Handwritten Digits Classification

Team Number 25: **Course Number: CSE 574**
SHIVA MARISWAMY SUBRAMANI - 50133425
ANANDA KANAKARAJ SANKAR - 50133315
KARTHIK JANARTHANAN - 50133543

Implementation:

Neural Networks is organized into 3 layers. Input layer, hidden layer and Output layer respectively. Each layer has nodes connected to the next layers with weight vector applied on the interconnections. Implementation involves series of feed forward and backpropagation. The output layer's output is compared with the original result and checked for accuracy. The error is fed back into the network through back propagation and corrections to the weights are made according to the calculated error.

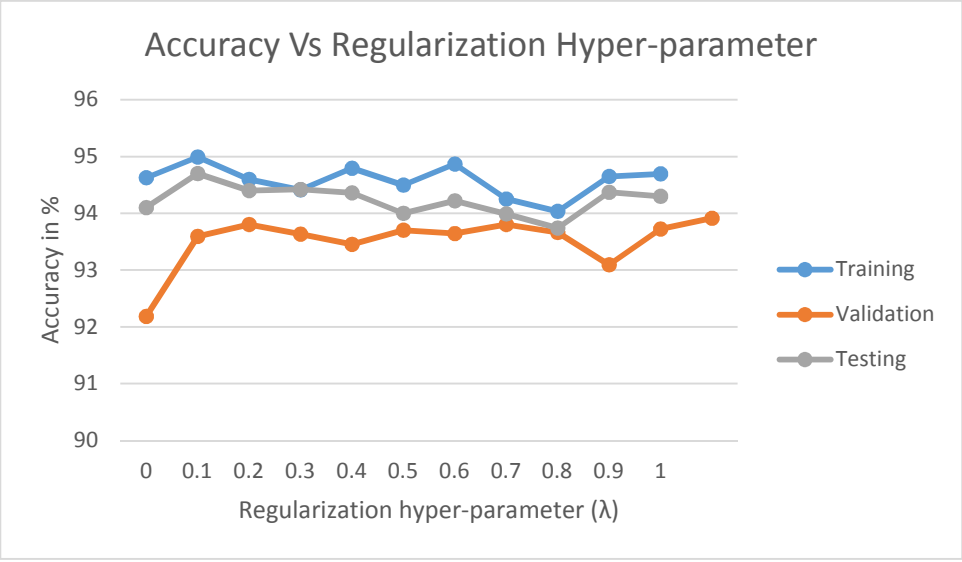
Performance Metrics:

Accuracy: 94.71 %

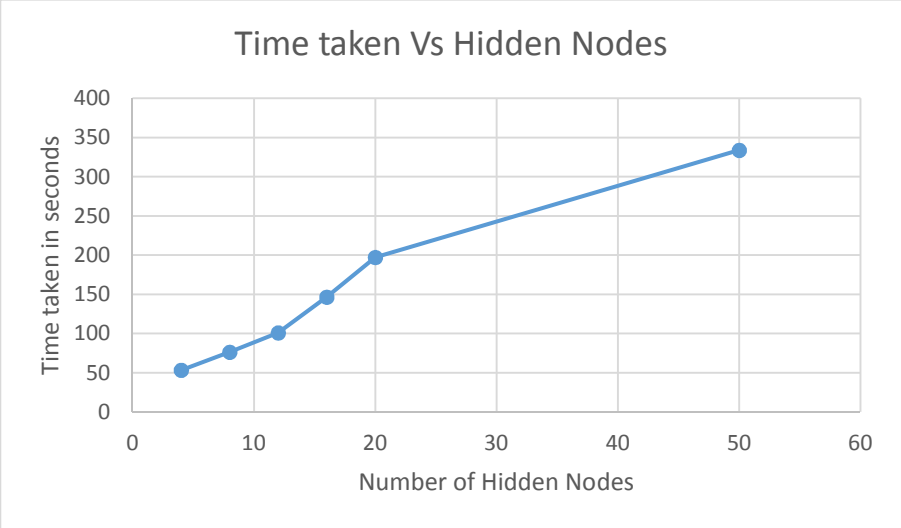
No of hidden nodes: 50

Regularization Hyper Parameter: 0.1

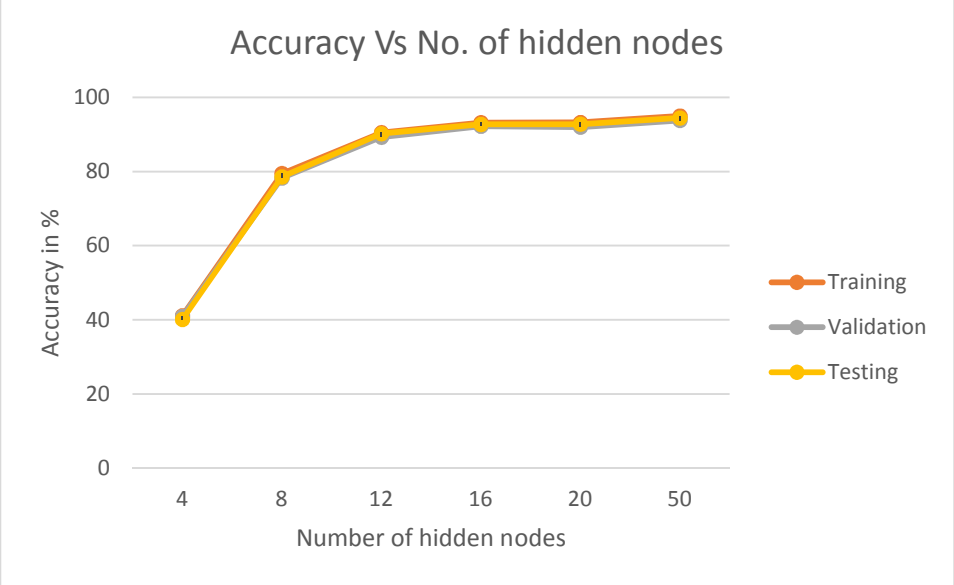
Accuracy Vs Regularization Hyper Parameter for 50 hidden nodes:



Time taken Vs Number of Hidden Nodes:



Accuracy vs Number of Hidden Nodes:



Inference behind choosing hidden nodes and hyper parametrization:

The number of hidden nodes is directly proportional to the prediction accuracy and inversely proportional to time. The increase in the number of hidden nodes neural network helped significantly increase the prediction accuracy from average of 67% at 4 hidden nodes to 94.7% at 50 hidden nodes. The time taken also increased from 56 seconds to 303 seconds due to the number of calculations involved. But the huge increase of 27% in accuracy makes the time taken worthwhile. Thus we get fixed the optimal number of hidden nodes at 50 where we achieved the highest prediction accuracy consistently.

Regularization is used to reduce over-fitting and increase classification accuracies. From our analysis after experimenting with different regularization values, non-regularized runs will often take huge number of iterations when compared with regularized runs. In all these non-regularized runs, the weight vector was growing by a large margin compared with regularized runs. We believe the non-regularized runs made it hard for the machine learning algorithm to properly learn the weight vectors and harder to converge on the local minima.

Over-fitting was a problem where there was a large gap between training accuracy and tested accuracy. With regards to tuning regularization parameter and number of hidden nodes, we empirically observed that the gap between testing and training accuracy was larger when hidden nodes were smaller and the gap reduced significantly for the same regularization parameter, when the hidden nodes were increased. Thus regularization parameter should be chosen empirically by testing with different parameters values and choose the one which gives maximum consistency in prediction result.