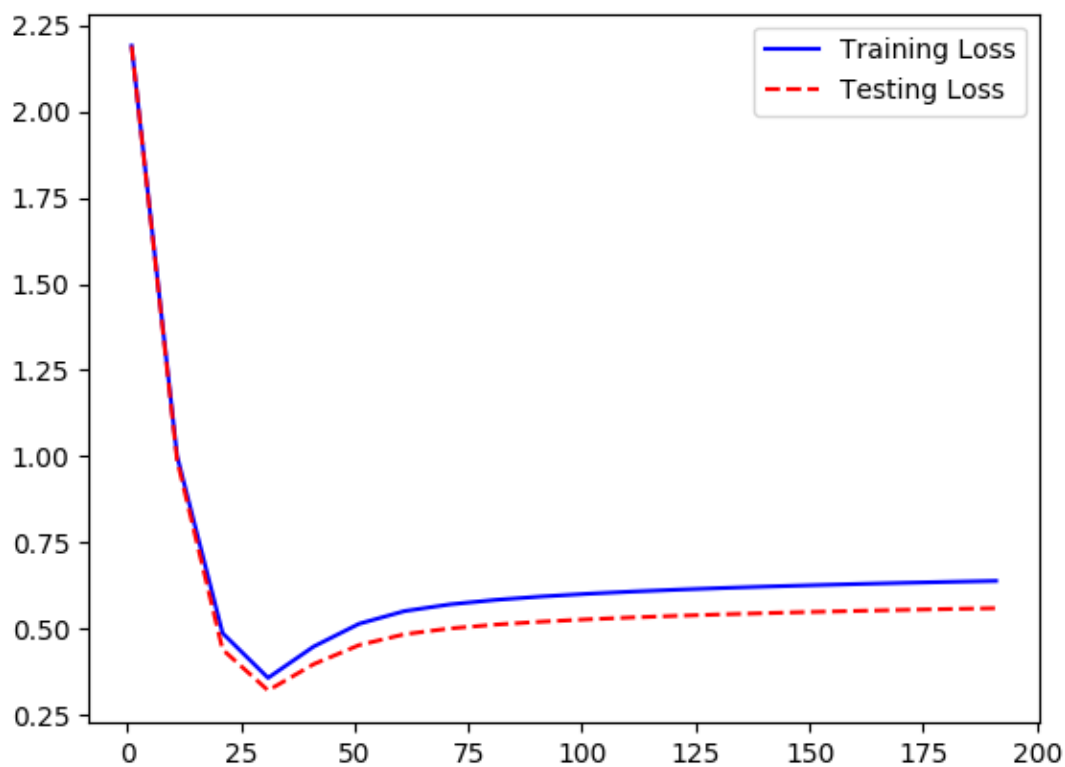
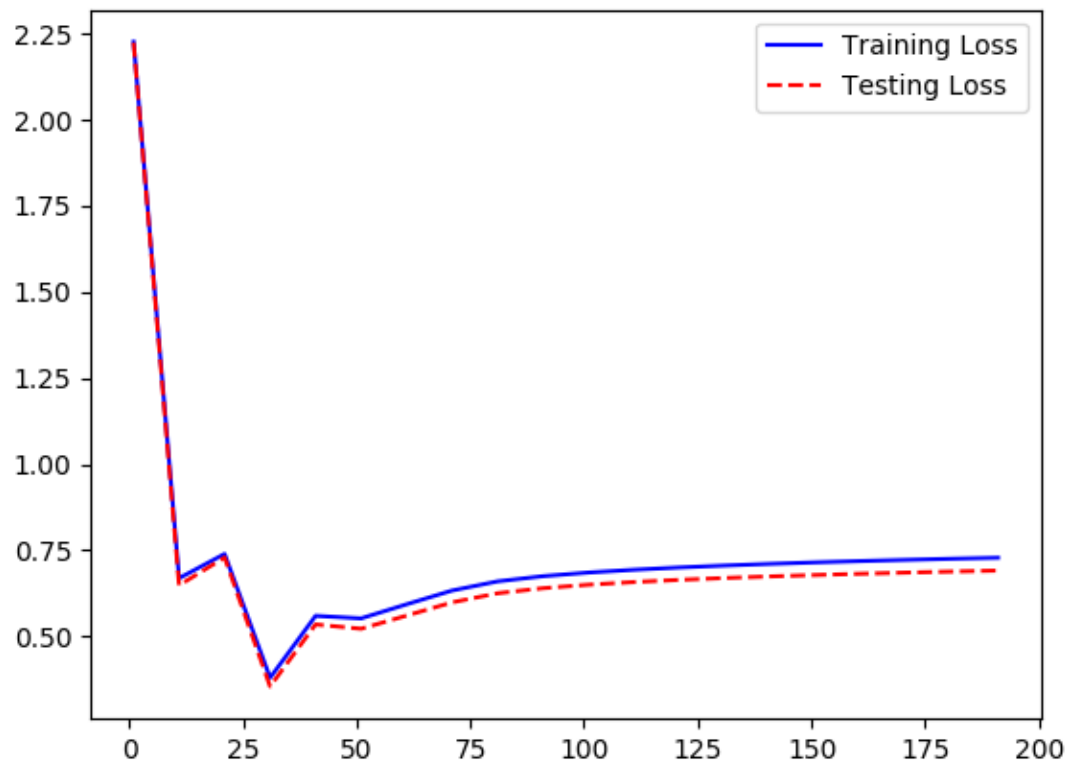


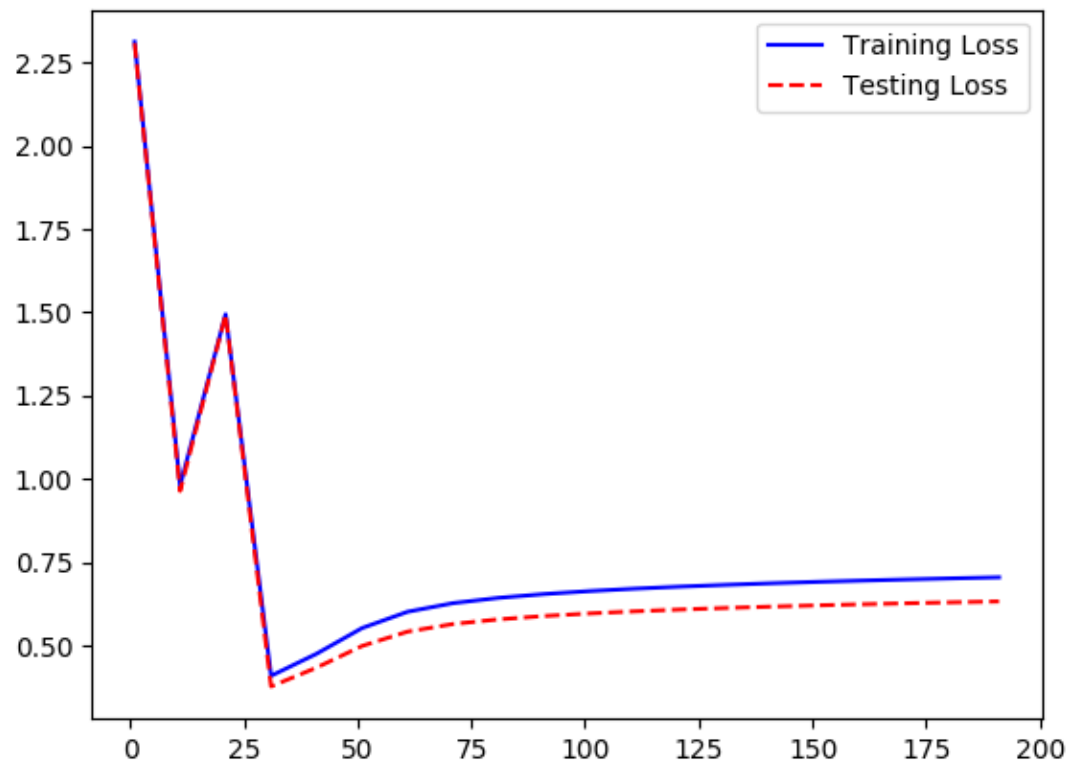
1. A) (2048,1024,512,128,32)



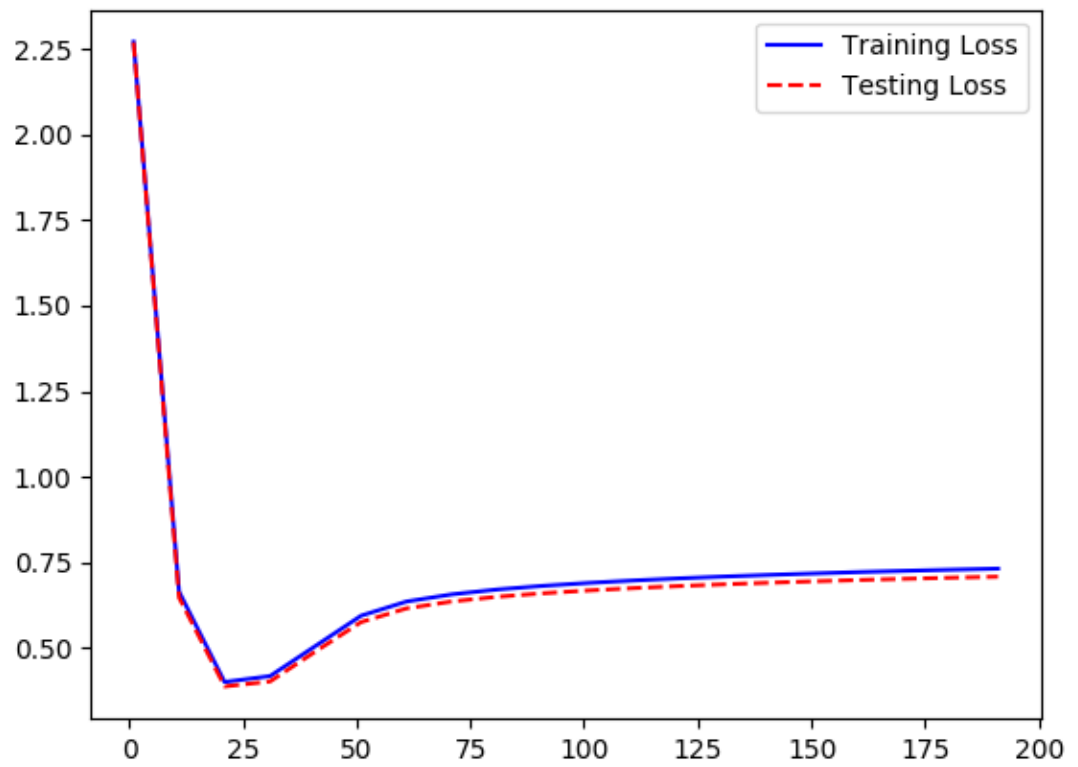
b) (1024,512,128,64,32)



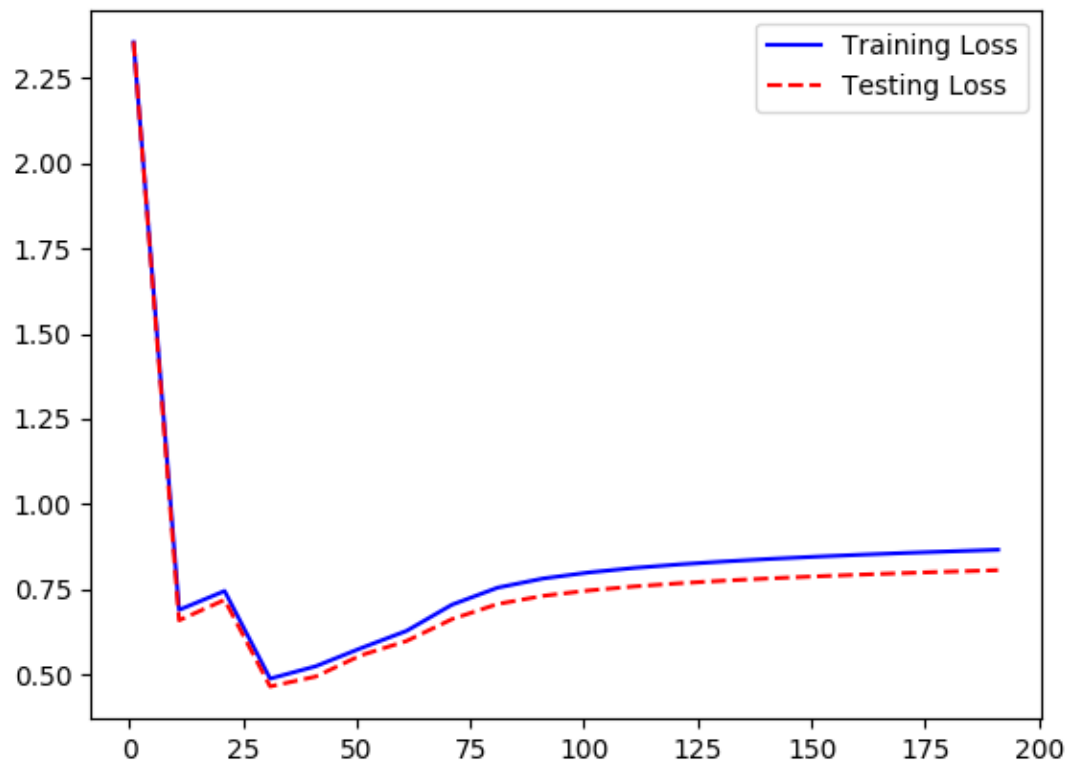
c) (512,256,128,32,16)



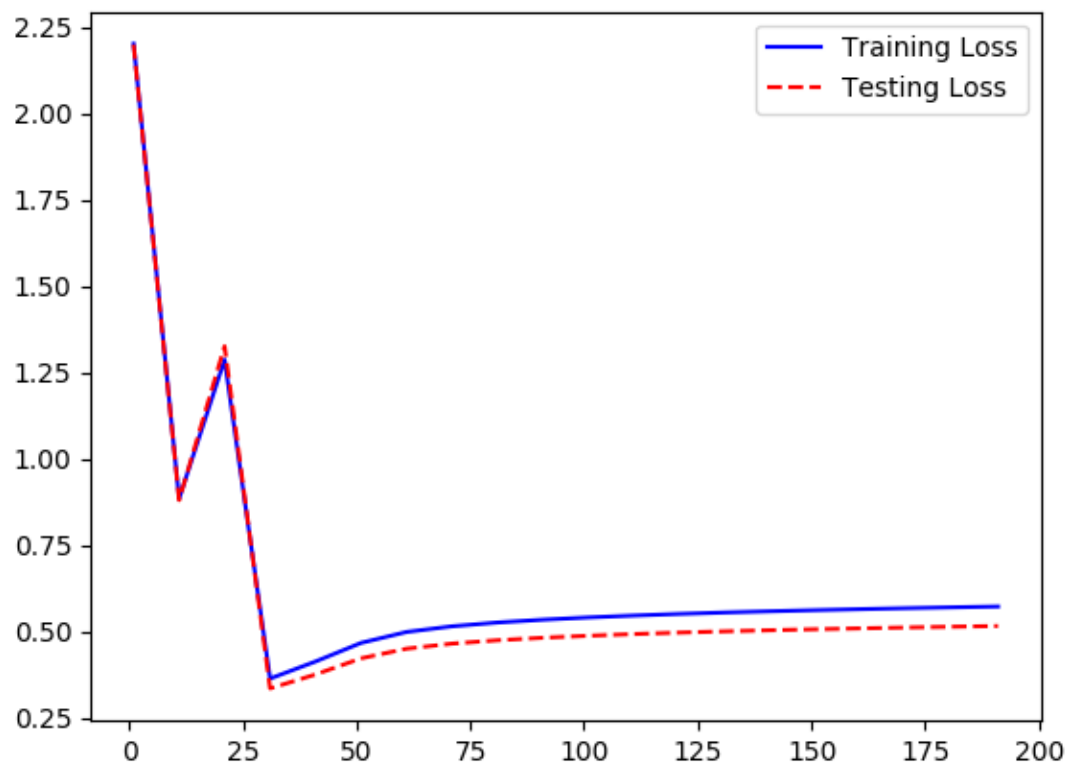
d) (256,128,32,16,8)



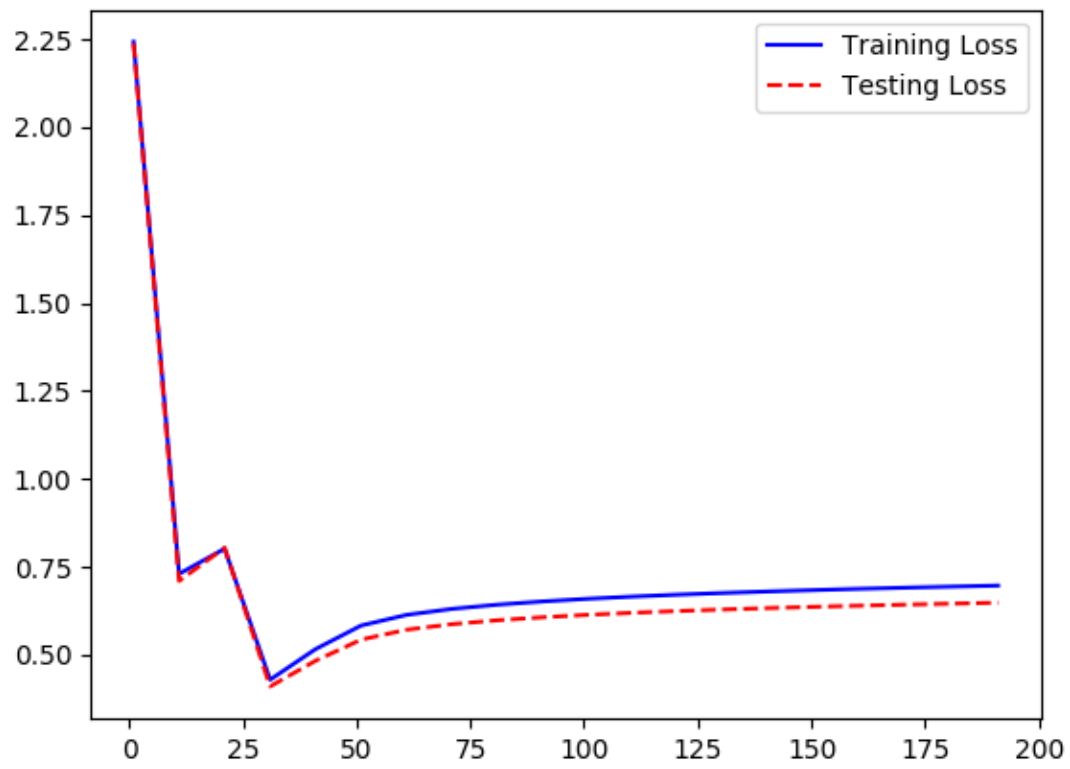
e) (128,64,32,16,8)



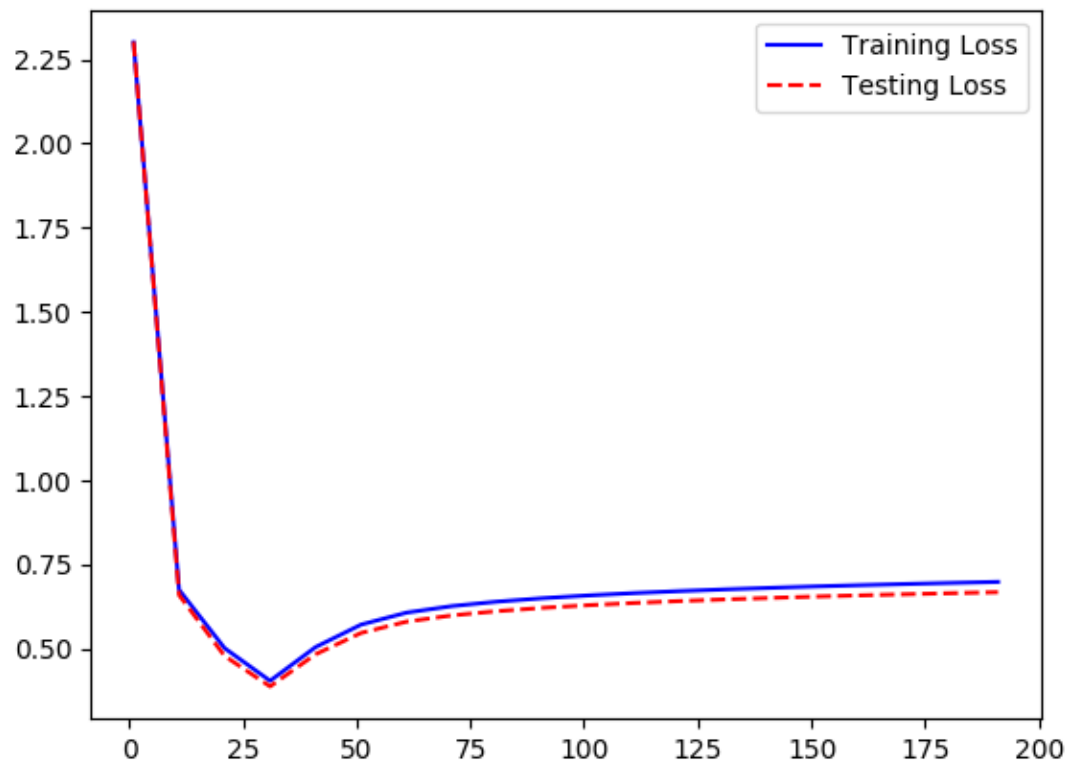
2. a) (2048)



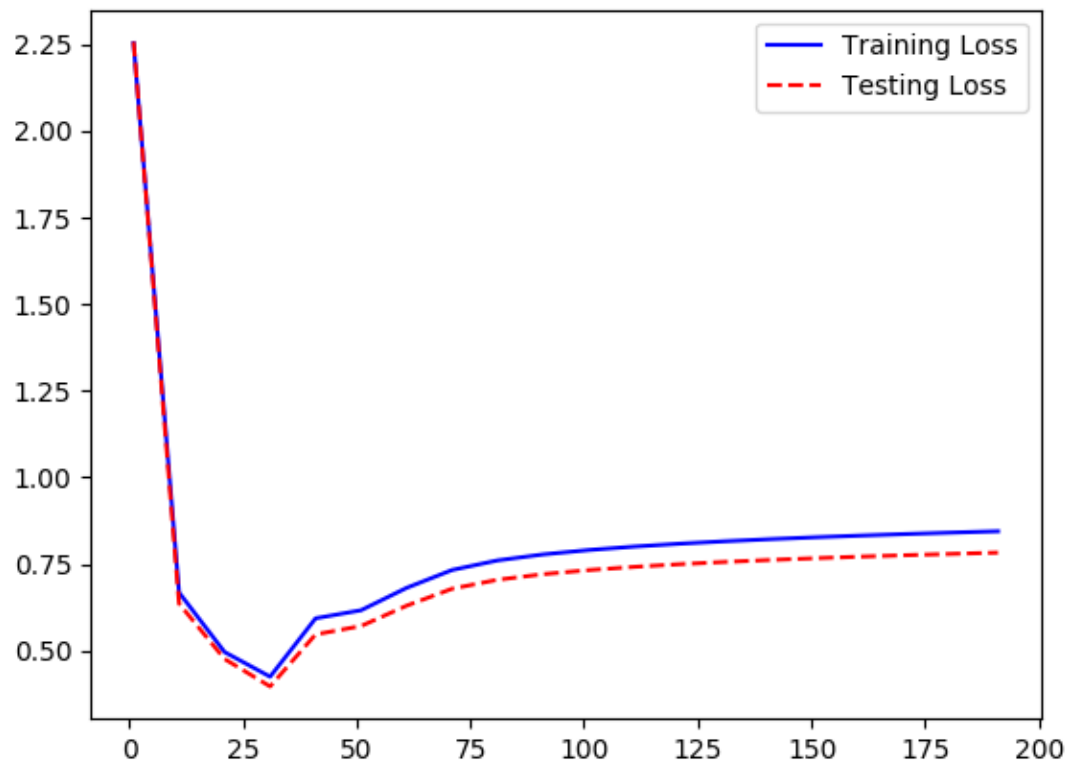
b) (1024, 512)



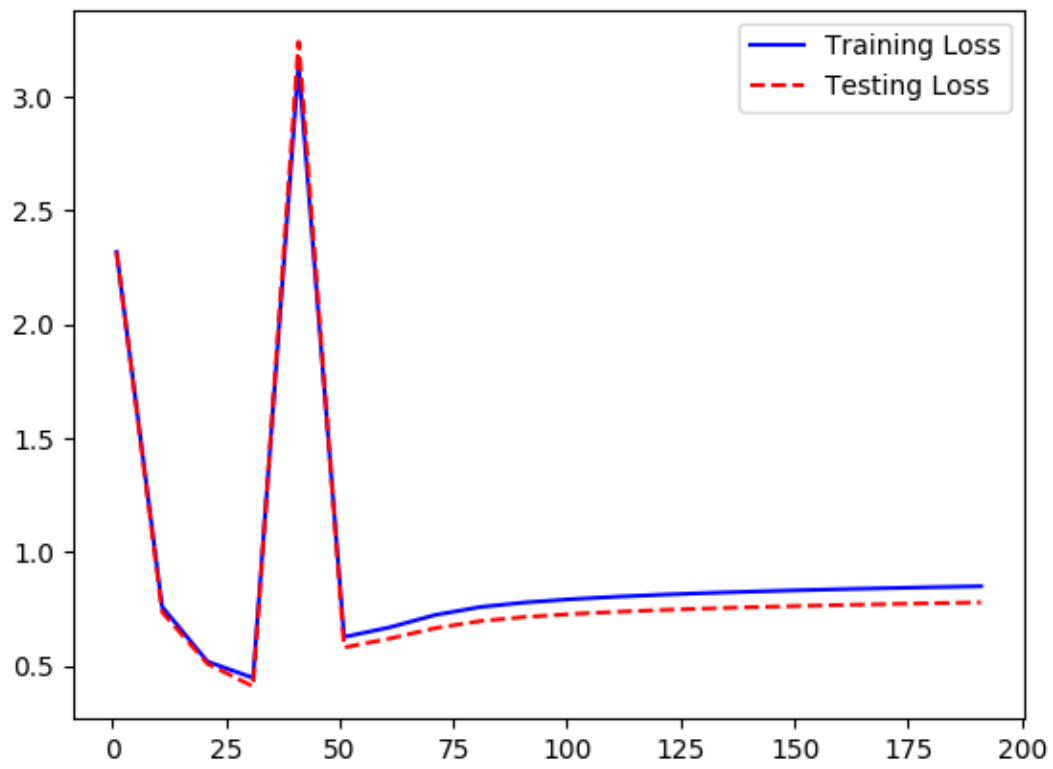
c) (512, 256, 128)



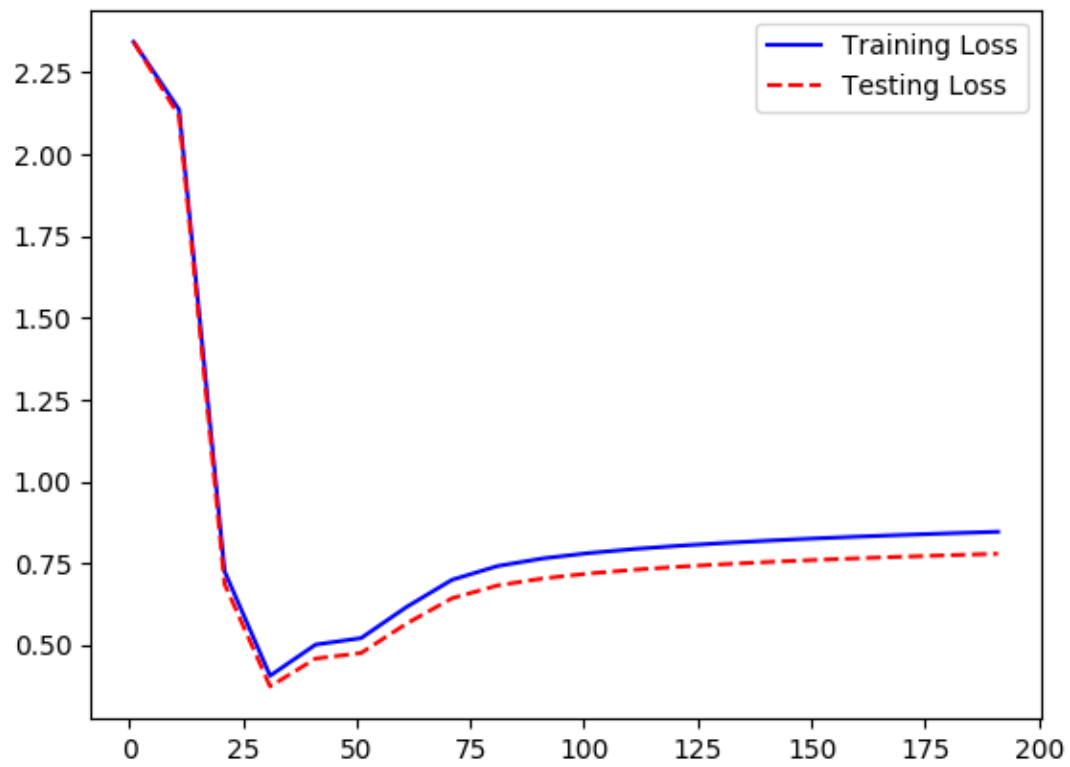
d) (256,128,64,32)



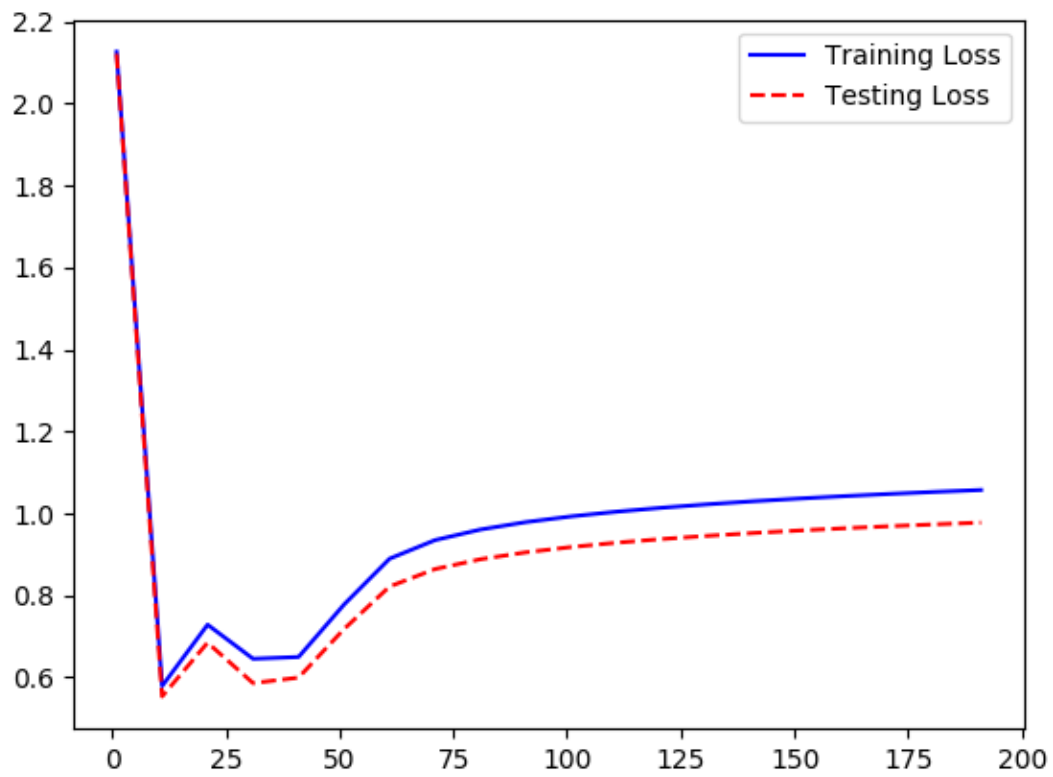
e) (256,128,64,32,16)



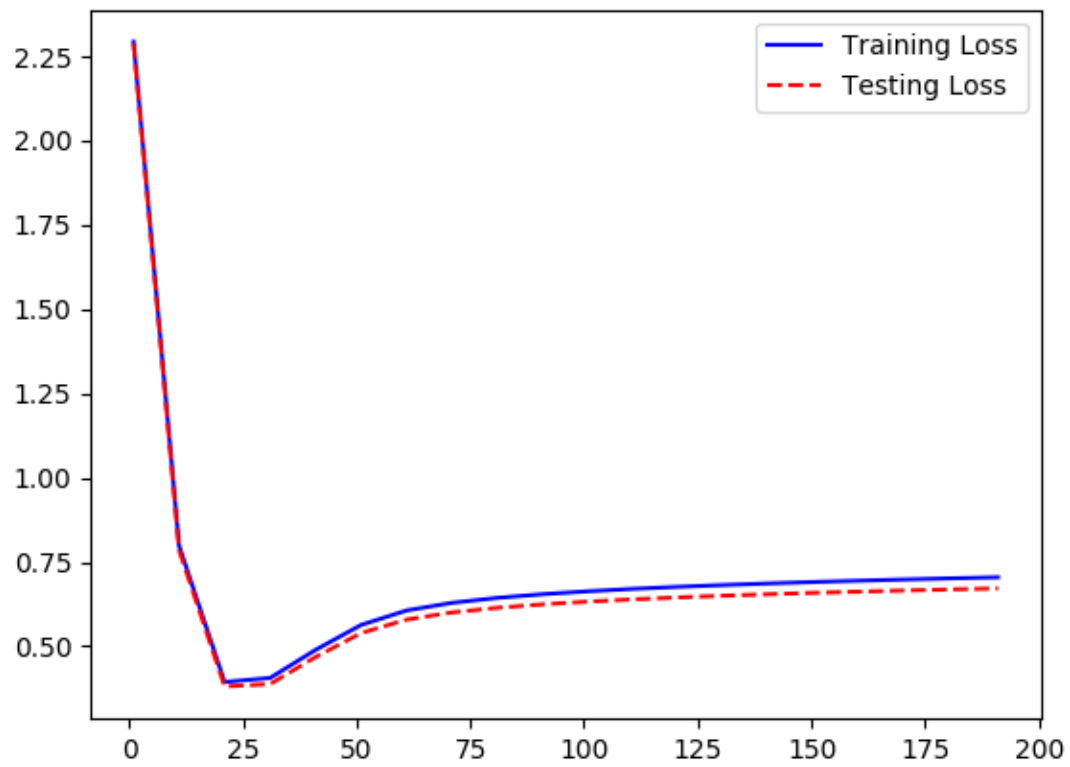
3. a) Sigmoid



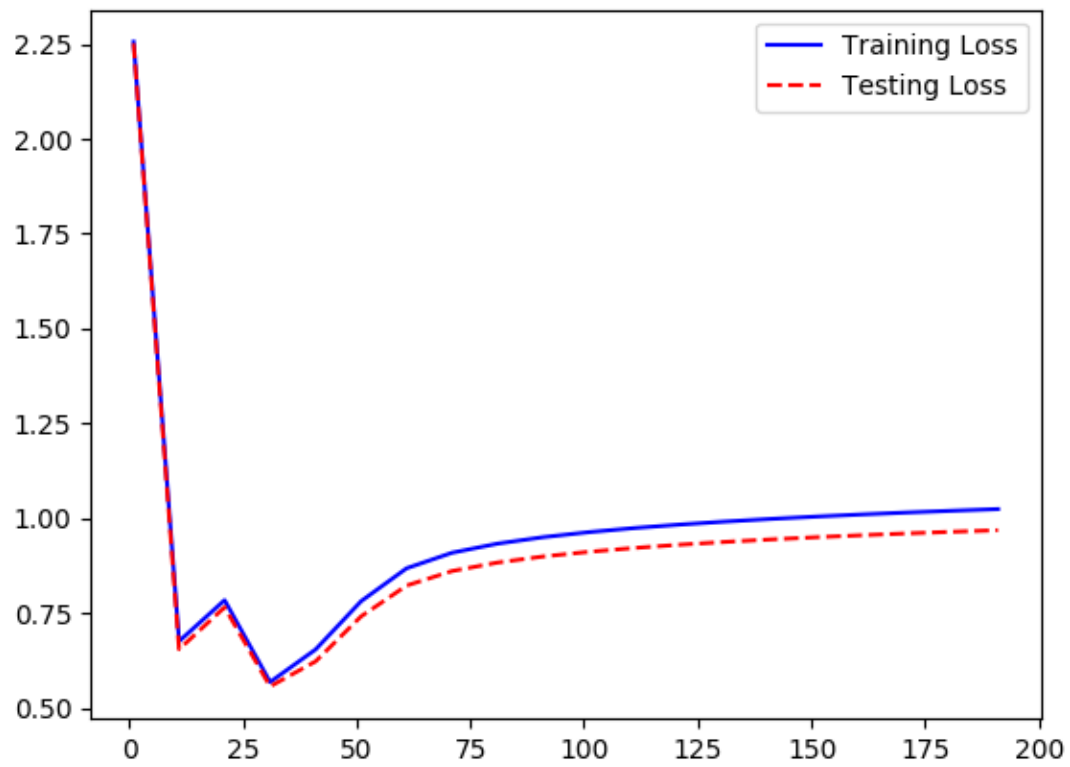
b) Tanh



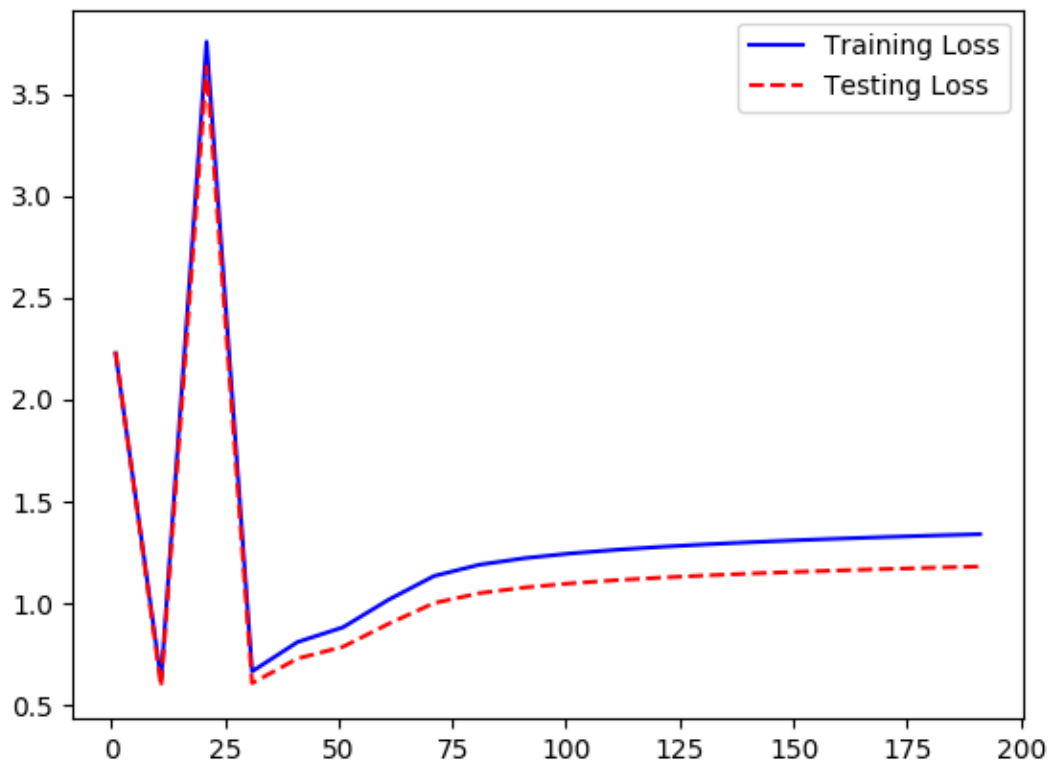
c) Relu



d) Leaky Relu



e) ELU



4. For question 1, changing the number of neurons affects the performance of the network, keeping the number of layers fixed by increasing the training accuracy and lowering the training loss for the most of the first 30 iterations then it gradually increases. For question 2, changing the number of layers affects the accuracy of the network by lowering the training accuracy as more layers were applied. For question 3, the activation function affects the performance of the networks with the Sigmoid the training accuracy didn't deviate much and the training and test loss were nearly identical for the first 30 iterations, the Tanh had a sharp decrease in training and test loss and then a short spike upwards, the Relu the training and test loss were nearly identical, the leaky relu has a plot similar to the tanh and the loss gets slightly more positive, and the ELU has a sharp spike in the plot and has higher accuracy.