

# Module 5 Report - Johannes Moskvil

15. november 2015 19:02

## Constructing networks

To begin with I tried random configurations of number of hidden layers, number of nodes at each layer, activation function and learning rate. Early results suggested I got best performance using three layers where each layer contained 30-50 nodes and using the Tanh activation function. However, in order to make sure I found one of the best solutions in the range [5, 50] nodes using three hidden layers, I created a script that would test every possible combination of layer configurations in that range, record the result to a file that I could later search for the best configurations. The script ran on my computer for a whole week. It would train every network for 5 epochs on the whole 60000 examples and recording the performance on both training and test data. This resulted in roughly 800 different network configurations when consider every combination with a 5 node interval. The script would first try [5, 5, 5], then [5, 5, 10], [5,5,15]...[30,10,50],[30,15,5]...until [50, 50, 50].

The data collected from the whole week of testing configurations yielded the following network as the best:

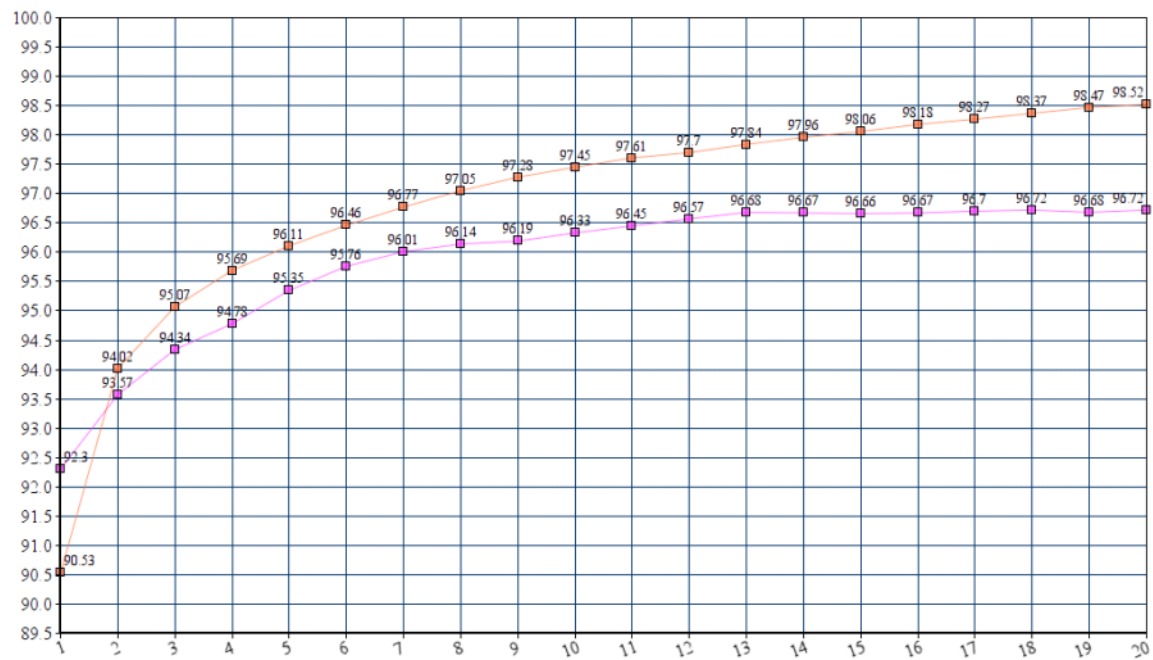
- [50, 40, 50]
- Learning rate = 0.015
- Activation function = Hyperbolic Tangent

Other network configurations considered were:

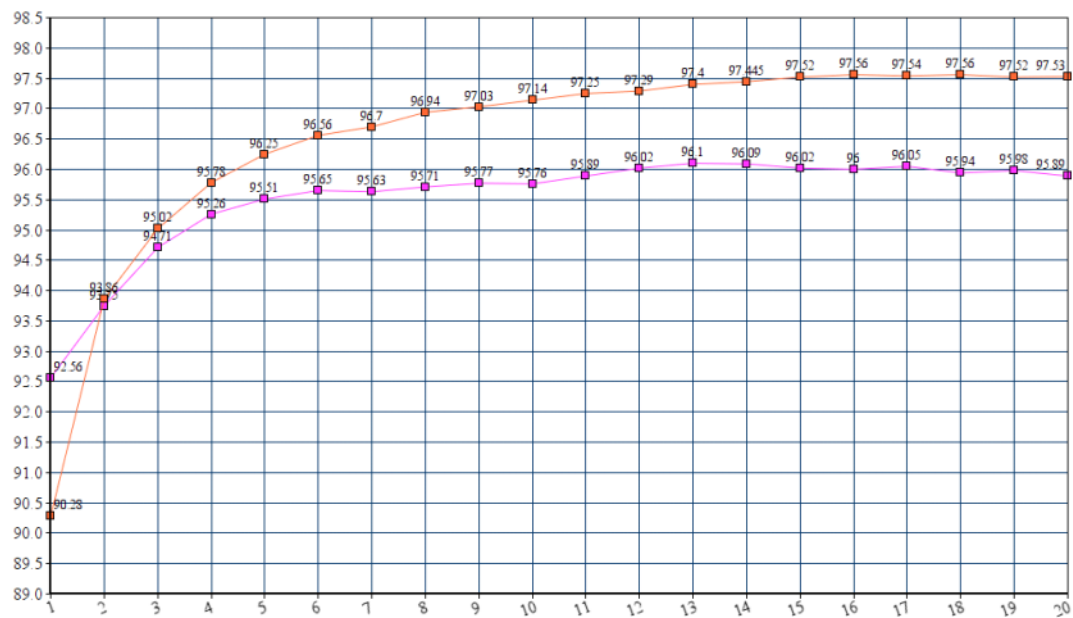
- [45, 30, 40]
- Learning rate = 0.015
- Activation function = Hyperbolic Tangent
- [50, 30, 40]
- Learning rate = 0.015
- Activation function = Sigmoid
- [600]
- Learning rate = 0.015
- Activation function = Hyperbolic Tangent
- [50, 40, 50]
- Learning rate = 0.015
- Activation function = Sigmoid

Following are some graphs demonstrating the results of each of these networks under training, comparing the training and test performance. The graphs show the correct classification of training and testing data for each epoch, each graph shows the progression over 20 epochs over all 60000 examples.

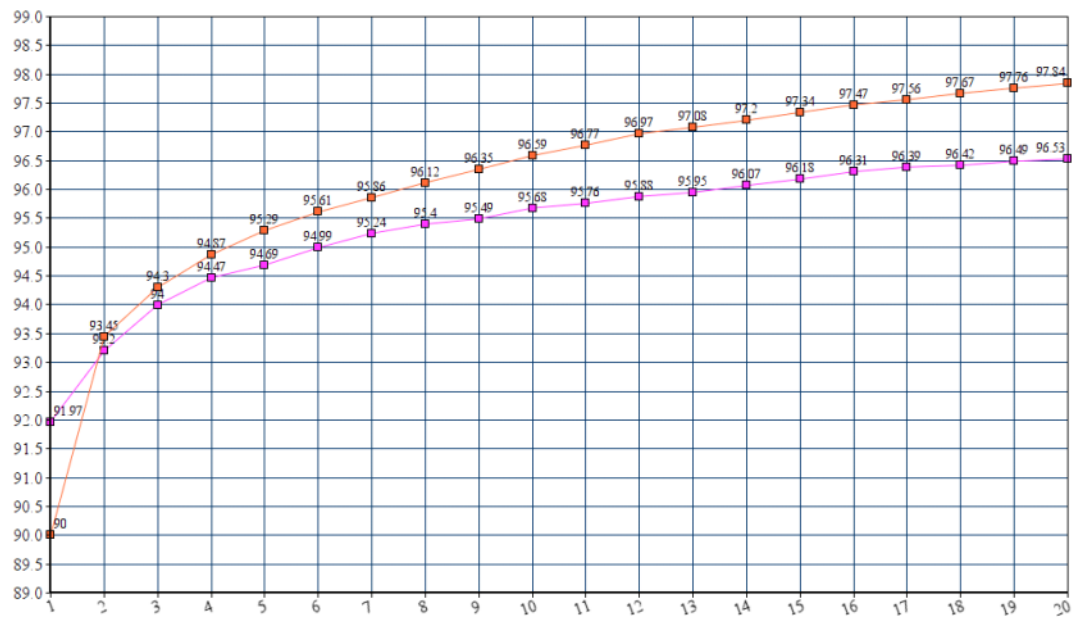
### [50, 40, 50] - Hyperbolic Tangent:



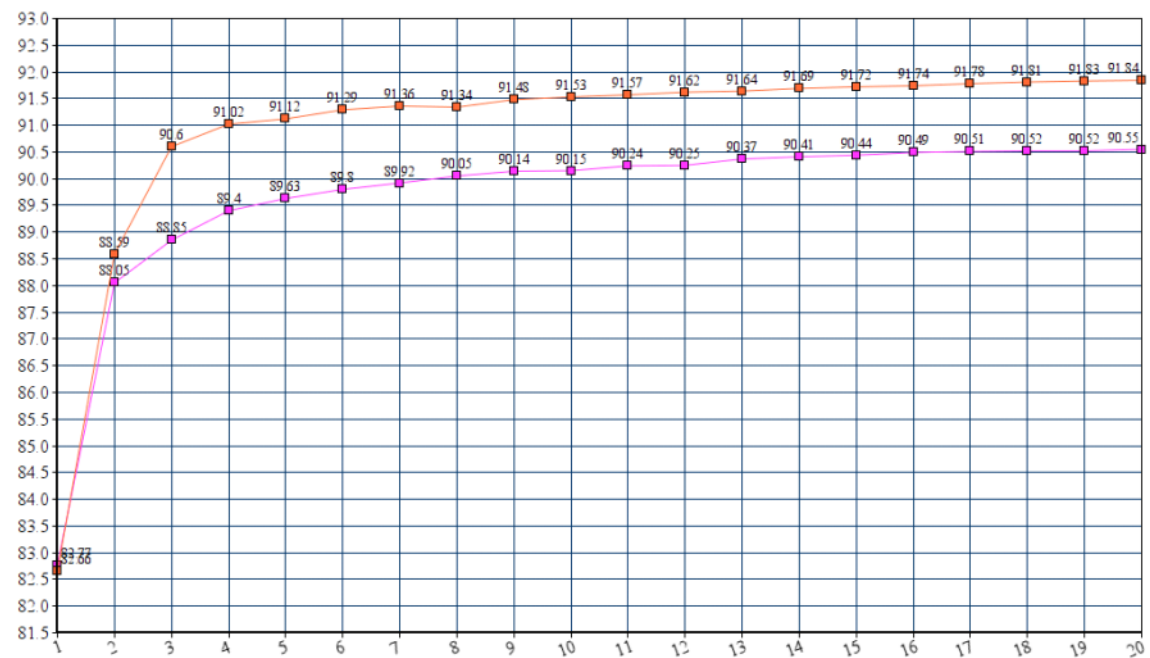
### [45, 30, 40] - Hyperbolic Tangent:



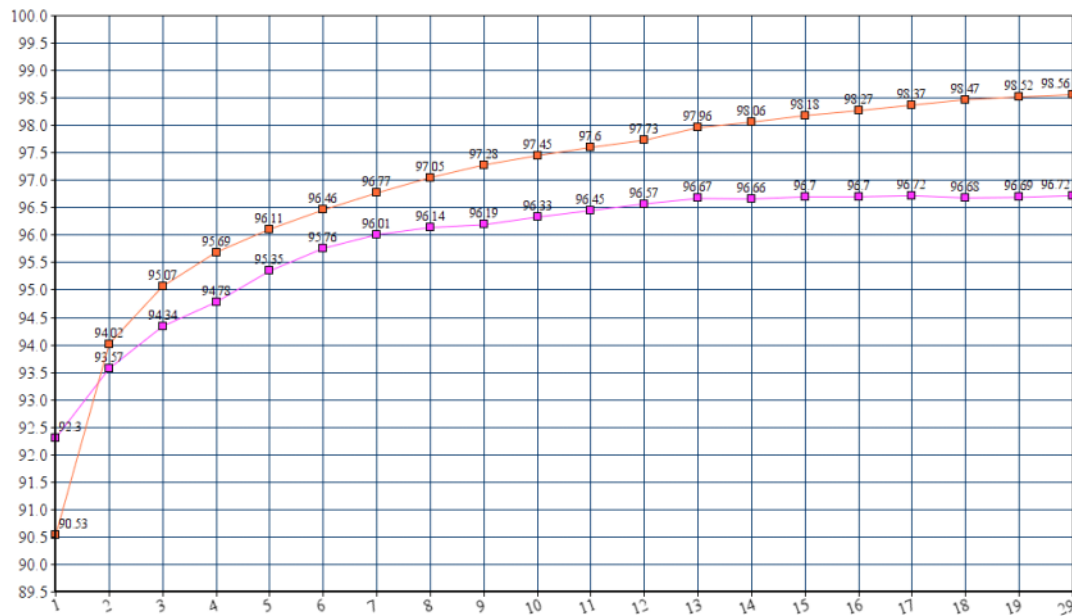
## [70, 60] - Hyperbolic Tangent



## [650] - Sigmoid



## [50, 40, 50] - Sigmoid



As we can see from the graphs, the chosen network configuration ([50, 40, 50] - **Hyperbolic Tangent**) yields the best result, peaking at 96.72. Actually [50, 40, 50] - **Sigmoid** had the exactly same peak, seems the topology matters far more than activation function for this problem. After 20 epochs, all of these examples converges to the value at the 20 epoch, or overfits. For example [45, 30, 40] - **Hyperbolic Tangent** seems to start overfit already; towards the end the training accuract increases, while the testing accuracy decreases.

Final note: the [50, 40, 50] sigmoid and tanh graphs start out EXACTLY the same which is maybe a bit too much of a coincidence, maybe I messed up creating the graphs (though I believe I did not).