Handwritten Digit Recognition Neural Network

Akshay Kumar, Rohin Mittal and Tushar Kulkarni

Summary

This report summarizes various test approaches for selecting hyper parameters λ *and* m .

Based on various test runs, the optimum value of λ **was found to be 0.4** and optimum value of m **was found to be 88**. For these hyper parameters, the accuracy of **93.88%** was achieved.

Observation

We started analysis by running script for all values of λ between 0 and 10 and m between 2 and 100. Following were the observations:

1. Accuracy drastically reduced as the number of hidden nodes fell below the number of output nodes.
2. For λ, maximum accuracy was observed in the range of range of 0.3-0.5 for different values on m
3. For m, maximum accuracy was observed at m=88 in the above range of λ

On further analysis and scraping through various articles on internet, to understand the above observation, we found that :

1. Number of hidden nodes should be between the size of input layer and the size of output layer.
2. According to [1] optimum value of hidden nodes is $\sqrt{x * o}$ where *x* is number of features in input layer and *o* is number of nodes at output layer.
3. A model with extremely high regularisation, i.e high lambda, would not appropriately fit the training data. The 2-D surface making the classifications will be too smooth, and there are bound to be incorrect predictions. However if the lambda is too small, it would overfit the training data. It would adapt to the irregularities in the training set, causing the surface to have arbitrary ripples which would give bad predictions for the test data.

For our neural network, we had 785 input nodes (including bias) and 10 output nodes. So according to [1] optimum value of hidden nodes should be $\sqrt{785 * 10} = 88.6 \approx 88$, which is coherent with our observation.

Following are the graphs depicting our analysis and results:

We started with 100 hidden nodes, since a few extra nodes in the hidden layer won't do any harm, convergence will still happen.

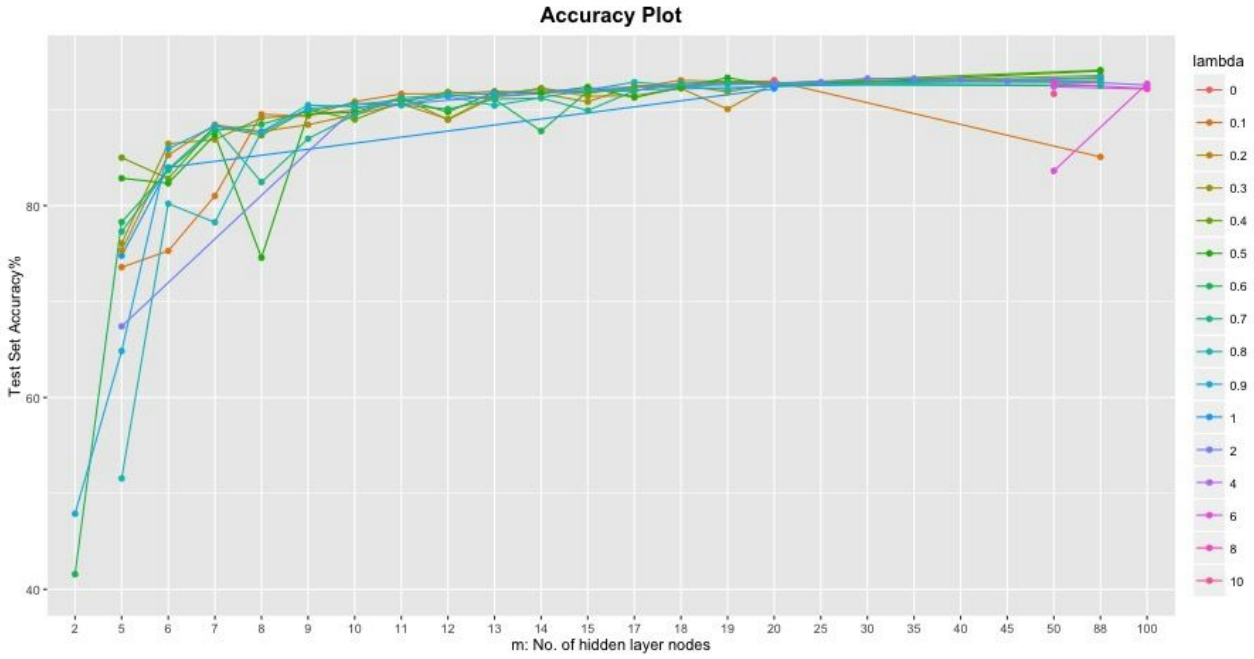

Fig. 1 Accuracy vs m $\subset$ (2,100) for all values of lambda

When the hidden nodes are less than 10 (number of output nodes) the accuracy reduces drastically. This may be due to underfitting. Based on the observation we started pruning the hidden nodes and found the values as shown in Fig. 2.

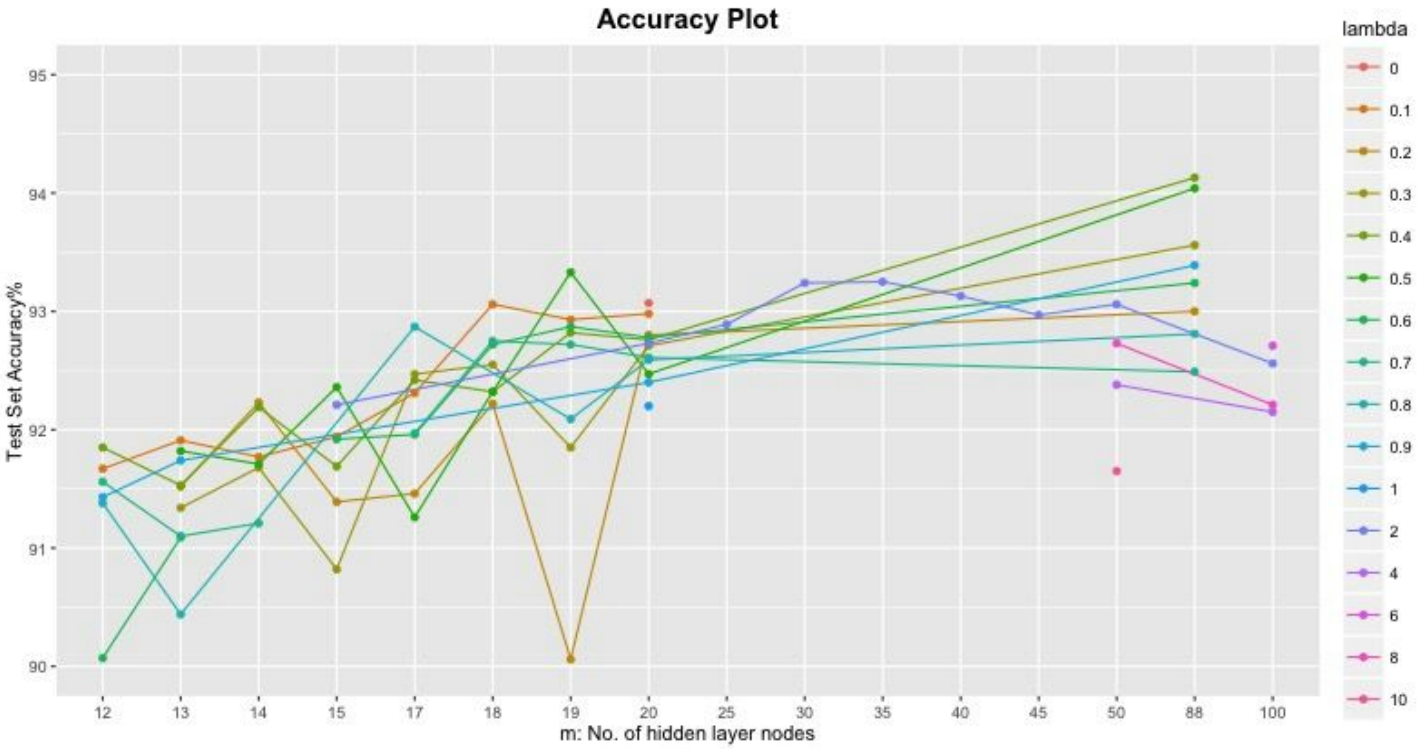

Fig. 2 Accuracy vs selective $m \subset (12,100)$ for all values of lambda

After reducing m to a handful of promising values, we observe that *m*=88 has the highest accuracy for *lambda* ranging from 0.2 to 0.6. This reinforces the heuristic of the ideal number of hidden nodes being equal to $\sqrt{x * o}$. Fig.3 depicts this observation.

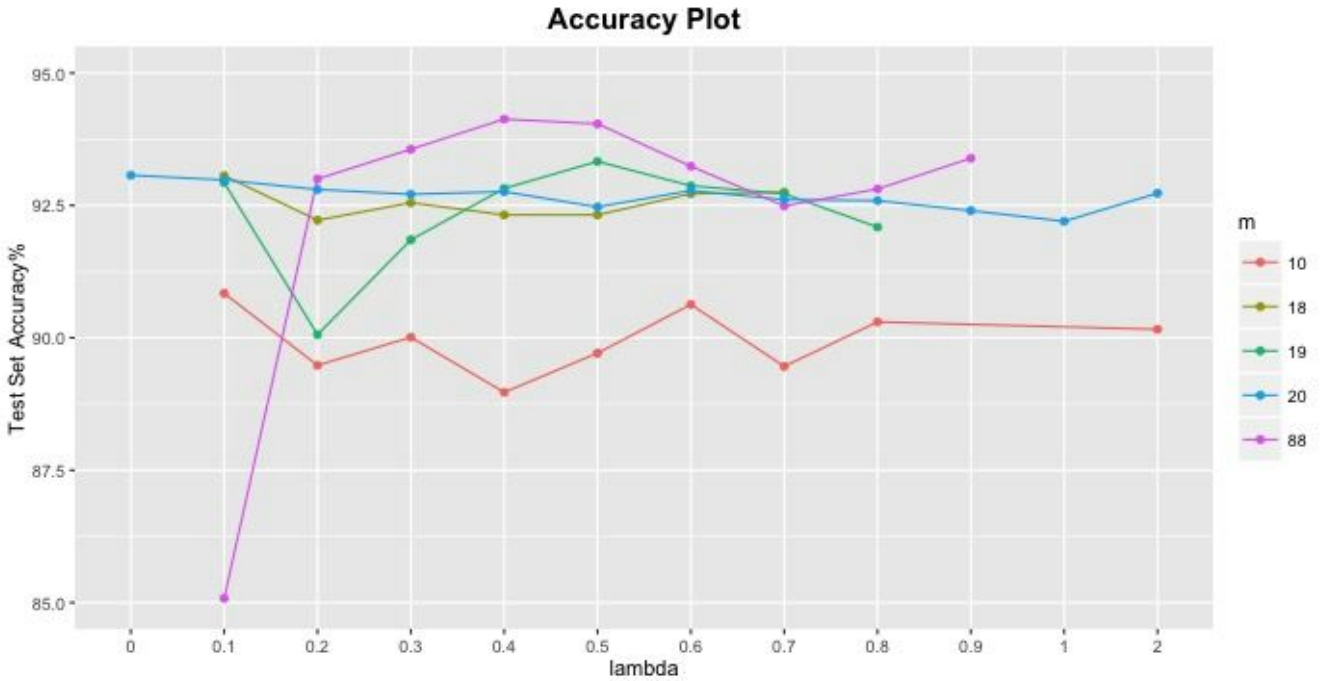

Fig. 3 Accuracy vs lambda $\subset (0,2)$ for selective values of m

During the training of the network we logged the changes in error values till it converged. Fig. 4 depicts the error minimization during the training phase of the neural network.

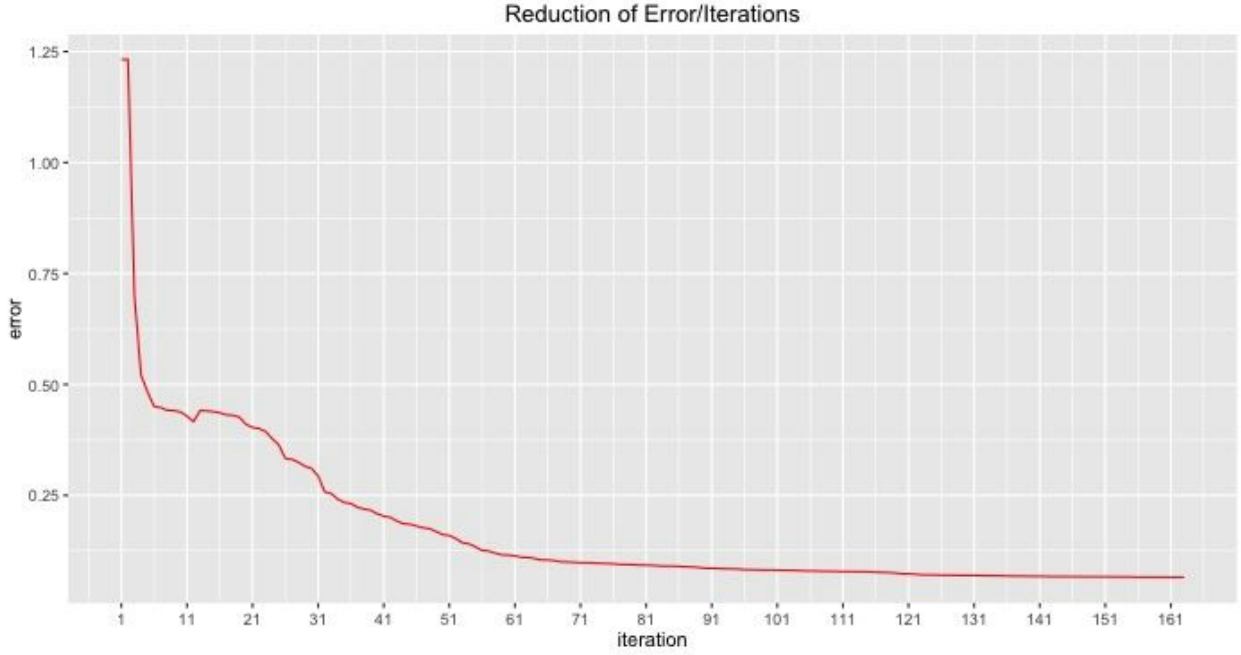

Fig. 4 Error vs No. of iterations for m = 19 and $\lambda = 0.5$

References:

[1]
http://stats.stackexchange.com/questions/181/how-to-choose-the-number-of-hidden-layers-and-nodes-in-a-feedforward-neural-netw