PROGRAMMING ASSIGNMENT 1 -GROUP NUMBER 43

COURSE NUMBER: CSE 574 – INTRODUCTION TO MACHINE LEARNING

TEAM MEMBERS: **INSTRUCTOR: Varun Chandola**

1) Arundhati Ravikumar Rao
2) Cuurie Athiyaman
3) Saumitra Sudhir Vaidya

NEURAL NETWORKS

Neural networks are multi layered perceptron models. It typically represents the working of a human brain and it's functioning. There are three layers in Neural Networks- input layer, hidden layer and output layer.

- The first layer comprises of (d + 1) units, where d represents the number of features of the input data with one additional bias node.
- The second layer in neural network consists of hidden units. Hidden units can be considered as the learned features extracted from the original data set. There is an additional bias node at the hidden layer as well.
- The third layer is also called the output layer where each node represents an output class.

The Generalization error is used as a measure to determine the accuracy of the model

RELATIONSHIP BETWEEN λ VALUES (REGULARIZATION TERM) AND ACCURACIES

A network, which is too closely fitted with the given data, will have a large variance and thus, also have large generalization error. This is called over fitting. But, if we attempt to minimize this variance, then bias becomes large, which again leads to large generalization error. This is called as under fitting.

The weights associated with the hidden layer units and the output layers units are directly proportional to the complexity of the network. If we increase the weights, the model becomes more complex (overfitting) and if we decrease the weights, the model becomes too simple (under fitting).

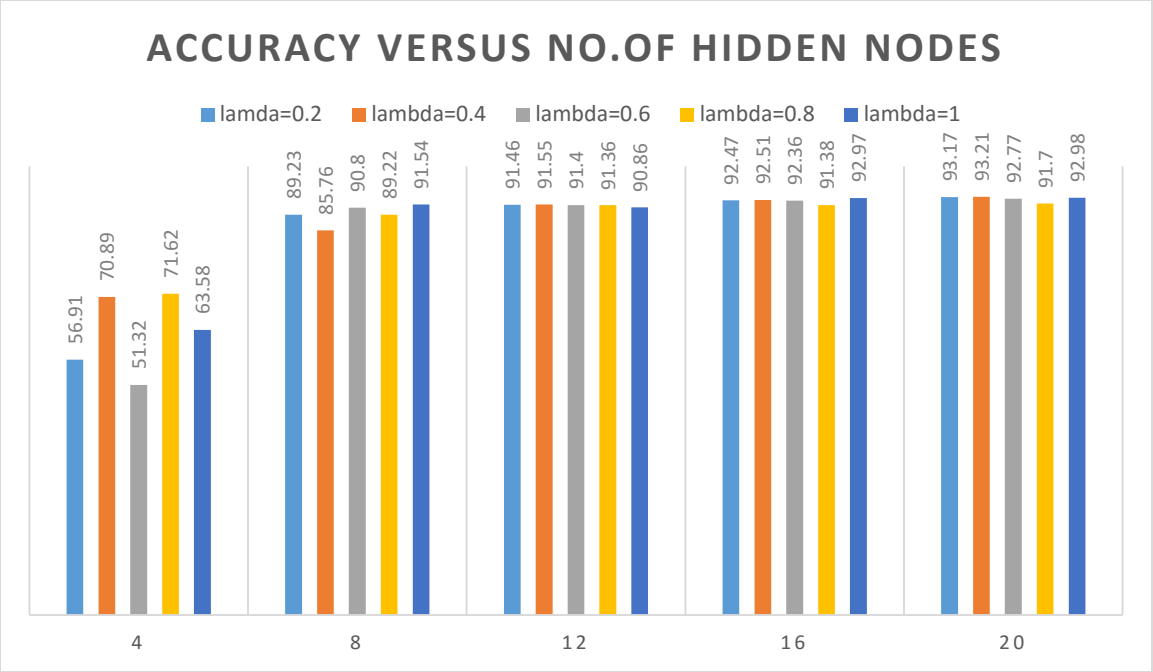
As a model becomes more and more complex, the error at first decreases and then increases. Hence, achieving a proper variance-bias tradeoff is crucial.

IMPLEMENTATION RESULTS

The regularization coefficient, λ, controls the emphasis put on the simplicity of the network, When we set λ=0, then it means the model only concentrates on minimizing the error. When λ is very high, then it means we put more emphasis on minimizing the complexity than error.

In our case, at λ = 0.6, it can be seen that irrespective of the number of hidden nodes, the value of accuracy is highest. Thus achieving a good bias-variance tradeoff.

Also, the overall accuracy with any value for λ is the best when we set number of hidden nodes = 20, as also seen in the chart below.



ANALYSIS OF RUNTIME:

During implementation, it can be observed that as the number of hidden nodes increased there was a corresponding increase in the runtime. This is because number of hidden nodes is directly related to the complexity of the Neural Network. The graph below illustrates this behavior.

CONCLUSION:

Thus, as can be observed from our analysis, the perfect combination of hyper parameters would be:

1) λ set to 0.6
2) Number of hidden nodes set to 20.