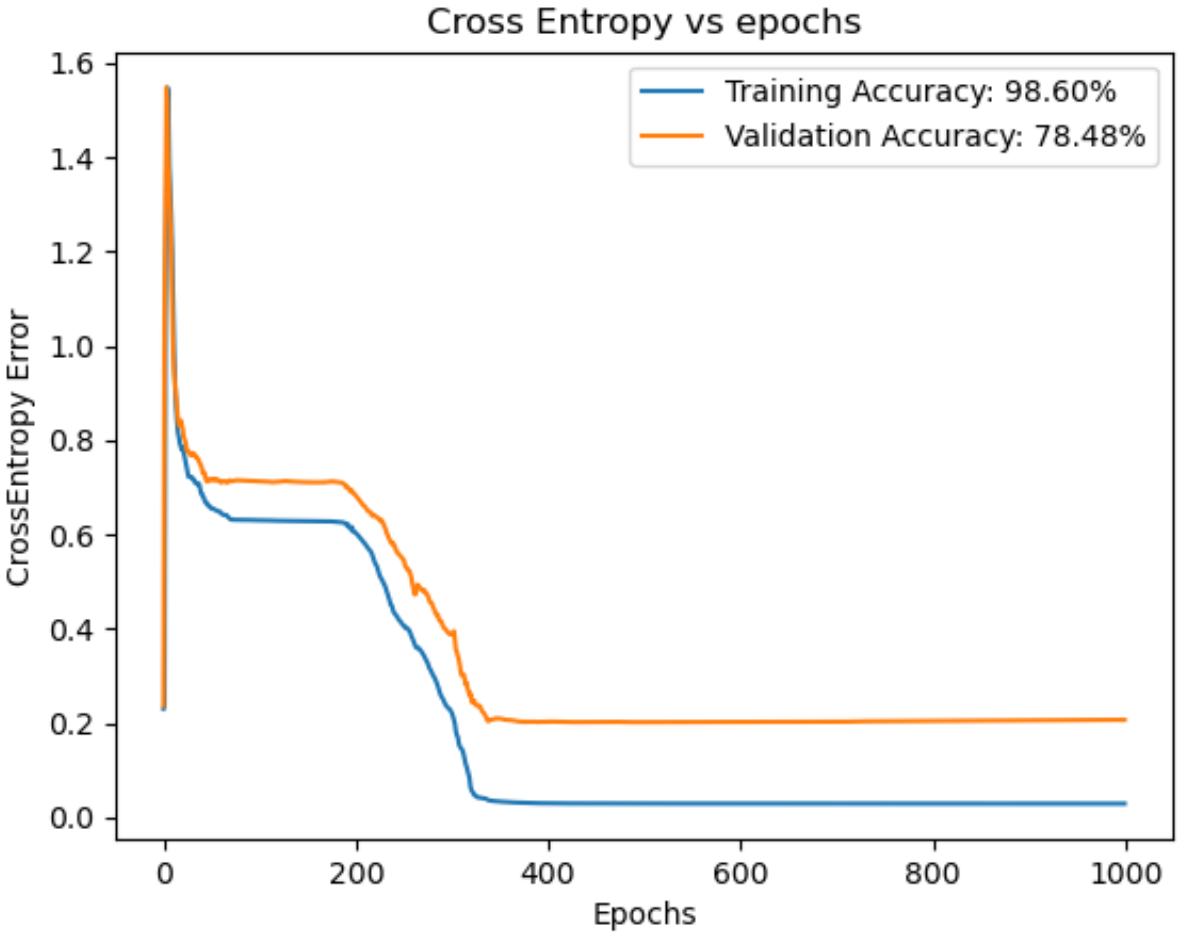

Figure 6: Architecture 2

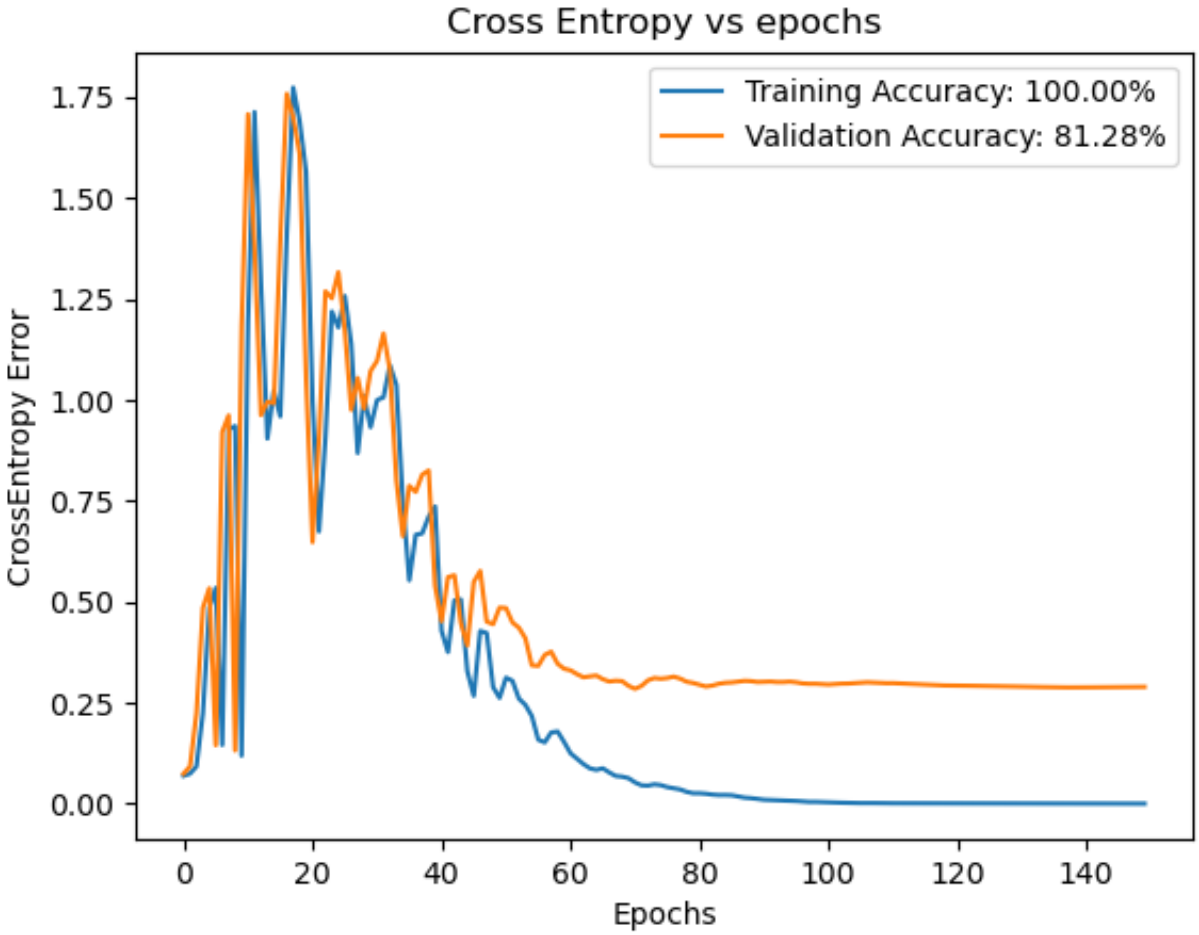

Figure 7: Architecture 3