schuler_5779_ece_527_report_07

November 14, 2024

1 GMU ECE 527 - Computer Exercise #07 - Report

Stewart Schuler - G01395779
20241114

1.1 The XOR DataSet 6.2

Questions:

1) What is the default activation function for the hidden layers?
ReLU
2) What is the activation function that is used in the output layer?
Logistic Sigmoid

The first experiment on the *XOR* dataset is to consider number of iterations required for the gradient descent training to converge, and the training accuracy at that convergence point. The first tested classifier consisted of of hidden layer with 5 neurons. This classifier was trained 100 different times, and was found to have an average converge of 129 iterations and an average training accuracy of 0.98. The results of each of these 100 classifiers are shown in *figures 1* and *2* respectively. It can be seen from *figure 2* that many of the classifiers trained acheived a 1.0 training accuracy, conversely there also existed example where SGD lead to convergence at a local minimum that didn't have a 1.0 accuracy.
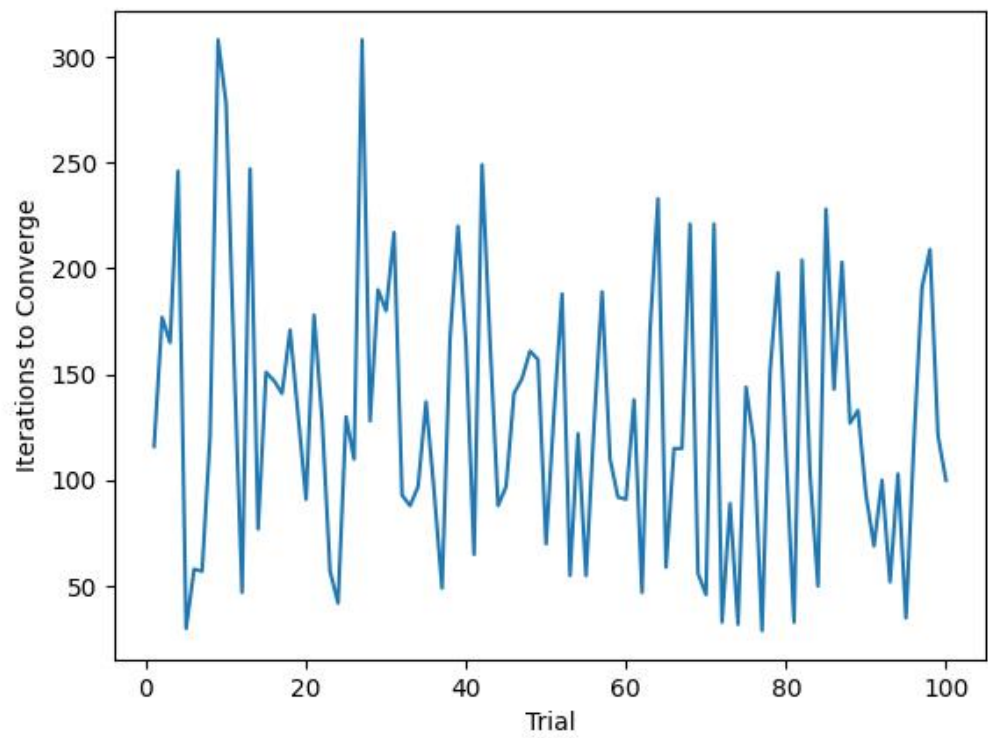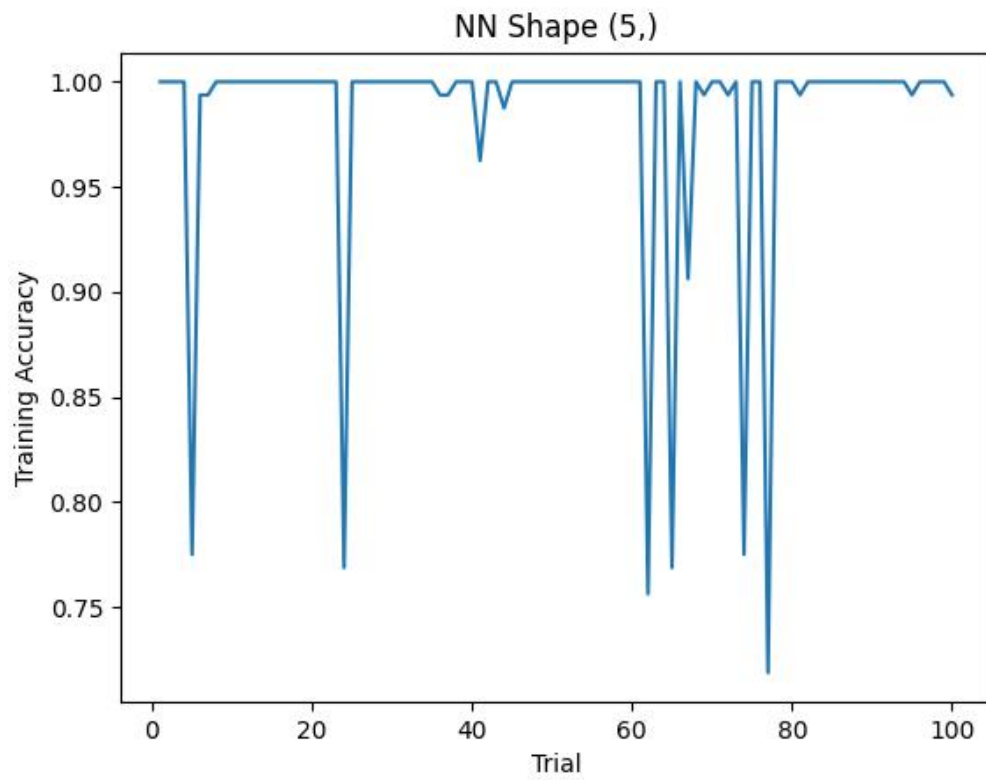
Figure 1. XOR Dataset 5 Neuron Convergence

Figure 2. XOR Dataset 5 Neuron Accuracy

The classification boundary of one such classifier with a 1.0 training accuracy is shown in *figure 3*.
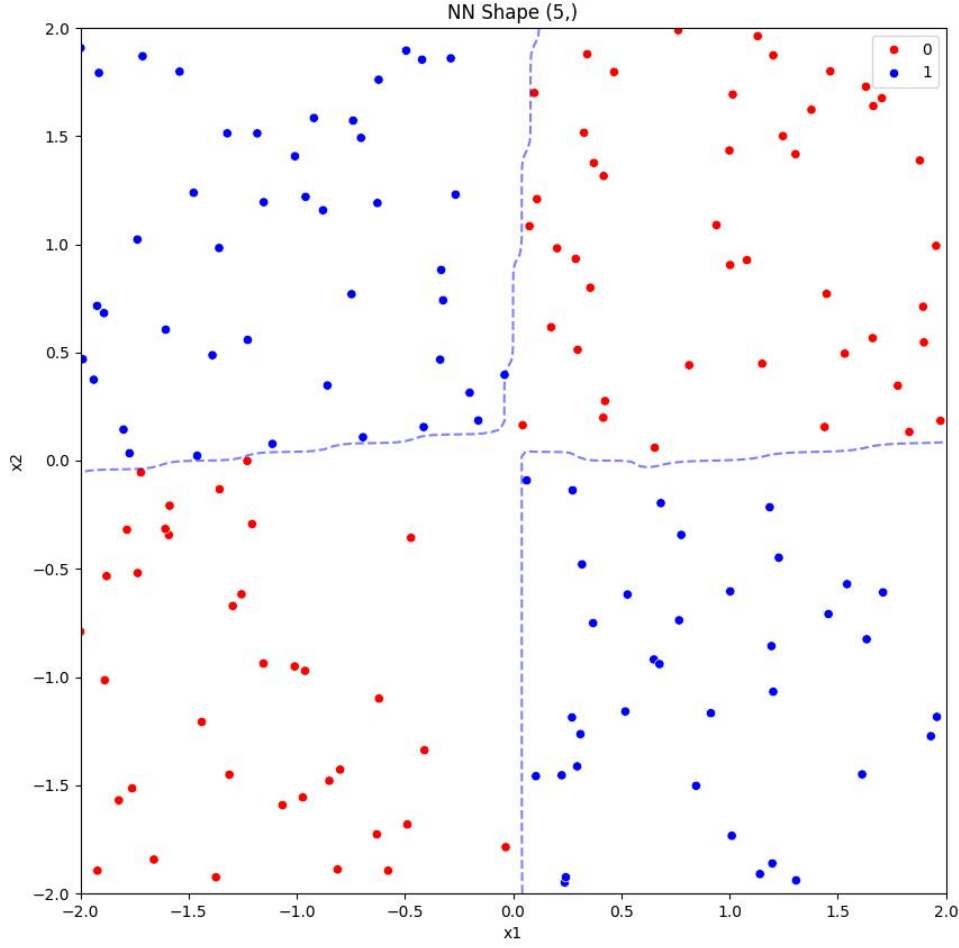
Figure 3. XOR Dataset 5 Neuron Classifier

The same experiments were then repeated for single hidden layer classifiers with 4, 3, and 2 neurons. The results of which are summarized in *table 1*.

Neurons	Mean Convergence	Mean Accuracy	Max Accuracy
5	129	0.98	1.0
4	117	0.95	1.0
3	100	0.90	1.0
2	36	0.77	0.88

Table 1 XOR Dataset Convergence and Accuracy for Number of Neurons

Unsuprisingly from *table 1* there is a clear trend that as the model complexity (number of neurons) decreases the convergence time decreases and the model accuracy goes down. The 2 neuron model is unable to represent the *XOR* dataset since the model complexity is to low to model a boundary

such as the one created by the *XOR* dataset. The highest training accuracy 2 neuron model is shown in *figure 4*, and it can be seen that the boundary is two straight lines. Any orientation of those lines will be unable to fully model the dataset. Thus we need a more complex model if we want to classify this boundary.
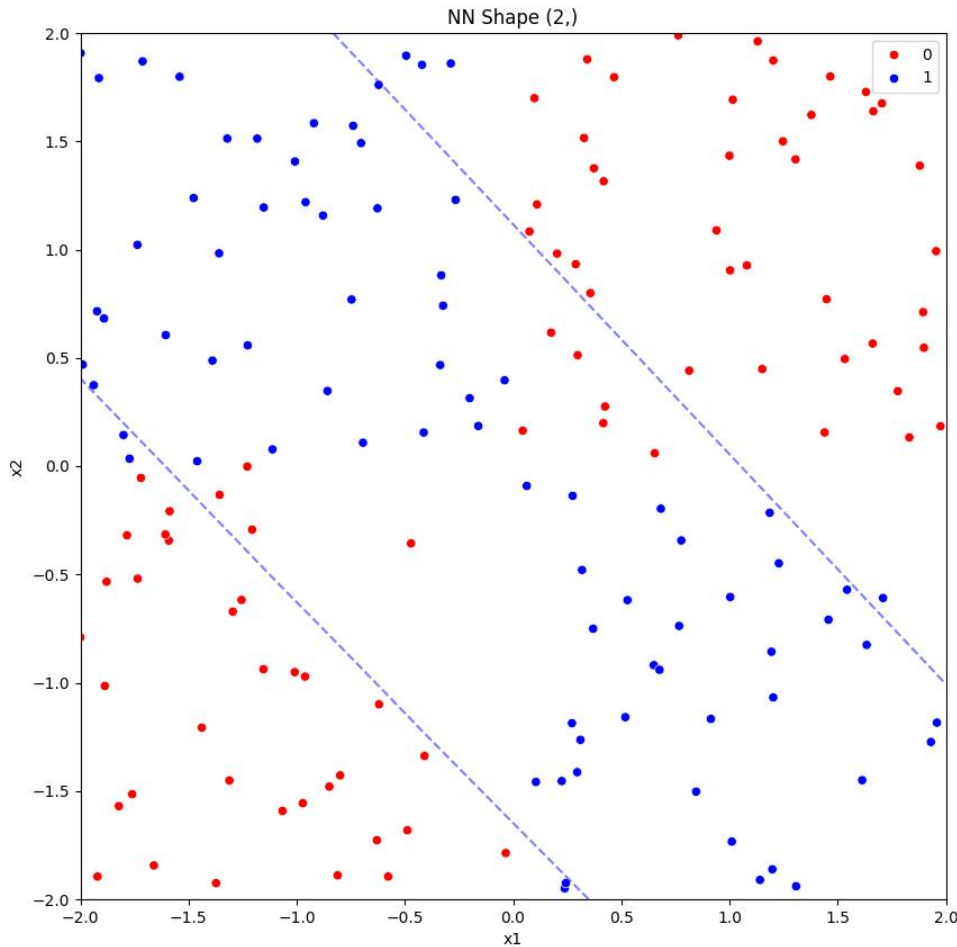

Figure 4. XOR Dataset 2 Neuron Classifier

Next we consider a deeper model. It has already been demonstrated that a single hidden layer with 3 or more neurons is sufficient to model the dataset. For this experiment a network with 3 hidden layers each consisting of 5 neurons was used. The results found an average convergence of 133 iterations, more or less the same as the single layer 5 neuron model. And a mean accuracy of 0.95 with a maximum of 1.0. With a lower mean accuracy, and an (very) slightly slower average convergence when comparted to the 1 layer 5 neuron model it can be said that increasing model complexity beyond what is required for the dataset can hurt performance.
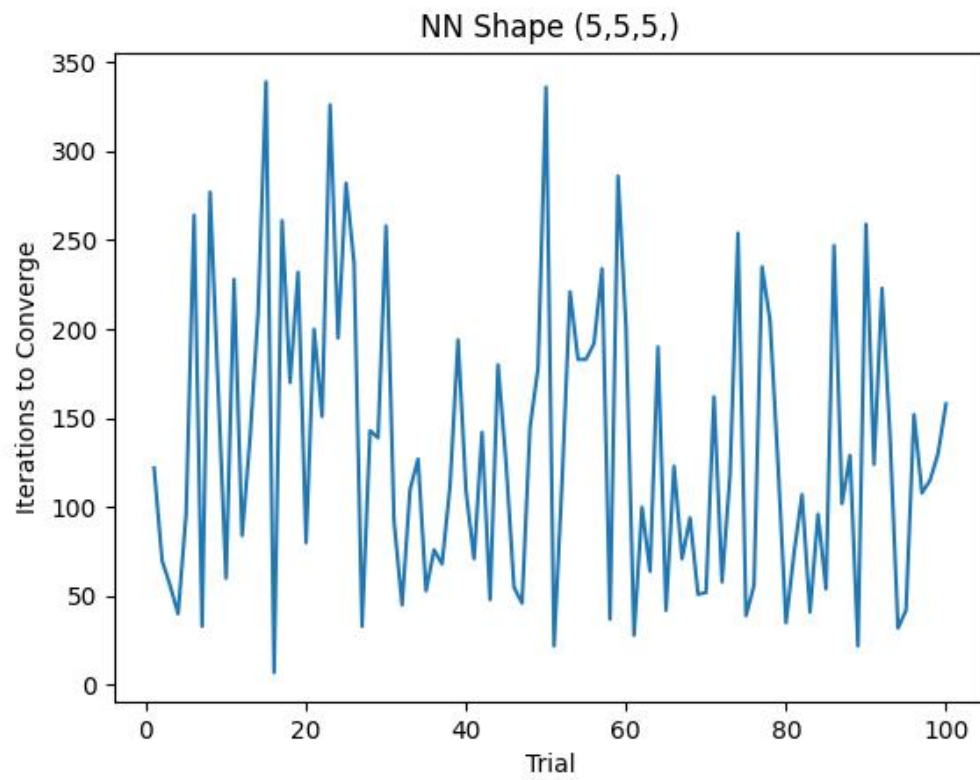
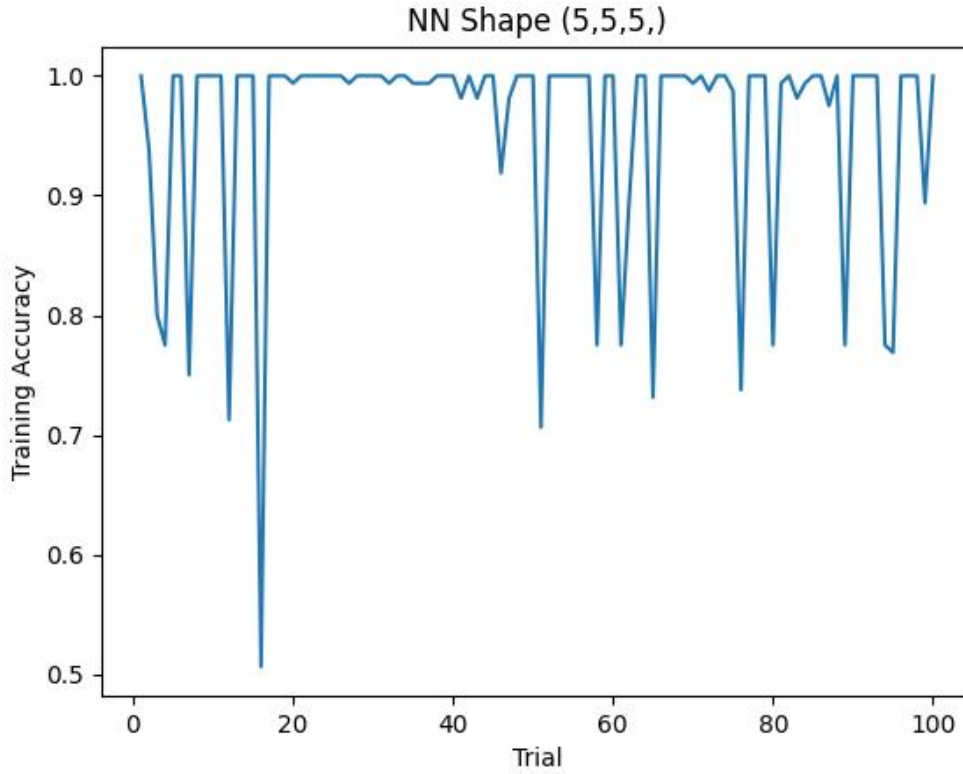**Figure 6.** XOR Dataset 5,5,5 Neuron Convergence

Figure 7. XOR Dataset 5,5,5 Neuron Accuracy

Lastly with *XOR* dataset we consider the impact of different neuron activation functions. *scikit-learn* supports three non-linear activations, *ReLU*, *logistic*, and *tanh*. All previous models thus far have used the *ReLU* activation. The result of the 3 tested activation functions are shown in *figure 8* and *9*, and summarized by *table 2*. It can be seen that the *ReLU* activation converges much faster than the other two. All 3 activation function were able in most trials to find a model with a 1.0 training accuracy. From *figure 9* it can be seen that when the *tanh* models converge before reaching a 1.0 accuracy they do so with accuracies far higher than the other two. However it can also be seen that this improvement in accuracy comes at the cost of convergence iterations i.e. training time.
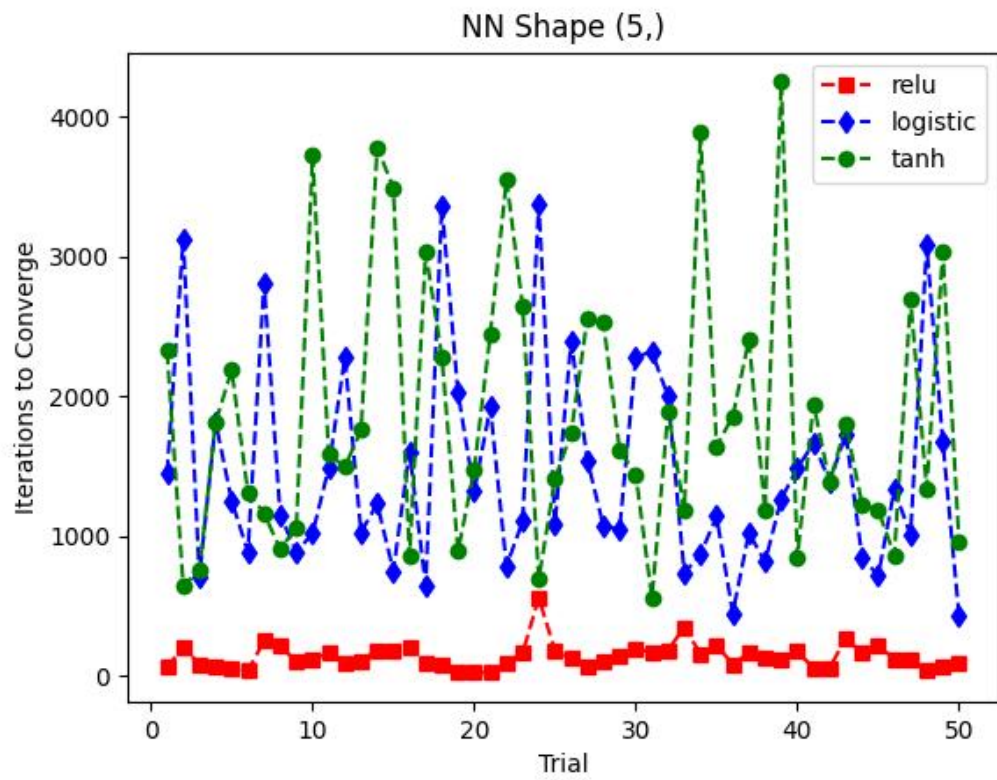
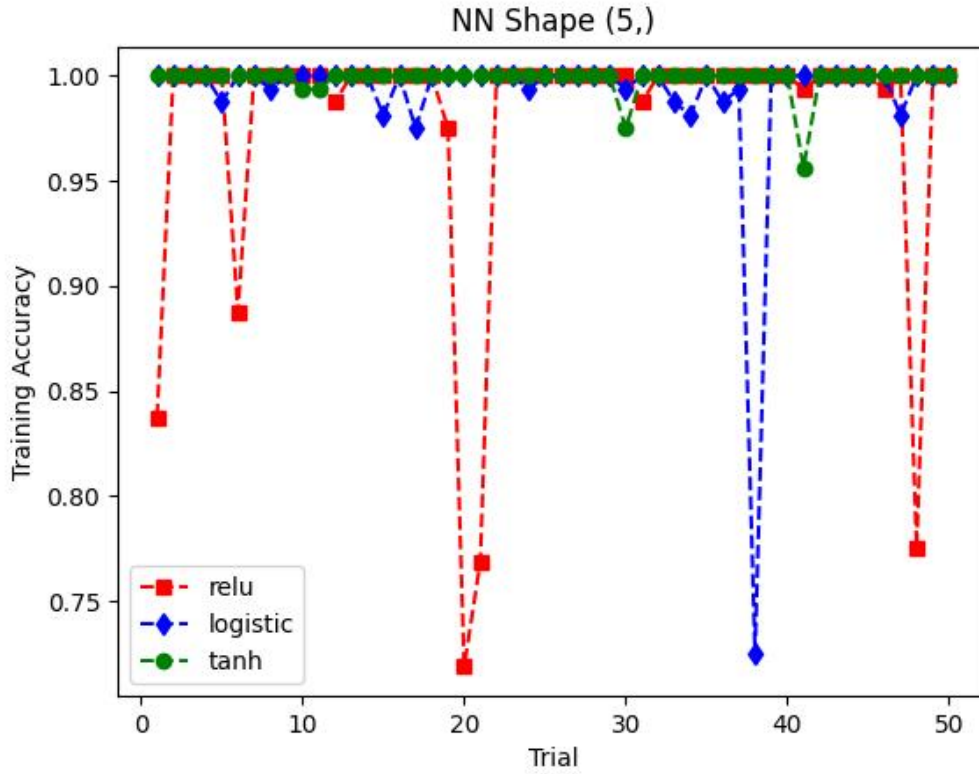**Figure 8.** XOR Dataset 5 Neuron Convergence - Different Activations

Figure 9. XOR Dataset 5 Neuron Accuracy - Different Activations

Activation	Mean Convergence	Mean Accuracy	Max Accuracy
ReLU	139	0.978	1.0
logistic	1469	0.991	1.0
tanh	1865	0.998	1.0

Table 2 XOR Dataset Convergence and Accuracy for Activation Functions

1.2 The Circles DataSet 6.3

Next we consider the circles dataset, first with a radius factor of 0.5. We repeat the same 100 trail classifiers tested on the *XOR* data and find the convergence iterations and accuracies summariazed in *table 3*.

Neurons	Mean Convergence	Mean Accuracy	Max Accuracy
5	111	0.93	1.0
4	95	0.90	1.0
3	68	0.81	1.0
2	32	0.69	0.85

Table 3 Circles (0.5) Dataset Convergence and Accuracy for Number of Neurons

Like with the *XOR* dataset it can be seen that a model with 2 neurons is unable to accuracy model this dataset. The highest training accuracy models on this dataset with 5 and 2 neurons are shown in *figures 10* and *11* respecitvly. Again the complexity of a 2 neuron model is unable to match that of the dataset.
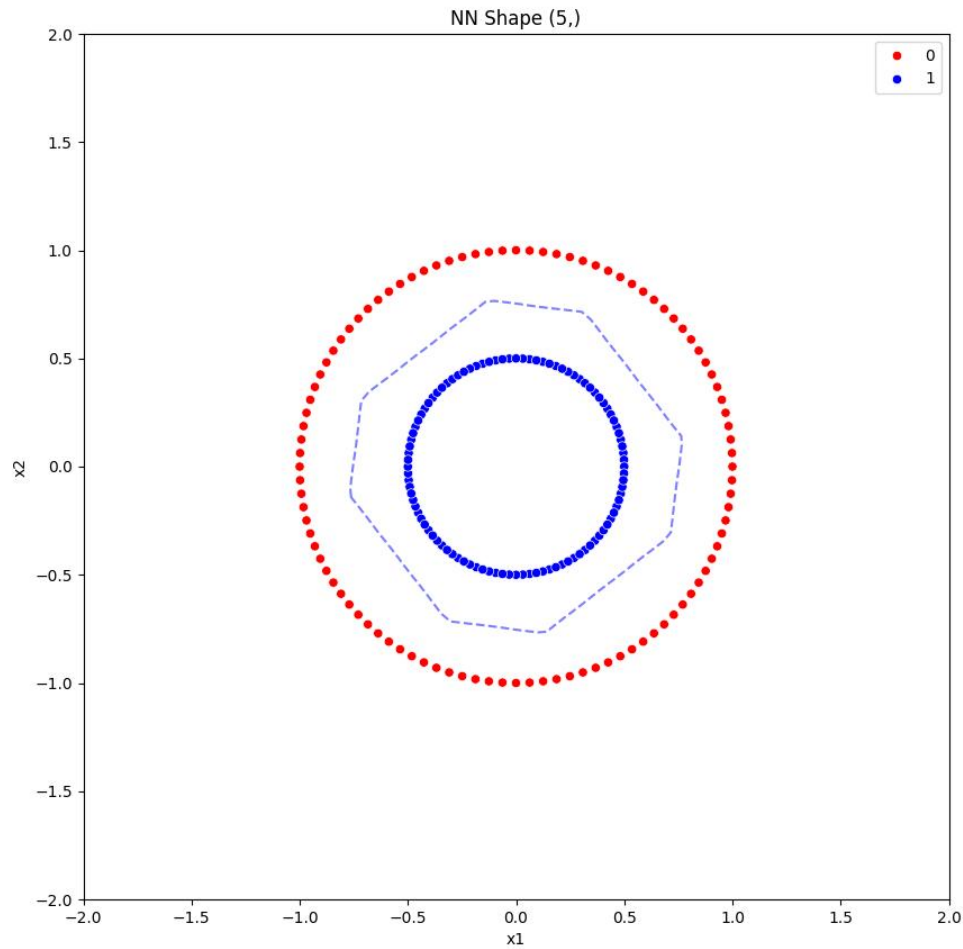

Figure 10. Circles (0.5) Dataset 5 Neuron Classifier
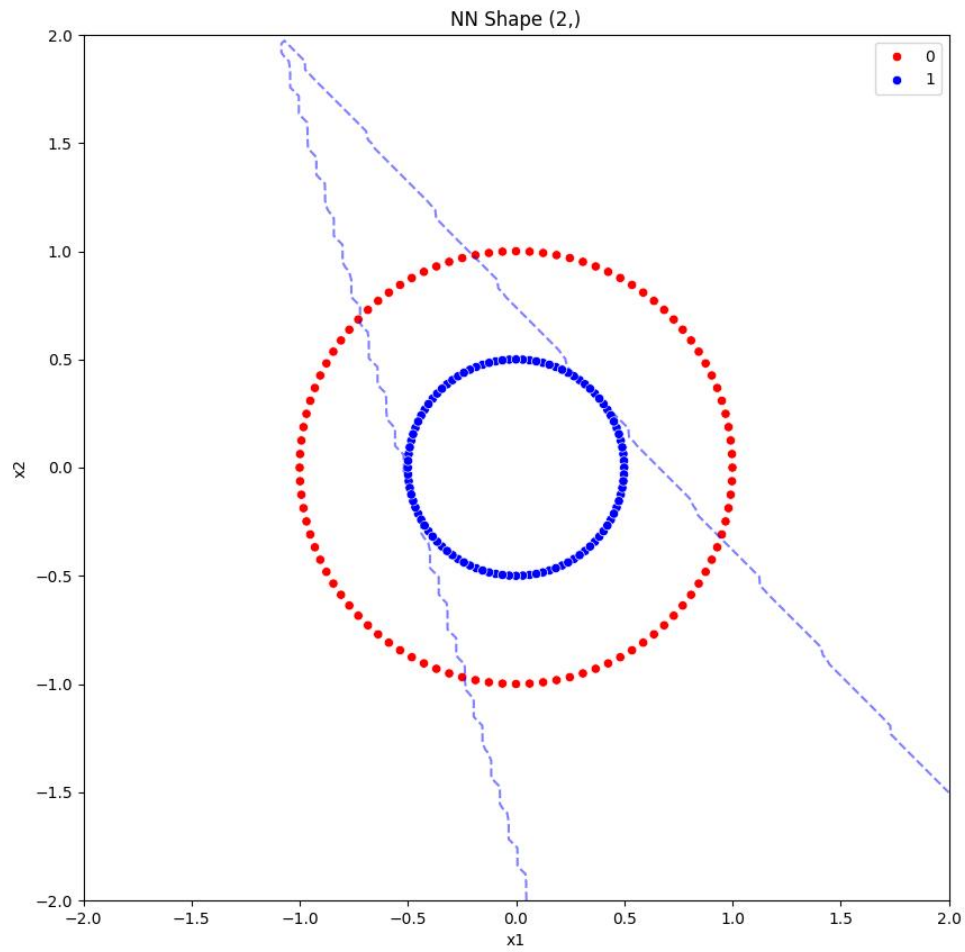
Figure 11. Circles (0.5) Dataset 2 Neuron Classifier

We then repeated the different activation functions test on the Circles 0.5 dataset. Like with the XOR datset the 3 different activation functions follow the same trend. However for this experiment the difference between convergence times is fast less extreme.
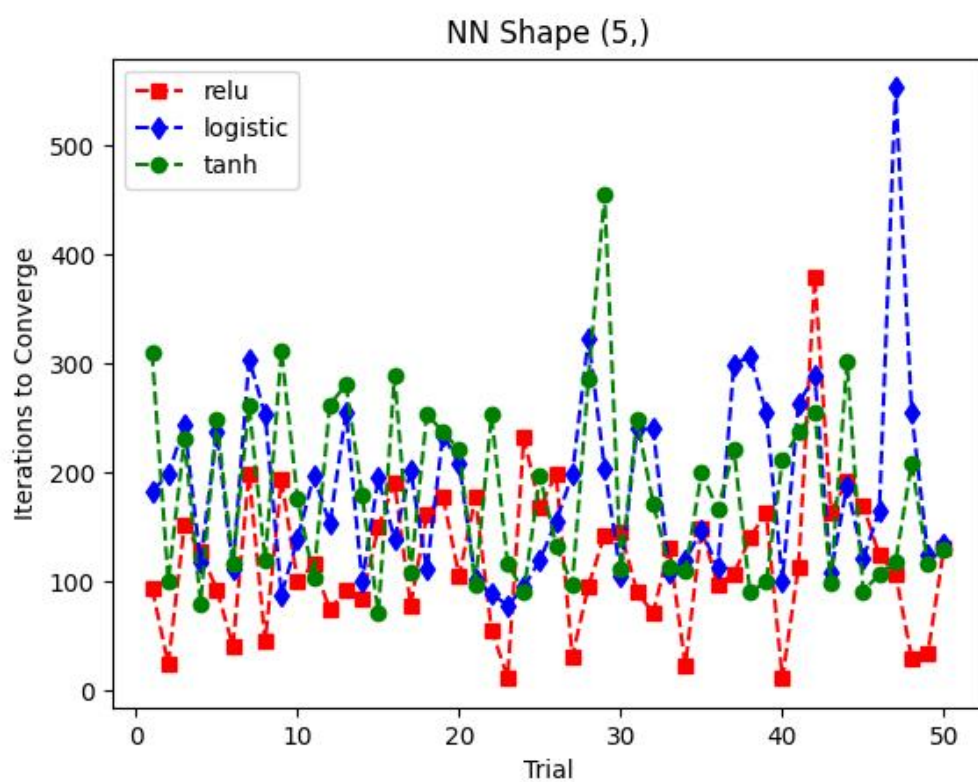
Figure 12. Circles (0.5) Dataset 5 Neuron Convergence - Different Activations
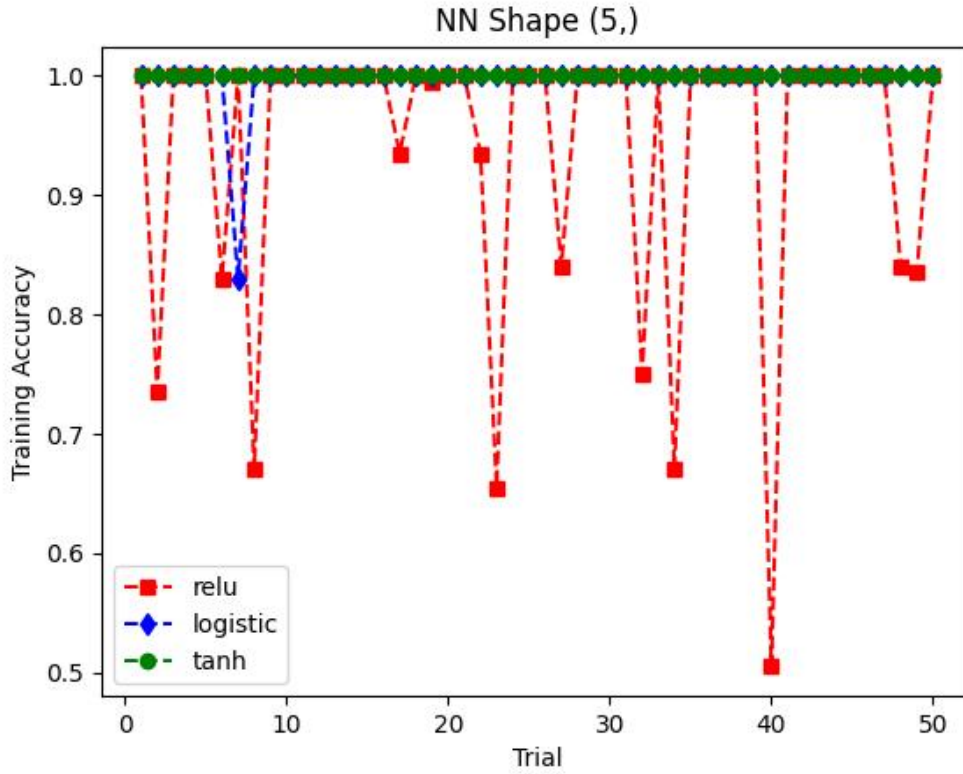
Figure 13. Circles (0.5) Dataset 5 Neuron Accuracy - Different Activations

Activation	Mean Convergence	Mean Accuracy	Max Accuracy
ReLU	111	0.959	1.0
logistic	205	0.979	1.0
tanh	182	0.993	1.0

Table 4 Circles (0.5) Dataset Convergence and Accuracy for Activation Functions

Next we modify the dataset to have a radius factor of 0.8 rerun the same experiement and produce the convergence iterations and accuracy number in *table 5*. An across the board increase in convergence iterations and a decrease in mean accuracies is observed. This is because the data set is "harder" to model since the buffer between classes where the boundary must be drawn is tighter.

Neurons	Mean Convergence	Mean Accuracy	Max Accuracy
5	204	0.87	1.0
4	132	0.80	1.0
3	63	0.69	1.0
2	35	0.61	0.71

Table 5 Circles (0.8) Dataset Convergence and Accuracy for Number of Neurons

A comparison between single layer classifier boundaries is shown in *figure 14* and *15*. In this case we looked at the 3 neuron which as the average accuracy indicates often converged before reaching a 1.0 accuracy. But in the case where it does such as *figure 15*, it produces a very close boundary to that of the 5 neuron model.
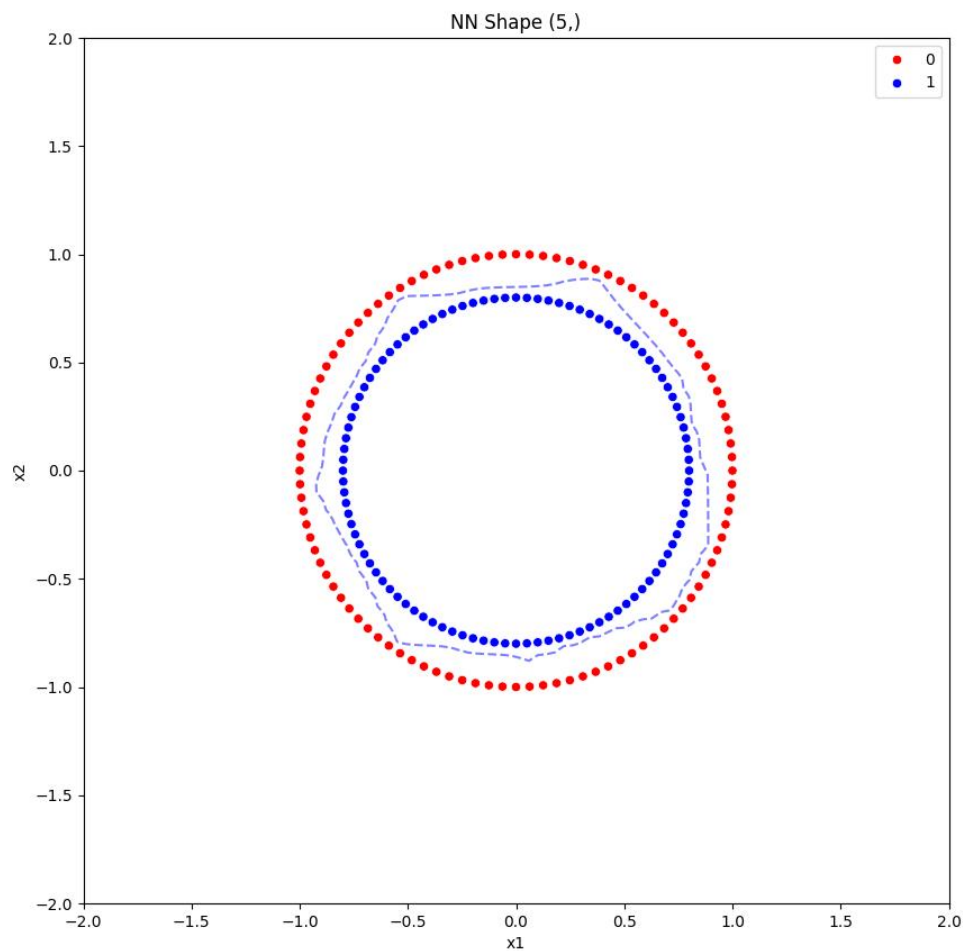

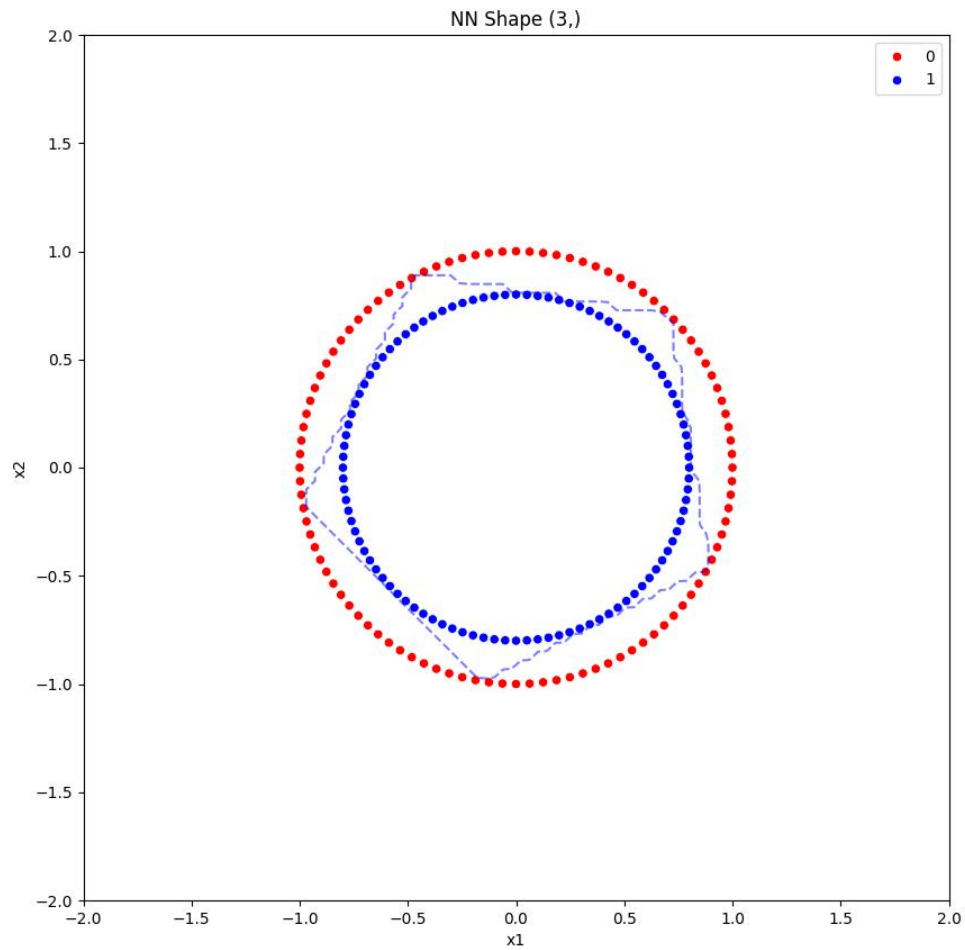**Figure 14.** Circles (0.8) Dataset 5 Neuron Classifier

Figure 15. Circles (0.8) Dataset 3 Neuron Classifier

Next we consider a deeper model with 3 layers of 5 neurons, over 100 trails it was found to have an average convergence of 189 and mean accuracy of 0.85. Which is very similar performance to the 1 layer 5 neuron model. However from the plotted decision boundary we see it is able to model a more complex (more sides) shape which does a better job modeling that halfway point between the two classes.
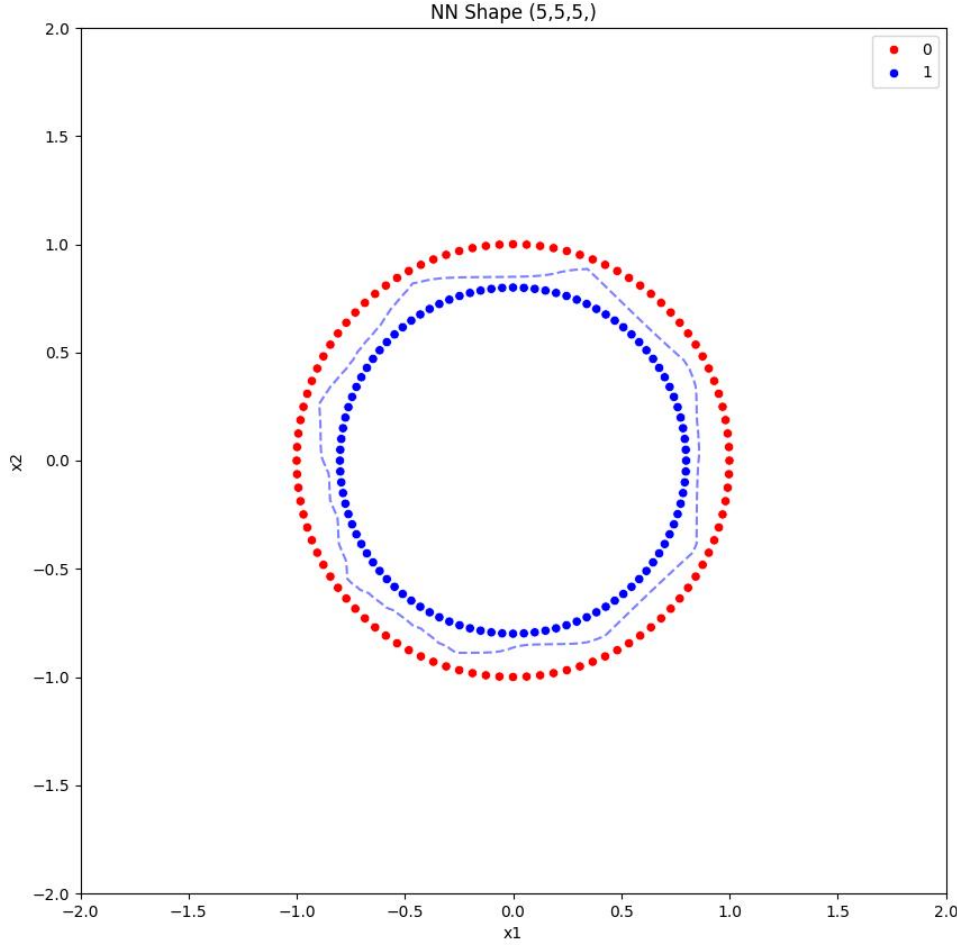
Figure 16. Circles (0.8) Dataset 5,5,5 Neuron Classifier

The experiment with different activation functions yielded the results in *table 6*, which interestingly align closer with the results observed on the *XOR* dataset than those from the *Cirlces 0.5* dataset.

Activation	Mean Convergence	Mean Accuracy	Max Accuracy
ReLU	158	0.869	1.0
logistic	747	0.971	1.0
tanh	1080	0.977	1.0

Table 6 Circles (0.8) Dataset Convergence and Accuracy for Activation Functions

Finally the last dataset tested it the *Circles* dataset with a radius factor of 0.9. The results of varying the number of neurons in the single hidden layer are shown in *table 7*. Again, unsuprisingly, as the dataset got more difficult average accuracies decreased. Noteable for this dataset is now that 3 neuron model wasn't able to converge to a 1.0 training accuracy for an of the 100 trails ran.

Neurons	Mean Convergence	Mean Accuracy	Max Accuracy
5	248	0.77	1.0
4	162	0.70	1.0
3	61	0.64	0.94
2	35	0.57	0.71

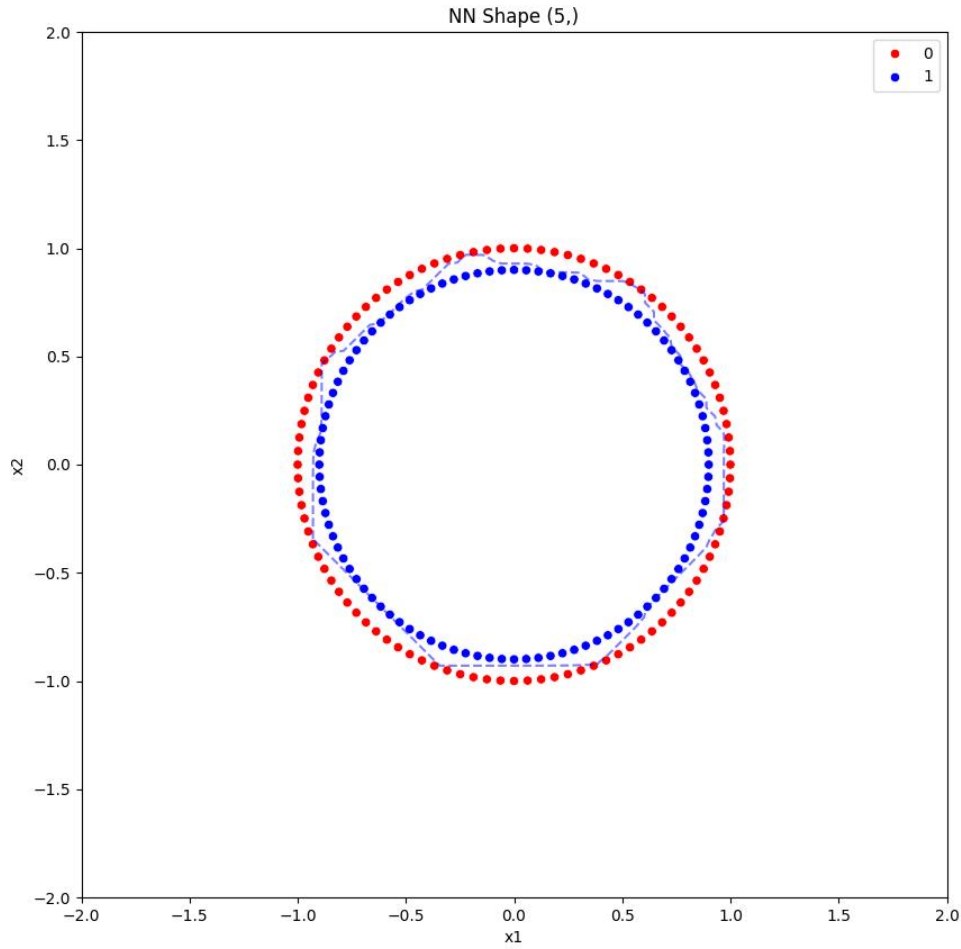**Table 7** Circles (0.9) Dataset Convergence and Accuracy for Number of Neurons

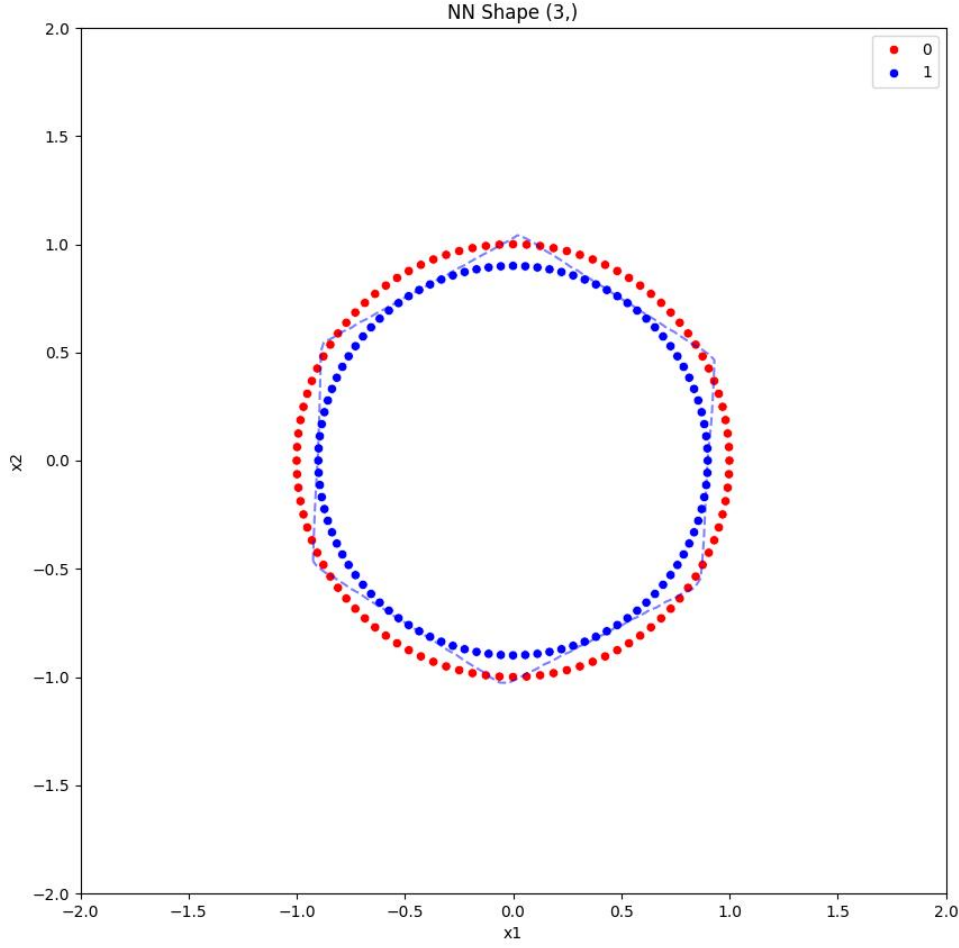**Figure 17.** Circles (0.9) Dataset 5 Neuron Classifier

Figure 18. Circles (0.9) Dataset 3 Neuron Classifier

The results of the 3 layer 5 neuron model demonstrated the same conclusions as with the *Circles 0.8* dataset in that it performed nearly identically to the 1 layer 5 neuron model on the same dataset.

When comparing activation function we again see the same trend that *ReLU* converges fastest with the worst average training accuracy, where as the *logistic* and *tanh* activation fucntions take substantially longer, and produce higher average accuracies. Both the *logistic* and *tanh* functions have performed similar to one another when tested on all datasets.

Activation	Mean Convergence	Mean Accuracy	Max Accuracy
ReLU	290	0.827	1.0
logistic	2158	0.953	1.0
tanh	1991	0.947	1.0

Table 8 Circles (0.9) Dataset Convergence and Accuracy for Activation Functions