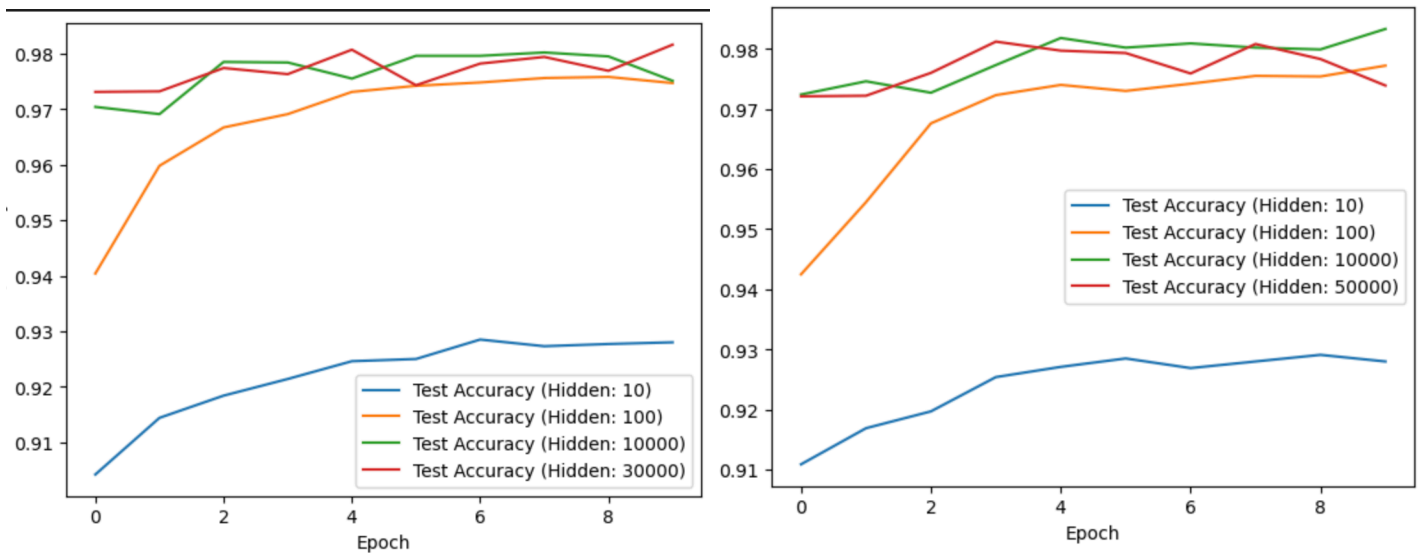


## 1.1 MLP with varying hidden layer



The Test Accuracy is represented by the red line in the above figures.

In order to determine how wide the network needs to be before it starts to overfit on the MNIST training dataset, I tested various number of hidden units within the 500,000 range. The left diagram uses 30000 as the maximum hidden units, the “Test Accuracy” is still increasing.

Next, when I increased the hidden units to 50,000 the “Test Accuracy” started decreasing by the end of the epoch, this indicates overfitting.

At this point the model is too complex and starts to memorize rather than learn the patterns.

So, although the training accuracy increases as the number of hidden layer increases, the test accuracy helps us see the point at which overfitting starts to occur, which is around 50,000 hidden units.