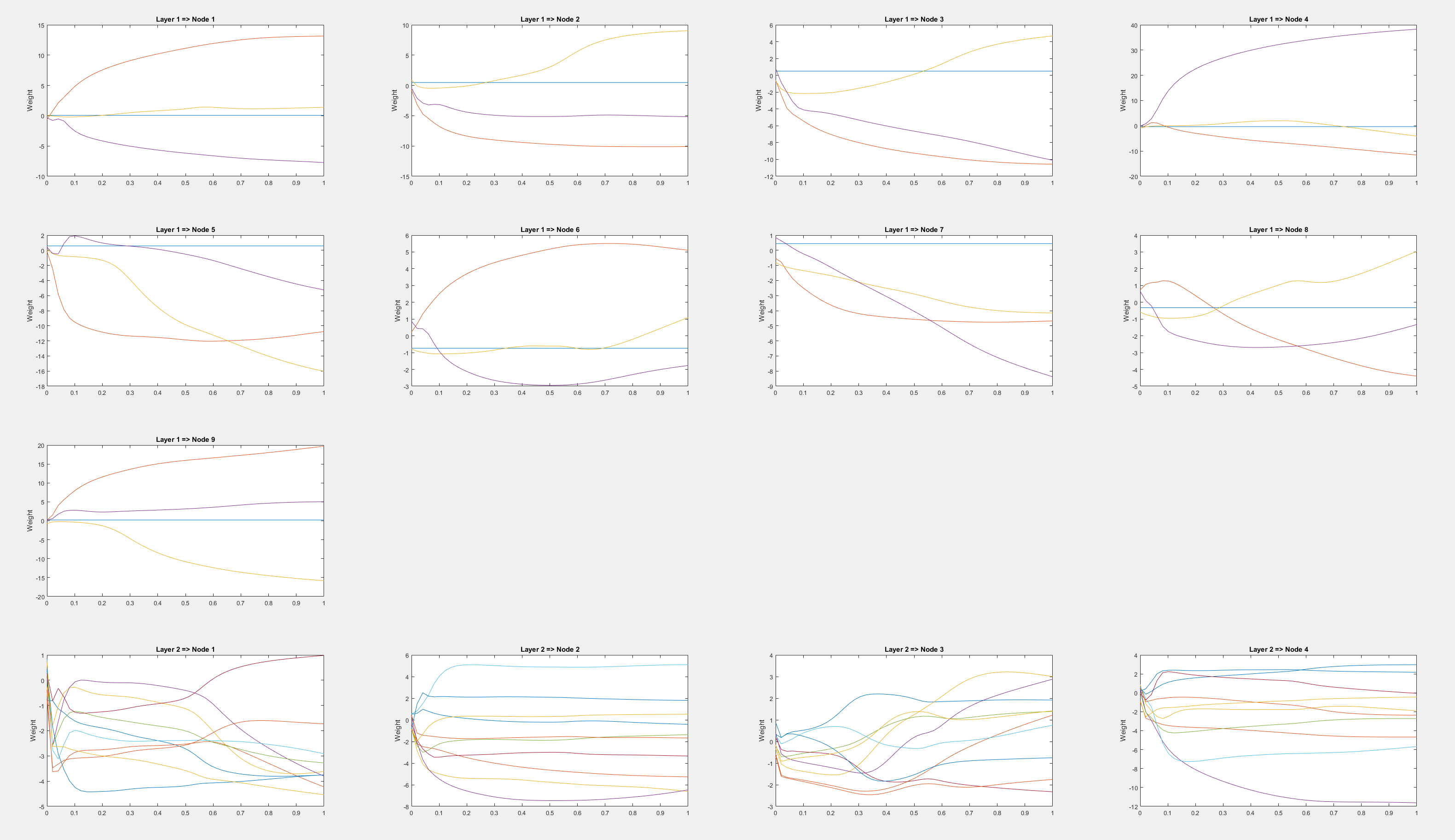
## ECS 171 Homework Set 2 Haozhe Gu 999200555

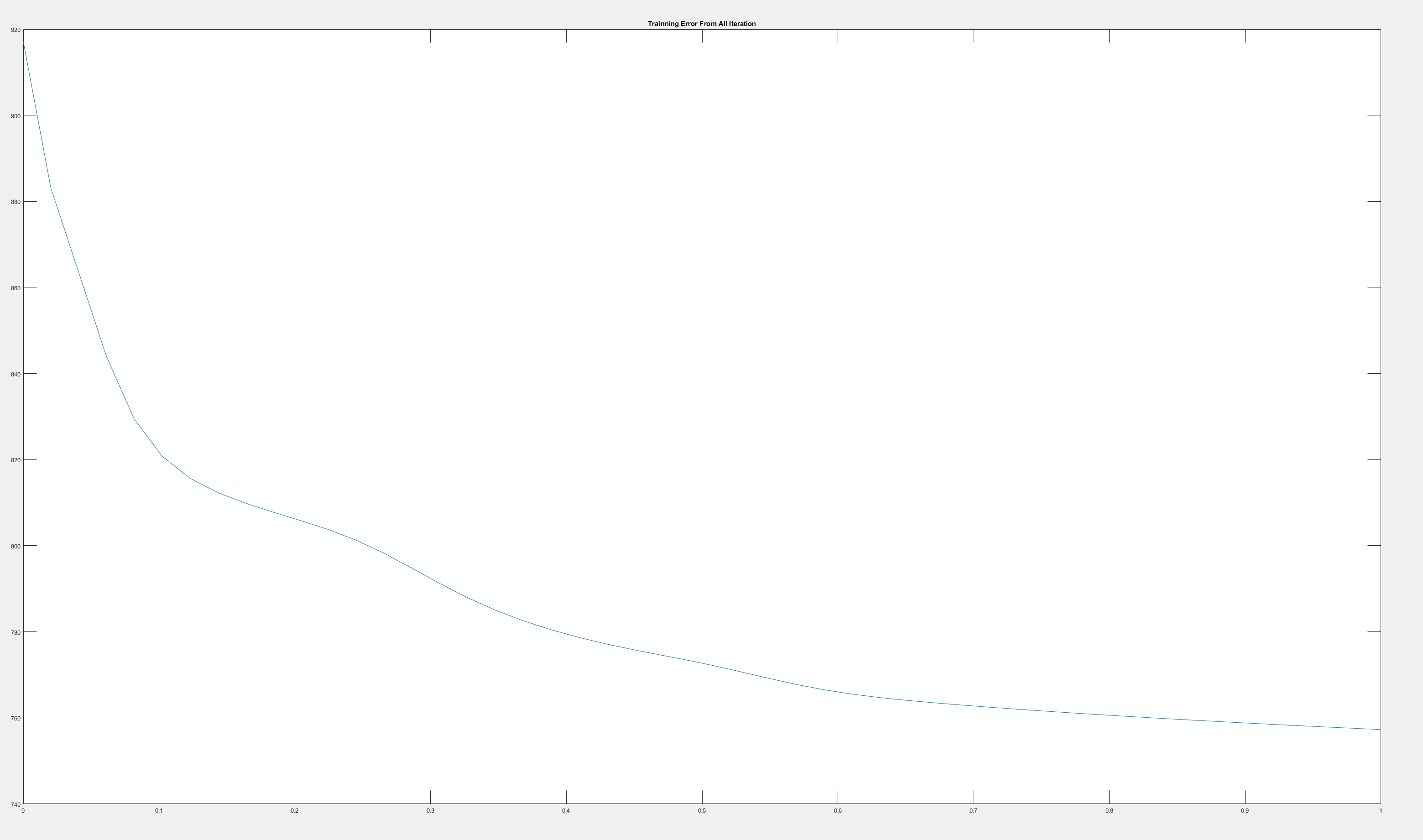
1. Only Training have iterations, so Training Weight,Error are based on iteration. Training Output is based on the the single minimized error round.

For the testing case, there’s no iteration need. Also, weight does not change. So there are only plot for single minimized error round Error and Output.

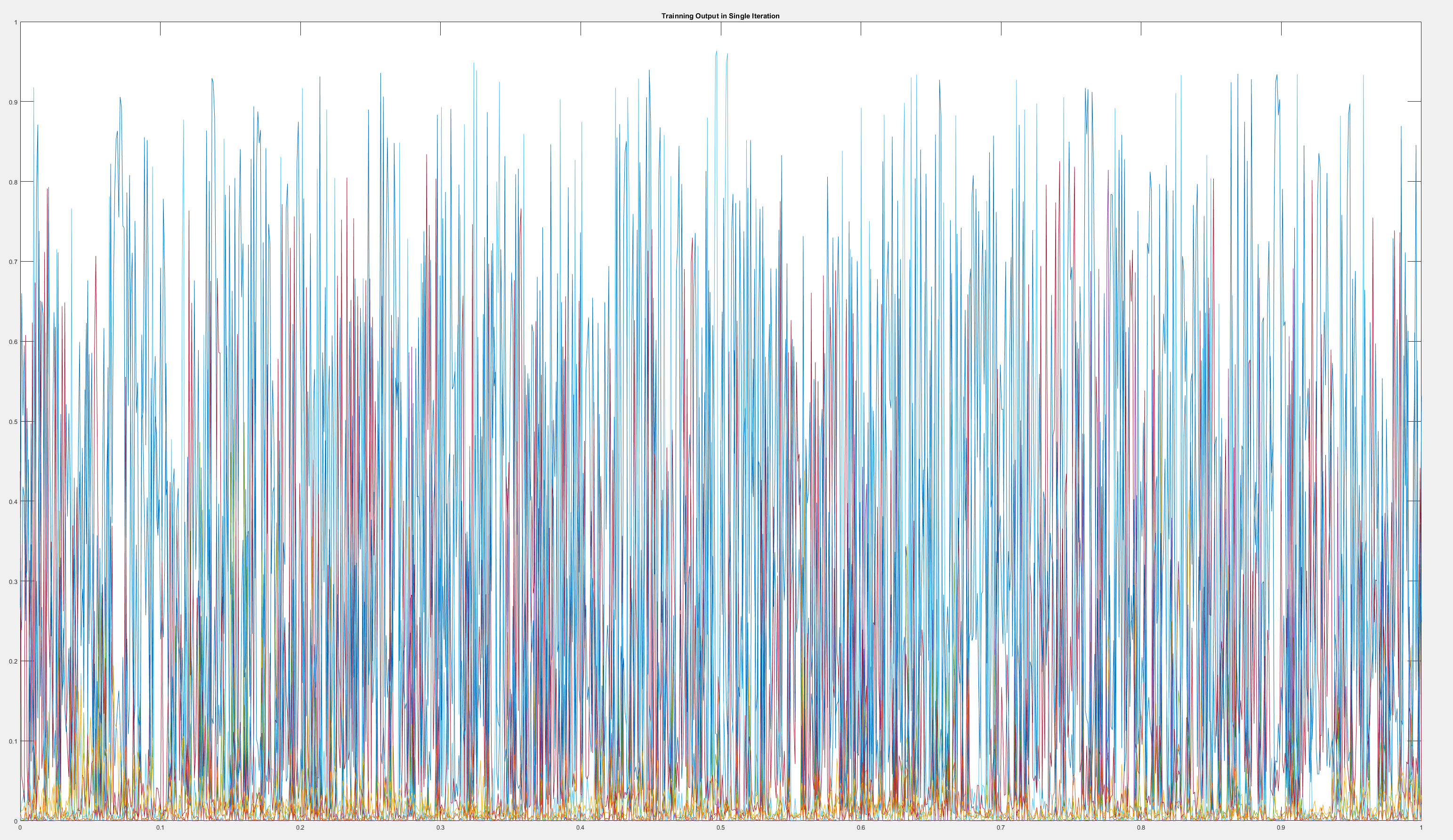
Training Weight Graph



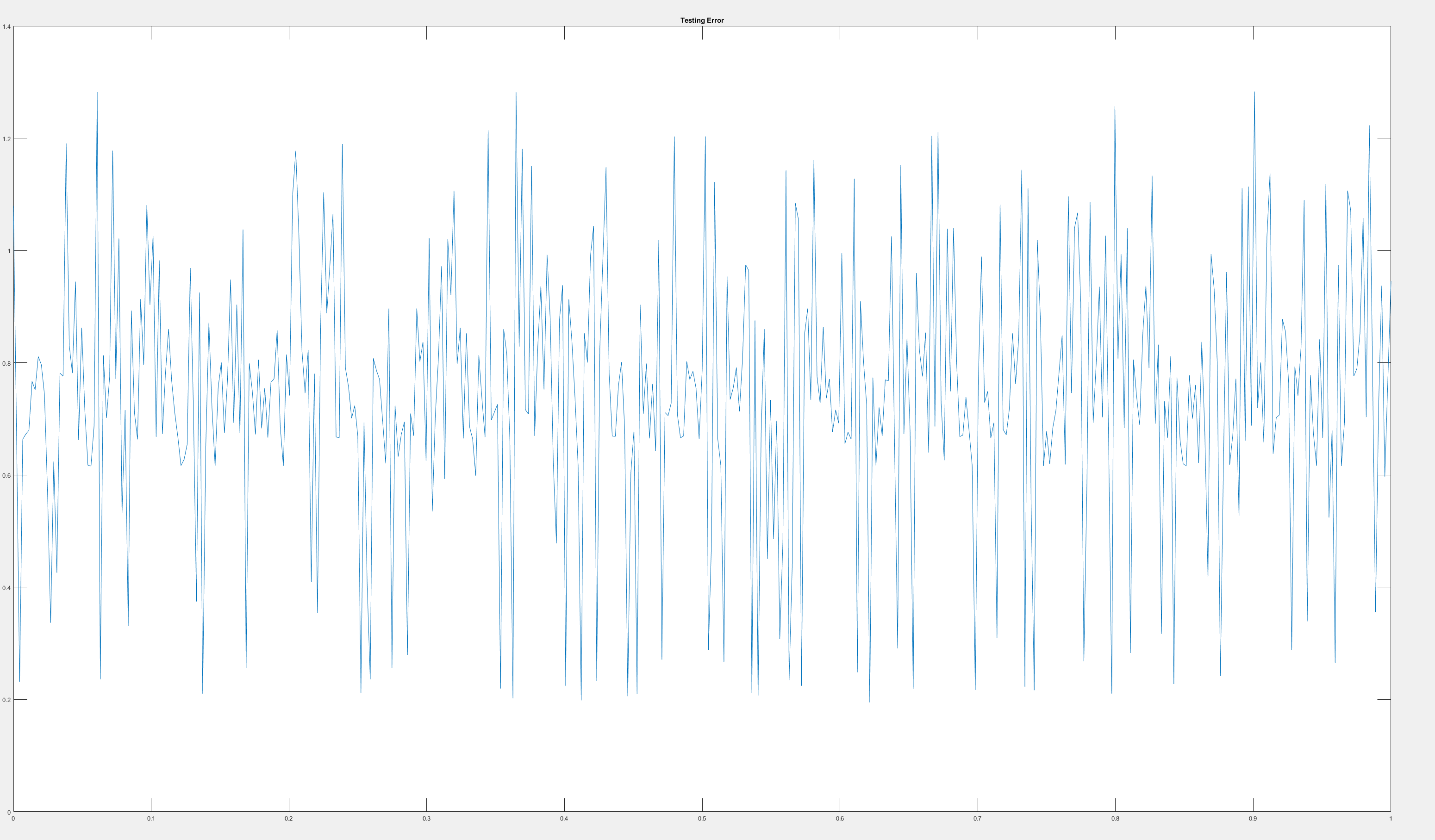
Training Error Graph



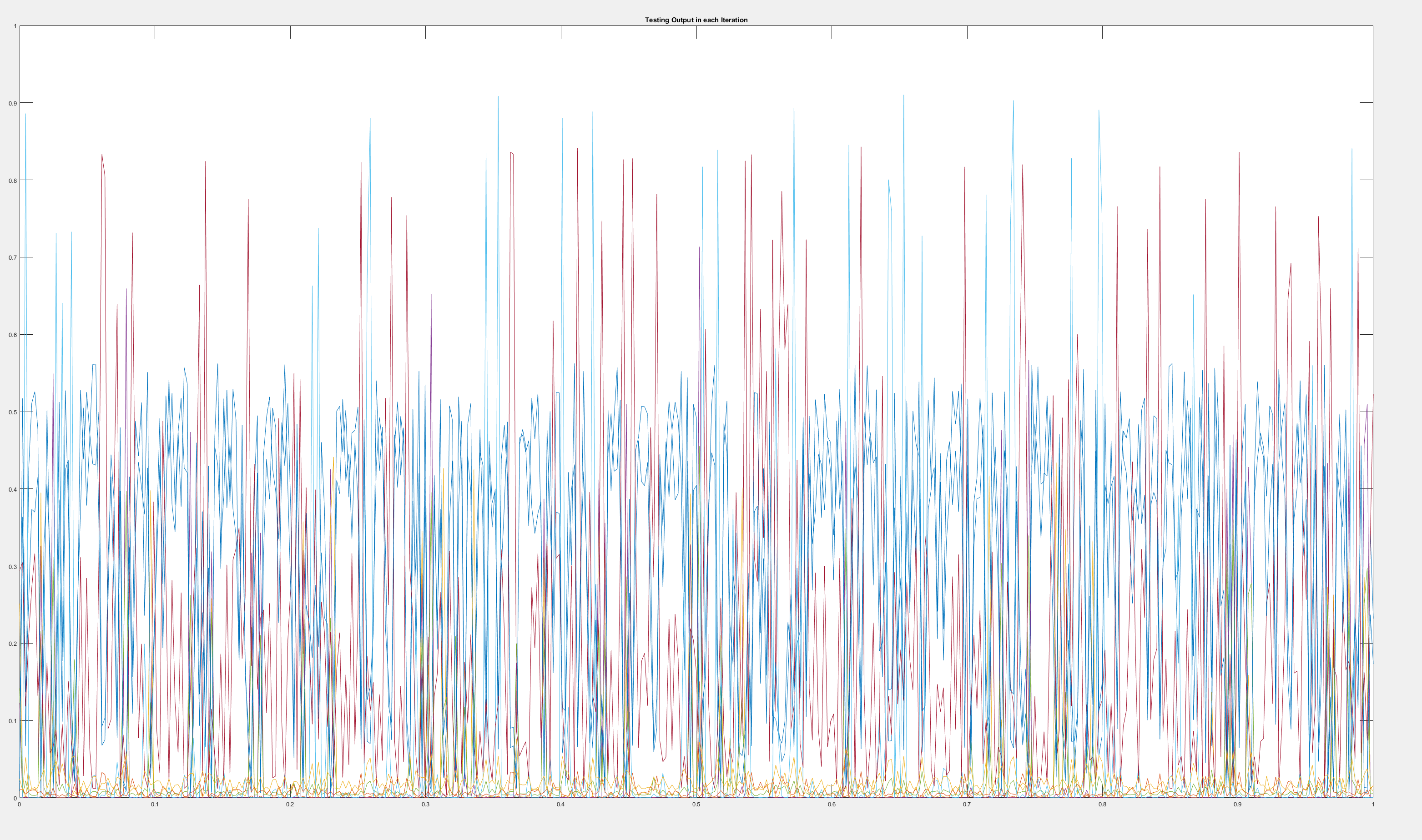
Trainning Output Graph

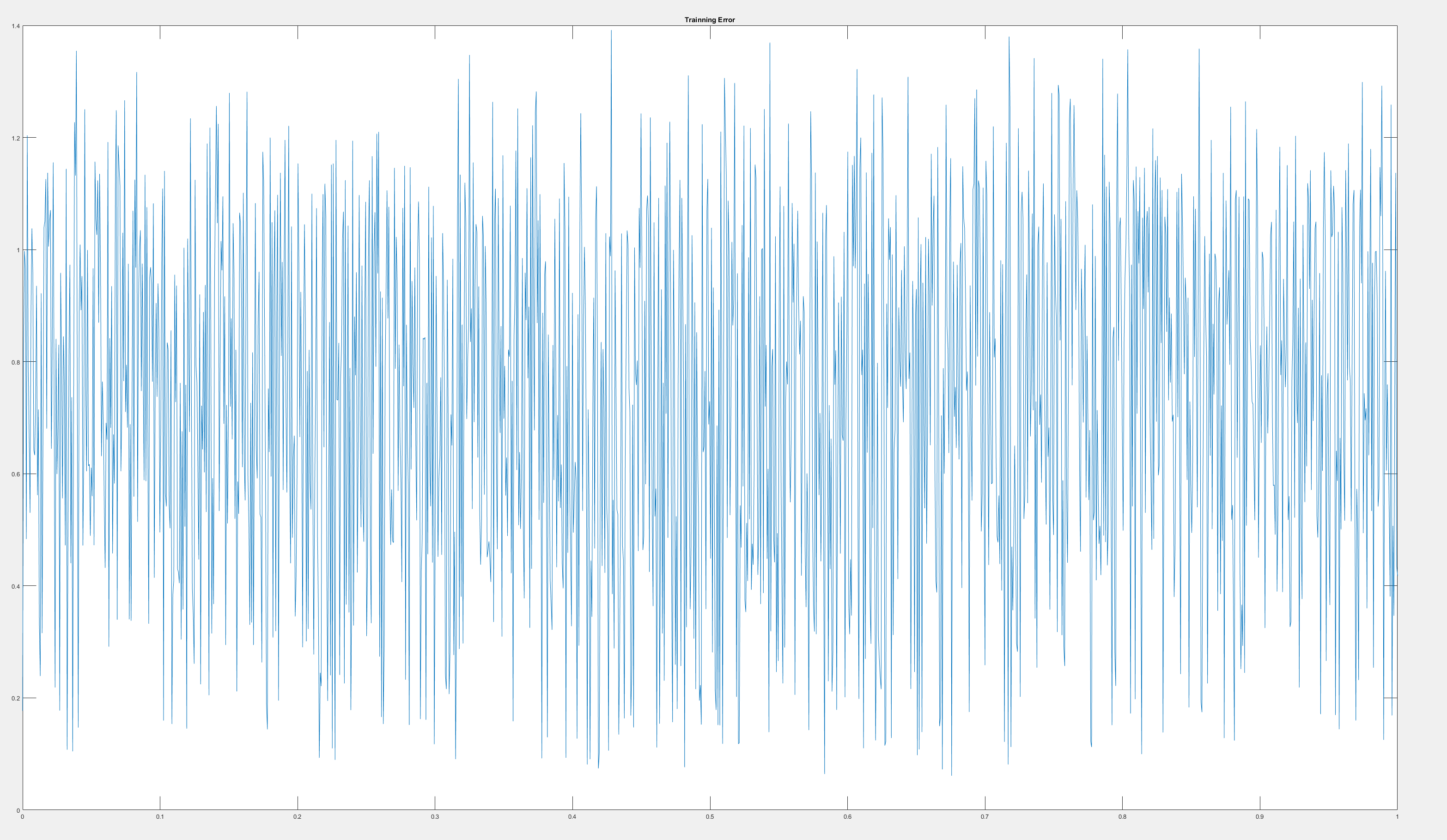


Testing Error Graph



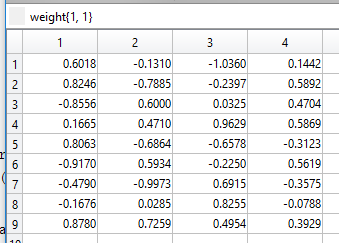
Testing Output Graph

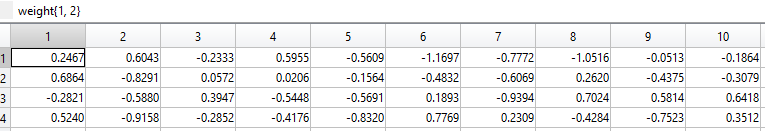


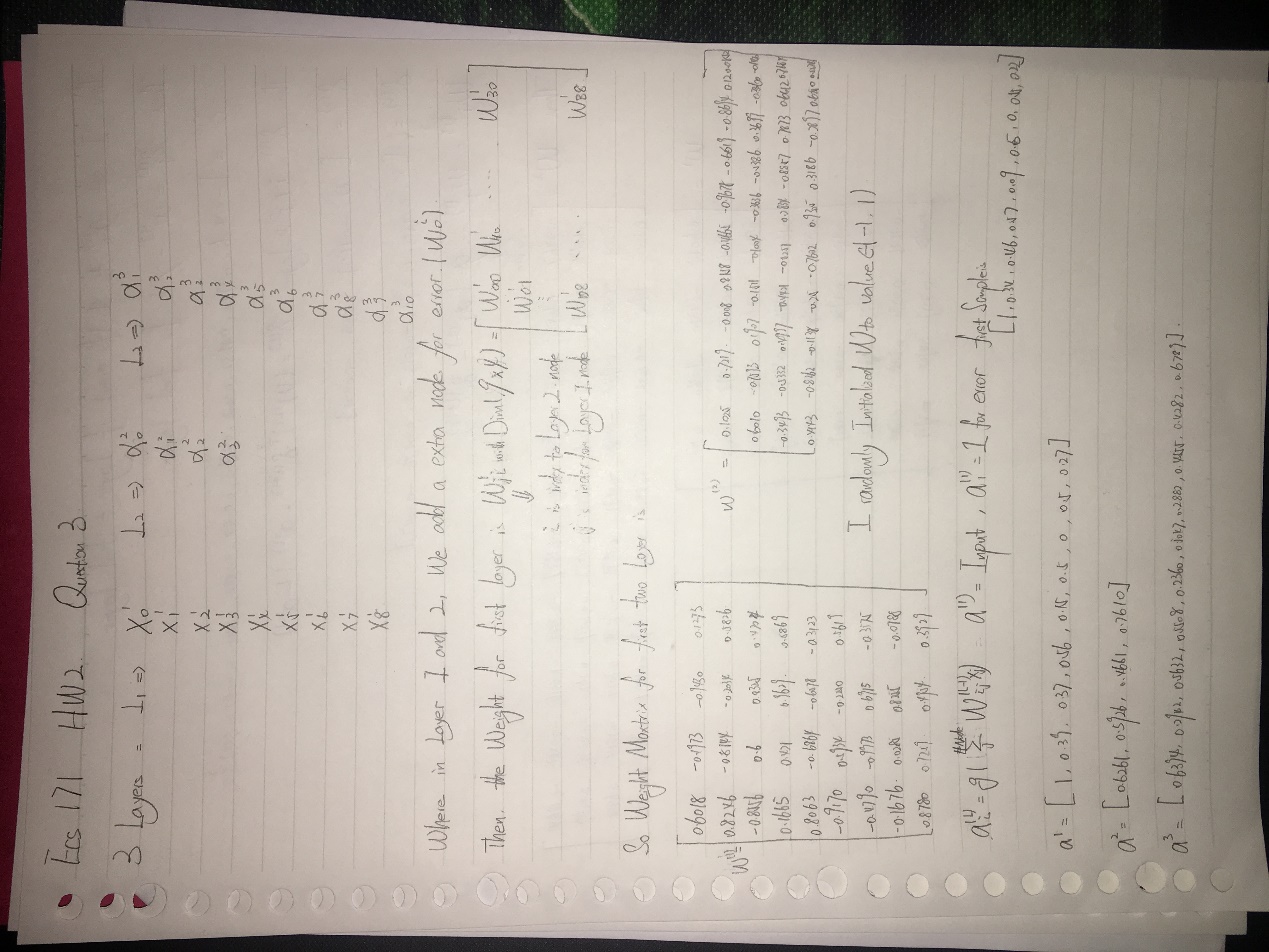
1. Training Error 

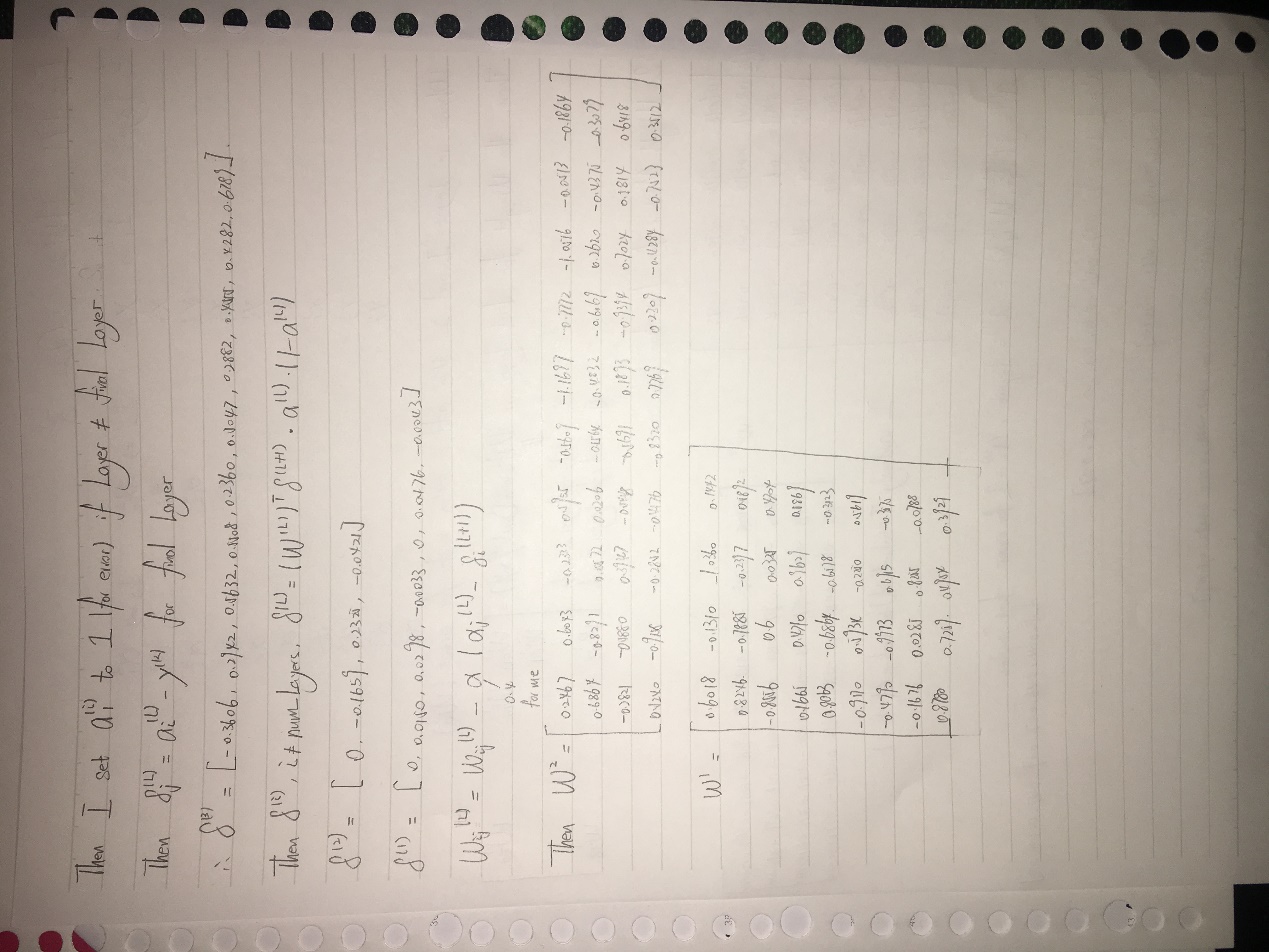


1. Output From Code ( First is Weight for first Layer, Second is Weight for Second Layer)





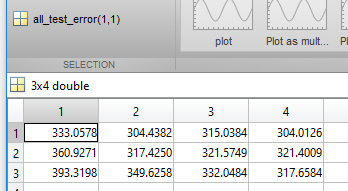


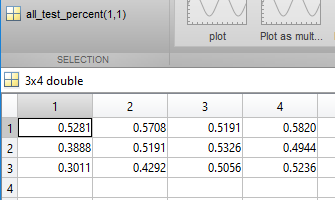


Both are in agreement.

1. Rows are Number of Layers {1,2,3}, Columns are Number of Nodes per Layer {3,6,9,12}

The Second Graph is the corresponding Testing Classification Rate





From the two matrix, the Optimal Configuration is 1 Hidden Layer with 12 Nodes per Hidden Layer.

The Relationship between these attributes can be summarized as: Generally, increase in Nodes per Hidden Layer will reduce the Error, while increase in Number of Layers might cause over-fitting which actually increase the Error in the test.



