Nathan Butler
ROB 537 | HW1 Report
10/10/2023

Homework 1: Neural Networks

Nathan Butler | butlnath@oregonstate.edu

Training Set 1 Discussion

Unless otherwise noted, all training for part one was conducted using 25 hidden units, a step size of 0.001, and 100 epochs.

1. Hidden Units

It can be observed in Figure 1 that greater numbers of hidden units yield increased accuracy and reduced loss during training. This is likely due to the ability of a larger network to capture more information about relationships between input data and the output classification. Greater numbers of internal units allow for the tuning of many more parameters.

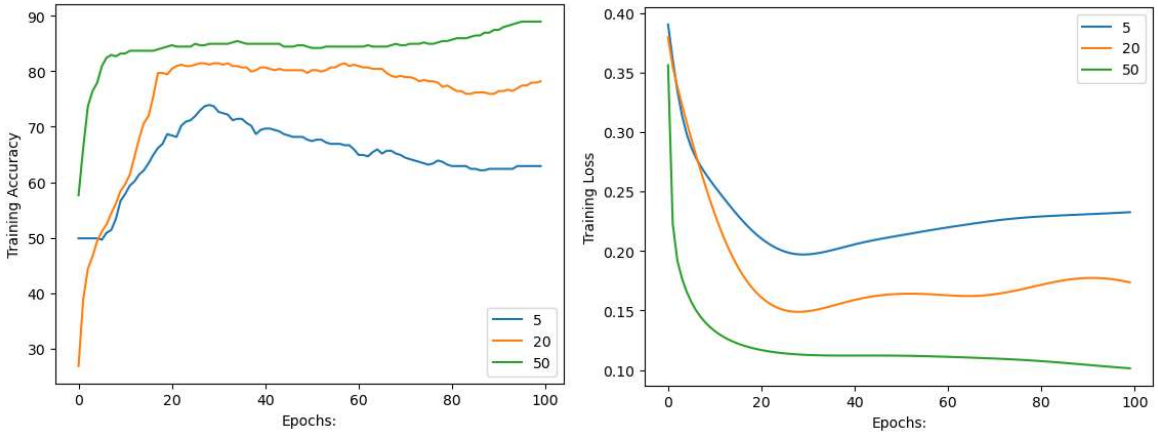

Figure 1. Accuracy and loss over training epochs for 5, 20, and 50 hidden layer units

These observations are further reinforced by the testing accuracy results stored in Table 1, which shows an accuracy of over 85% for a network with 50 hidden units and an accuracy of only 60.65% for a network with 5 hidden units.

Table 1. Testing accuracy for 5, 20, and 50 hidden layer units

# Hidden Units	Testing Accuracy
5	60.65%
20	76.44%
50	85.21%

2. Training Time

Contrary to the results obtained with increased hidden layers, an increase in training time is shown in Figure 2 to not necessarily yield better performance. In fact, the opposite effect was observed for training

set 1. After an initial peak in accuracy around 60 epochs, a downward trend appears in each training cycle. These results may be impacted by the other network parameters, such as the limited number of hidden units.

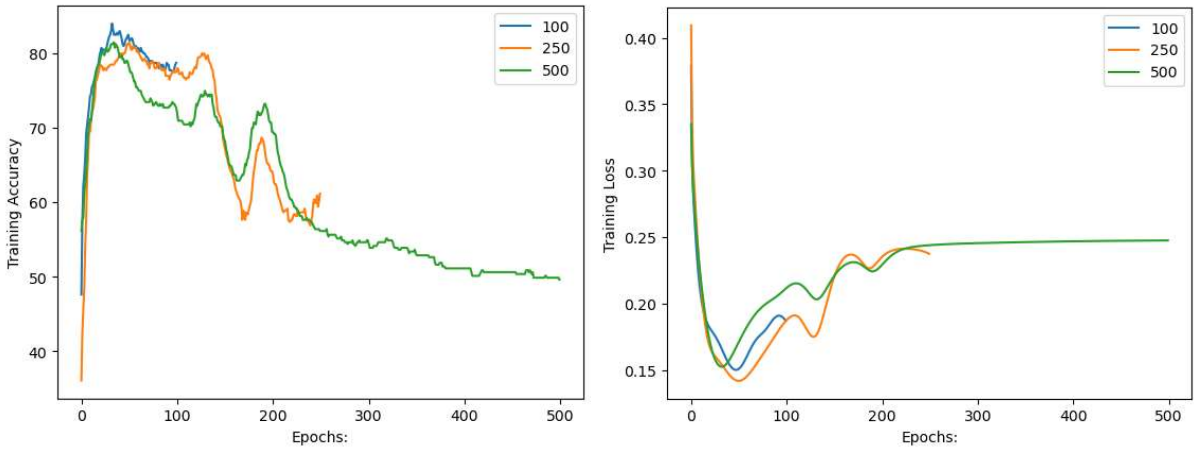

Figure 2. Accuracy and loss over training epochs for 100, 250, and 500 epochs

These observations are further reinforced by the testing accuracy results stored in Table 2, which shows an accuracy of over 75% for a network trained over 100 epochs and an accuracy of only 51.88% for a network trained over 500 epochs.

Table 2. Testing accuracy for 100, 250, and 500 epochs

# Epochs	Testing Accuracy
100	75.94%
250	58.65%
500	51.88%

3. Learning Rate

Learning rate results are somewhat skewed as an overflow error was often encountered in the sigmoid activation function at rates greater than 0.001. As a result, special attention will be given to the earlier training period of each chart displayed in Figure 3. These results indicate that the middle value of 0.01 may have been on the best trajectory towards optimal results before the error occurred. This may be due to the size of the network showing preference to the 0.01 step size at earlier time intervals. However, overall the smaller step size of 0.001 showed a steady increase in accuracy and decrease in loss over the course of training. While this smaller rate resulted in a longer training process, it also yielded the most steady training performance.

Nathan Butler
ROB 537 | HW1 Report
10/10/2023

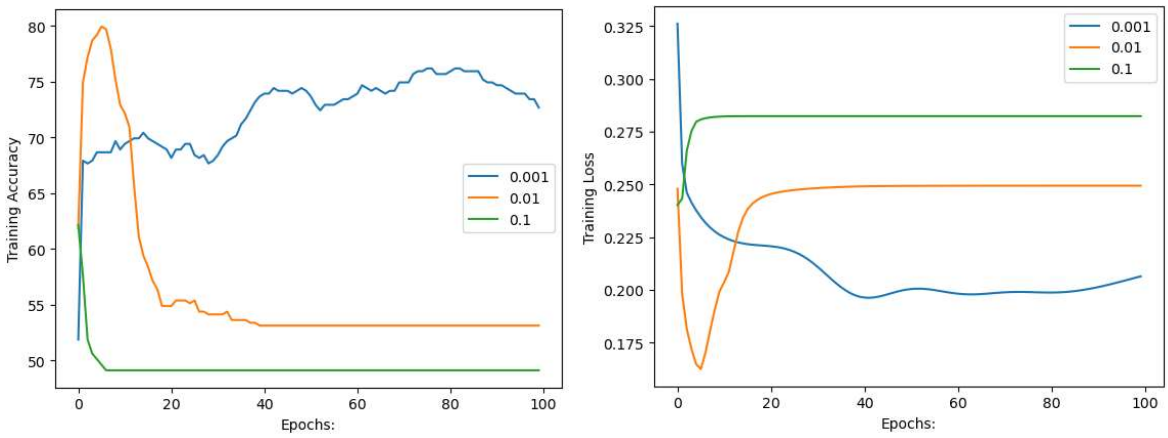

Figure 3. Accuracy and loss over training epochs for step sizes of 0.001, 0.01, and 0.1

Unfortunately, the failed learning from the overflow error for the 0.01 and 0.1 step sizes limits the accuracy results. Still, a 70% accuracy for 0.001 shows that the model was not overfit during training.

Table 3. Testing accuracy for 0.001, 0.01, and 0.1 step sizes

Step Size	Testing Accuracy
0.001	70.43%
0.01	50.13%
0.1	50.13%

4. Other Parameters

In training this network, weights and biases were initialized randomly. The initial accuracies and losses shown in the figures above illustrate the different effects that this initialization can have on the following training steps. While random sampling may be beneficial in certain settings, other approaches to initialization (such as sampling from a distribution) may result in more predictable training performances.

Additionally, the sigmoid activation function was used for the hidden and output layers of this network. There are many other types of activation functions, so future work may entail testing with ReLU, softmax, or other functions to determine which seems to have the best impact on training.

Training Set 2 Discussion

Unless otherwise noted, all training for part one was conducted using a 25 hidden units, a step size of 0.001, and 100 epochs.

1. Hidden Units

Figure 4a shows training accuracy and loss over training set 2 according to varying hidden unit sizes. Unlike set 1, there is not as noticeable of a difference in training performance according to hidden layer size. Dataset 2 is reportedly more complex, so perhaps increased training time may have shown greater separation between unit sizes and better performance overall.

Nathan Butler
ROB 537 | HW1 Report
10/10/2023

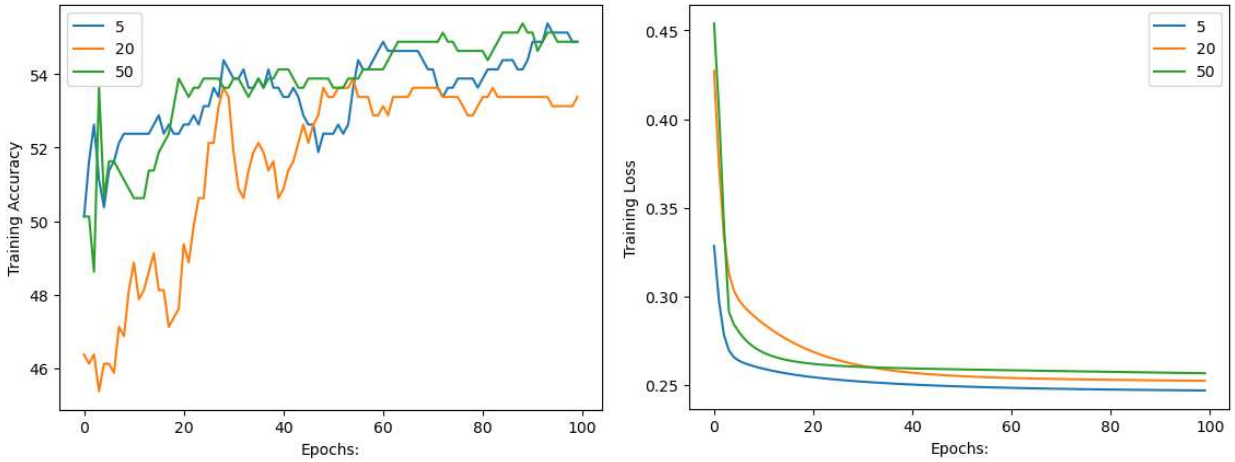

Figure 4a. Accuracy and loss over training epochs for 5, 20, and 50 hidden layer units

Because of these results, an additional test was conducted with larger hidden layers over 1000 training epochs. This appeared to have captured the complexity of training set 2 more effectively, and the results align better with those expected. Figure 4b and Table 4b show training and testing data.

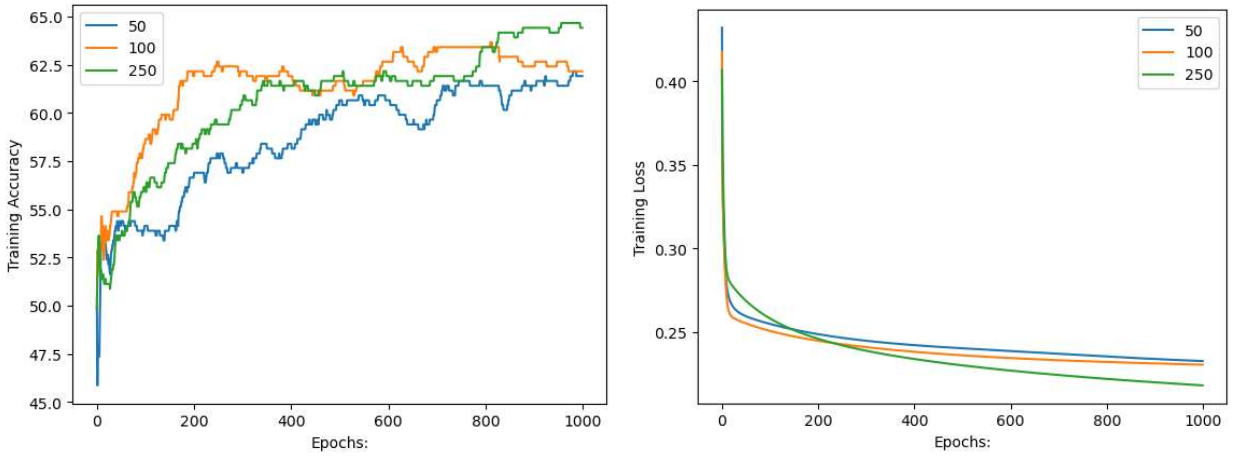

Figure 4b. Accuracy and loss over training epochs for 50, 100, and 250 hidden layer units over 1000 epochs

However, these training modifications may have resulted in some overfitting as the testing data performed poorly compared to training.

Table 4b. Testing accuracy for 50, 100, and 250 hidden layer units and 1000 epochs

# Hidden Units	Testing Accuracy
5	46.36%
20	47.87%
50	52.63%

Nathan Butler
ROB 537 | HW1 Report
10/10/2023

2. Training Time

Figure 5 shows that additional training time allowed the relatively basic networks to achieve better training performance, but like hidden units the simple network appears to not be capable of capturing the relationships present in the more complex data. Table 5 supports this conclusion with accuracy results around 50%. In order to effectively compare results between training sets 1 and 2, the remaining training on set 2 was conducted using the same simple network design implemented for set 1.

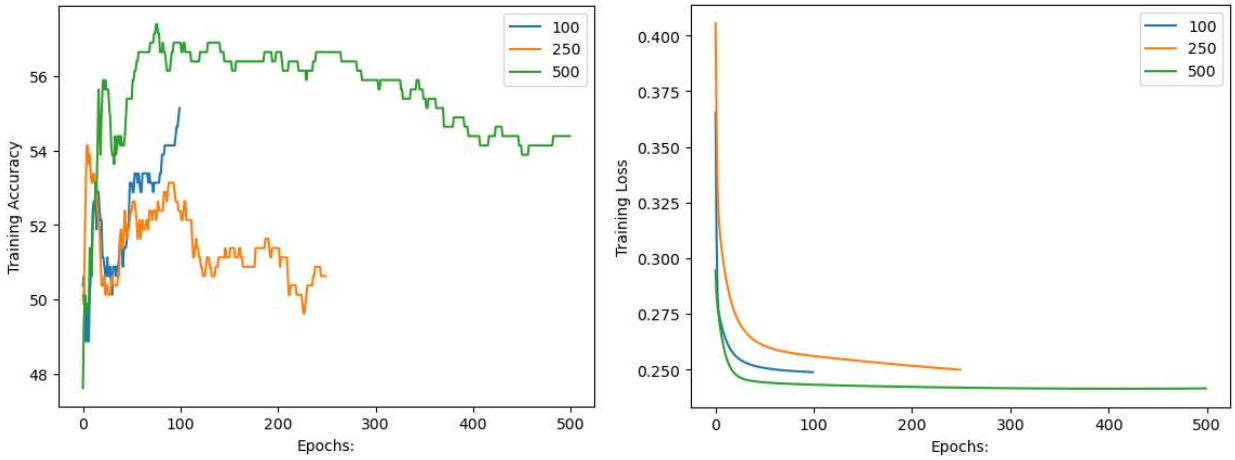

Figure 5. Accuracy and loss over training epochs for 100, 250, and 500 epochs

Table 5. Testing accuracy for 100, 250, and 500 epochs

# Epochs	Testing Accuracy
100	52.88%
250	54.88%
500	52.63%

3. Learning Rate

Learning rate results indicate that 0.01 may be a more optimal training rate for this network to capture relationships in the complex data in set 2. It might be the case the a rate of 0.1 resulted in changes that were too extreme while 0.001 was unable to produce results that could move out of local minima. Figure 6 shows the improved performance of the 0.01 rate, with similar testing performance reported in Table 6. It may be beneficial to train model at 0.01 for longer epochs, however the overflow error provides a constraint over more complex training configurations.

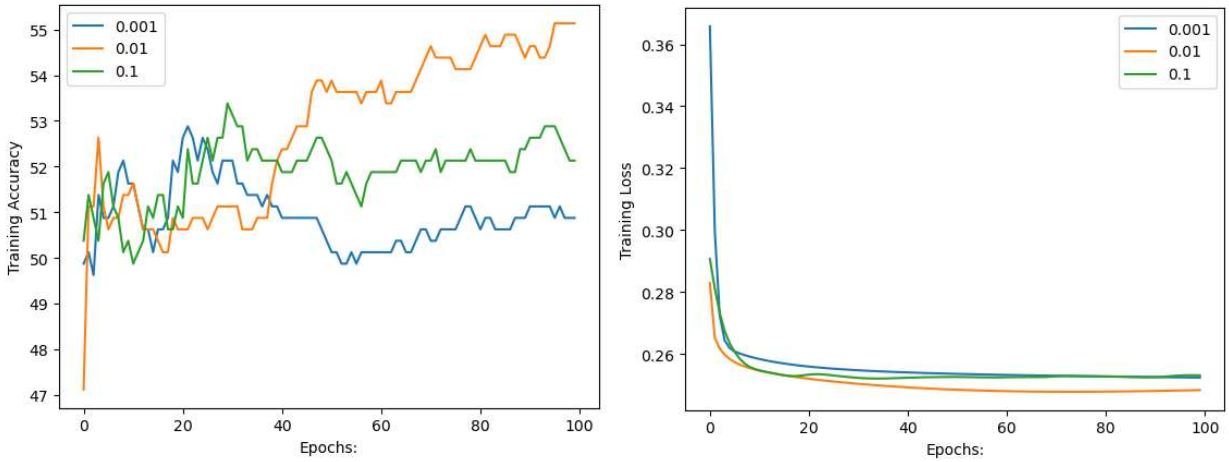

Figure 6. Accuracy and loss over training epochs for step sizes of 0.001, 0.01, and 0.1

Table 6. Testing accuracy for 0.001, 0.01, and 0.1 step sizes

Step Size	Testing Accuracy
0.001	52.13%
0.01	54.88%
0.1	48.87%

4. Other Parameters

In the case of this training, use of the sigmoid activation function provided a major bottleneck in computational abilities due to the overflow error observed in more complex training sizes. Other less computationally expensive activation functions like ReLU may be useful for expanding training capabilities.

Conclusions

Overall, I believe that the model parameters that yielded successful training results over training set 1 is too simple to capture the relationships present in the more complex set 2. 25 hidden layers cannot learn the more nuanced relationships, 100 epochs is too short of a training time, and the step size 0.001 may be too small for the training stage to break free of local minima. Going forward, I would be interest in observing the network's performance in training set 2 with over 100 hidden layers, at least 1000 epochs, and a step size around 0.01.