Task 1

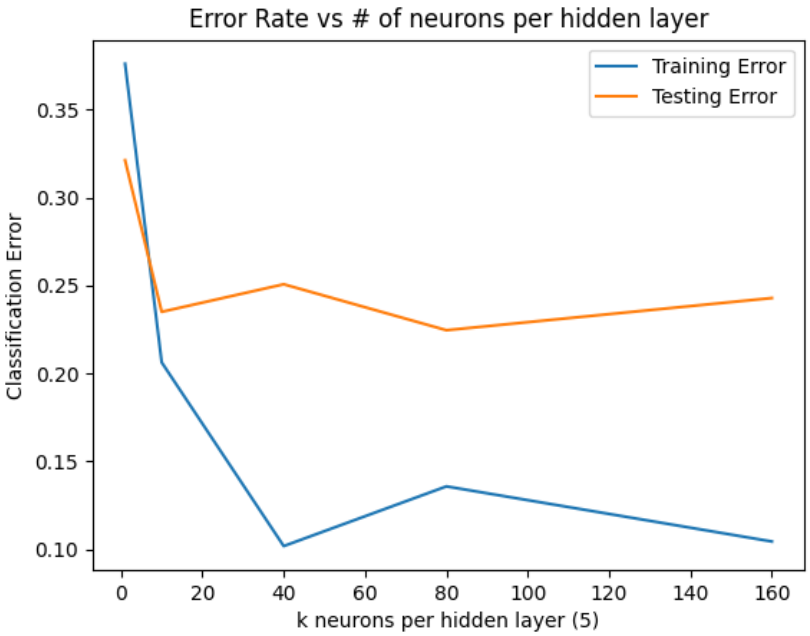

Figure 1: Error change with number of neurons per hidden layer (5 hidden layers)

At first, the model is underfitting with a high error rate in both training and testing. Then at around 80 neurons, the model looks to be the most correct fit, resulting in a lower testing error rate. Finally after the 80 neurons, and the 160 neurons the model error seems to be growing again while the training error is going down meaning the model is being overfit.

Task 2

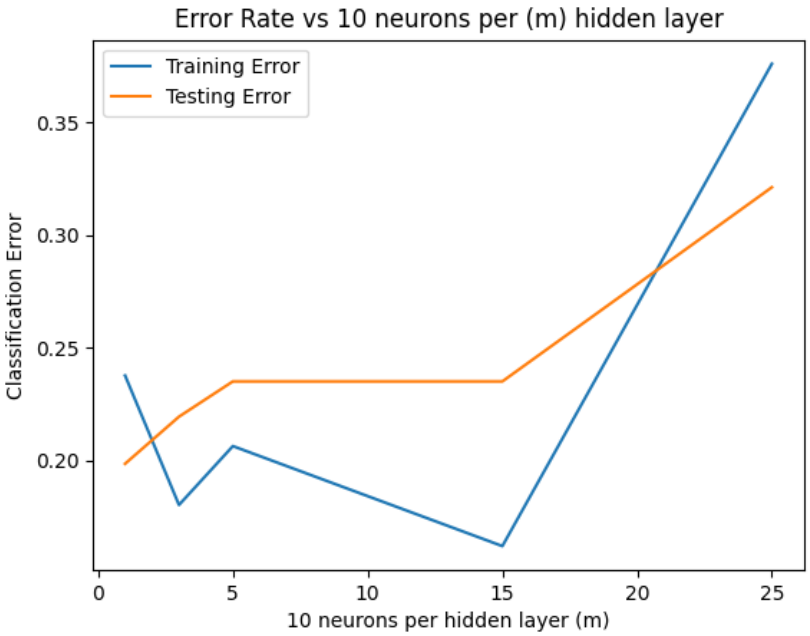

Figure 2: Error change with number of hidden layers (10 neurons per hidden layer)

At 1 hidden layer, the error of the test set seems to be at it's smallest. Then up to 15 hidden layers the training error goes down, but the testing error goes up which shows overfitting. Finally, at 25 hidden layers, the training error skyrockets and so does the testing error, which shows a potential underfitting that's worse than the base case.

Task 3

Activation Function	Training Error	Testing Error
Identity	0.230	0.201
Logistic	0.376	0.321
Tanh	0.235	0.196
reLU	0.136	0.225

Table 1: Error Values of Each Activation Function (80 neurons, 5 hidden layers)

Here we can see that the list of activation functions in terms of highest to lowest training error is logistic, tanh, identity, & relu. Then the list of activation functions in terms of highest to lowest test error is logistic, relu, identity, & tanh. Here we can just tell that the logistic model is the worst one. Then the identity function seems to match up the training and testing error the best while trying to lower the error. The relu tries to lower the training error really hard and results in a little bit of overfitting compared to the other functions. and tanh seems to get the best generalization resulting in the lowest test error.