CSE574 INTRODUCTION TO MACHINE LEARNING

PROGRAMMING ASSIGNMENT - 1
HANDWRITTEN DIGITS CLASSIFICATION

REPORT
TEAM-13

Submitted By -

Alok Asok (alokasok@buffalo.edu)
Mithun Nagesh (mithunna@buffalo.edu)
Vibhav Gupta (vibhavgu@buffalo.edu)

1. OBJECTIVE :

The objective of the programming assignment is to implement a Neural Network for handwritten digit classification. We implemented the neural network (feedforward and backpropagation) and trained the neural network with the MNIST training data and incorporated regularization on the weights. The validation set was then used to tune the hyper-parameters ,λ and the number of units in the hidden layer.

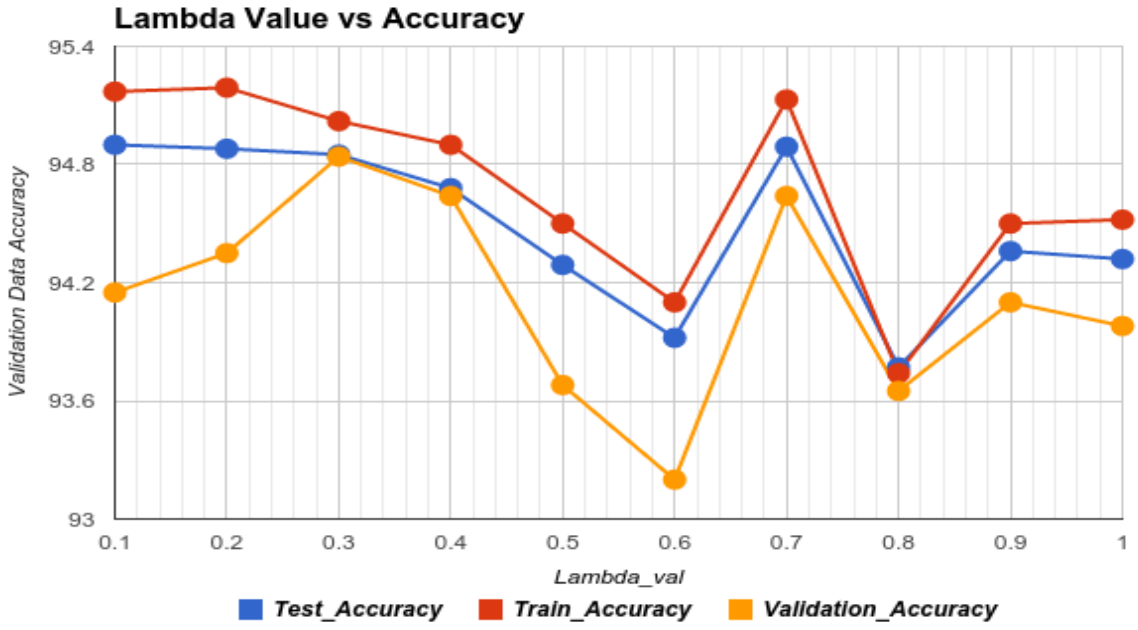
2. FEATURE SELECTION :

In the training data of the handwritten digits there are many features which does not provide any information about the variation between different data points. These features were ignored in the preprocessing step, taking the standard deviation across the data (columns). We considered 717 features for the training the neural network.

3. EXPERIMENTS :

In order to find the optimum value of the hyperparameters we ran the neural network with different values of the number of hidden units and the regularization parameter λ. The accuracy readings for the different set of values of these hyper parameters were chosen and based on the readings we arrived at optimum values for these parameters.

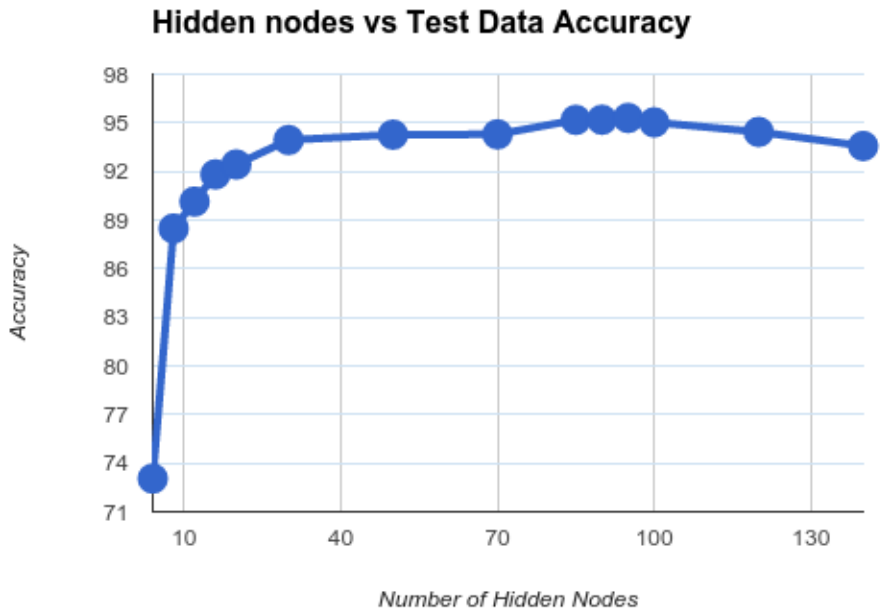
3.1. λ vs Prediction Accuracy



Plot : λ (Regularization hyper-parameter)VS Accuracy

The graph examines the change in the accuracy of the neural network in classifying the handwritten digits and the λvalue. The prediction accuracy of validation set keeps improving steadily as the λ value increases from 0 to 0.3. The highest prediction accuracy is obtained at λ value of 0.3 . With further increase in λthere is a gradual decline in the accuracy. This shows that the overfitting factor serves a good purpose between 0 and 0.3 i.e. 0<λ<0.3. Beyond 0.3 increase in λ doesn't improve accuracy so we are overfitting on the data.From 0 to 0.3 the accuracy is improving so λ<0.3 is underfitting on the given data.Since λ=0.3 is the optimum value which gives good accuracy for validation data set we can fix it as the best value for avoiding underfitting and overfitting problem.

3.2 Hidden Nodes (nn_hidden) Vs Prediction accuracy

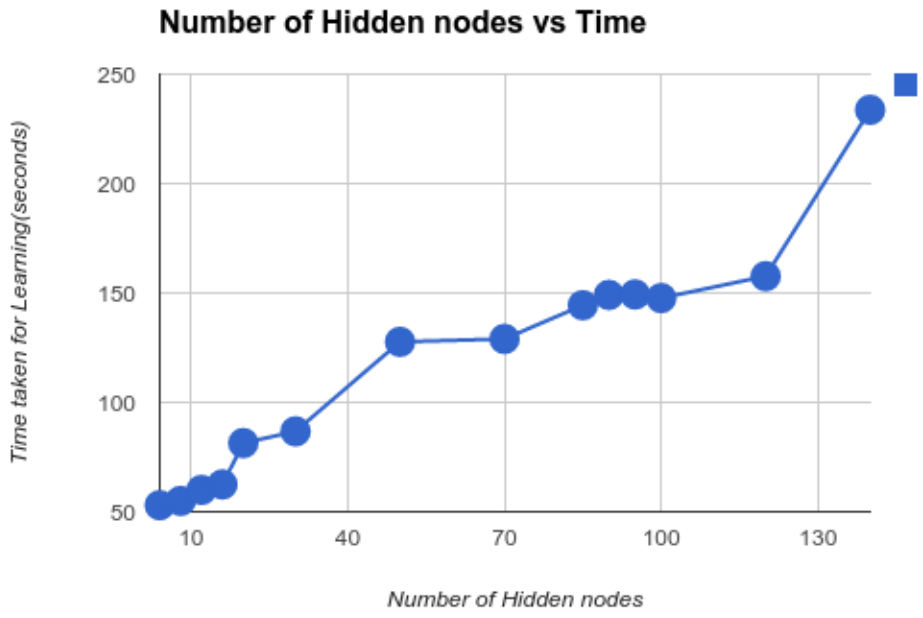


Plot : Number of Hidden Nodes VS Accuracy

Testing accuracy increases continuously as the number of hidden nodes are kept increasing till number of nodes is 95. This behaviour is found up to the point number of hidden nodes approach around 95, after which all the accuracies doesn't improve much and shows a small decline in accuracy as the hidden nodes are increased beyond 100. This shows that having a large number hidden nodes can cause overfitting. Also, we find that there is a steep rise in accuracy when n_hidden is increased from 10 to 20, which reflects that 10 hidden nodes are insufficient and

can cause underfitting and thus increases the training error. From n_hidden =20 to around 50, accuracy shows a gradual increase, that means these number of hidden nodes could serve the purpose but it is better to have higher number of those nodes.

Plot : Number of Hidden Nodes(X axis) VS Time taken for Learning(Y axis)



Analysis:

From the plot it is visible that as the number of hidden node increases the time taken also increases in a super linear way. Even though the prediction accuracy increases as the number of hidden nodes increases the learning rate decreases as the data size and the computations becomes huge.

4. Conclusion

From this assignment we learned that the accuracy of the prediction by the neural network depends on the features of the training example and the number of hidden nodes. We also found out that with an optimum values of the number of hidden nodes and the regularization parameter value λ we can achieve better results in less time.

REFERENCES :

LeCun, Yann; Corinna Cortes, Christopher J.C. Burges. “MNIST handwritten digit database”.

Bishop, Christopher M. “Pattern recognition and machine learning (information science and statistics).”