

Problem 1)

```
XOR(0,0): 0  
XOR(1,0): 1  
XOR(0,1): 1  
XOR(1,1): 0
```

Problem 2)

Below are screenshots of ReLU/Sigmoid test and training accuracy across 16, 32, and 64 hidden nodes, and across 50, 100, and 200 epochs.

Hidden Layer Size: 16 Number of Epochs: 50 ReLU Model Training Accuracy: 24.58 % ReLU Model Testing Accuracy: 24.8 % Sigmoid Model Training Accuracy: 77.55 % Sigmoid Model Testing Accuracy: 76.96 %	Hidden Layer Size: 32 Number of Epochs: 50 ReLU Model Training Accuracy: 80.51 % ReLU Model Testing Accuracy: 80.35 % Sigmoid Model Training Accuracy: 86.14 % Sigmoid Model Testing Accuracy: 85.49 %	Hidden Layer Size: 64 Number of Epochs: 50 ReLU Model Training Accuracy: 86.03 % ReLU Model Testing Accuracy: 85.56 % Sigmoid Model Training Accuracy: 88.6 % Sigmoid Model Testing Accuracy: 87.9 %
Hidden Layer Size: 16 Number of Epochs: 100 ReLU Model Training Accuracy: 46.97 % ReLU Model Testing Accuracy: 47.15 % Sigmoid Model Training Accuracy: 85.34 % Sigmoid Model Testing Accuracy: 83.85 %	Hidden Layer Size: 32 Number of Epochs: 100 ReLU Model Training Accuracy: 86.57 % ReLU Model Testing Accuracy: 85.7 % Sigmoid Model Training Accuracy: 89.46 % Sigmoid Model Testing Accuracy: 87.78 %	Hidden Layer Size: 64 Number of Epochs: 100 ReLU Model Training Accuracy: 89.71 % ReLU Model Testing Accuracy: 88.75 % Sigmoid Model Training Accuracy: 91.92 % Sigmoid Model Testing Accuracy: 90.12 %
Hidden Layer Size: 16 Number of Epochs: 200 ReLU Model Training Accuracy: 39.16 % ReLU Model Testing Accuracy: 38.82 % Sigmoid Model Training Accuracy: 88.01 % Sigmoid Model Testing Accuracy: 85.88 %	Hidden Layer Size: 32 Number of Epochs: 200 ReLU Model Training Accuracy: 90.91 % ReLU Model Testing Accuracy: 89.61 % Sigmoid Model Training Accuracy: 91.22 % Sigmoid Model Testing Accuracy: 88.93 %	Hidden Layer Size: 64 Number of Epochs: 200 ReLU Model Training Accuracy: 93.33 % ReLU Model Testing Accuracy: 91.46 % Sigmoid Model Training Accuracy: 93.62 % Sigmoid Model Testing Accuracy: 91.1 %

The 64 hidden node network gives the best testing and training accuracy, although only slightly better than the 32 hidden node network. Across all cases, the sigmoid function provides better or comparable performance to the ReLU activation function. The difference is especially present in the 16 node network, where the ReLU model is very inaccurate. The tradeoff between the two at a larger hidden layer size seems to be the sigmoid is slightly more accurate, while the ReLU function is known to be computationally cheaper and faster.

With too little nodes, the network may not be able to learn the pattern. This is shown by the ReLU model's inaccuracy with layer size 16, which drastically improves with 32/64 hidden nodes. With too many nodes, the network may become too heavy for the task at hand. It will take longer and use more computer resources than a network with less nodes for equal performance.

The ANN performance is likely to remain the same. Increasing layers allows an ANN to learn more complex relationships and functions between the data and output. Given the strong testing accuracy found in the 64 hidden node network (>90% accuracy), it's reasonable to assume the mnist dataset

doesn't have any complex relationship the ANN is currently missing. As a result, I believe adding more layers will be redundant and not capture any new relationship, keeping similar accuracy to the single hidden layer network.