

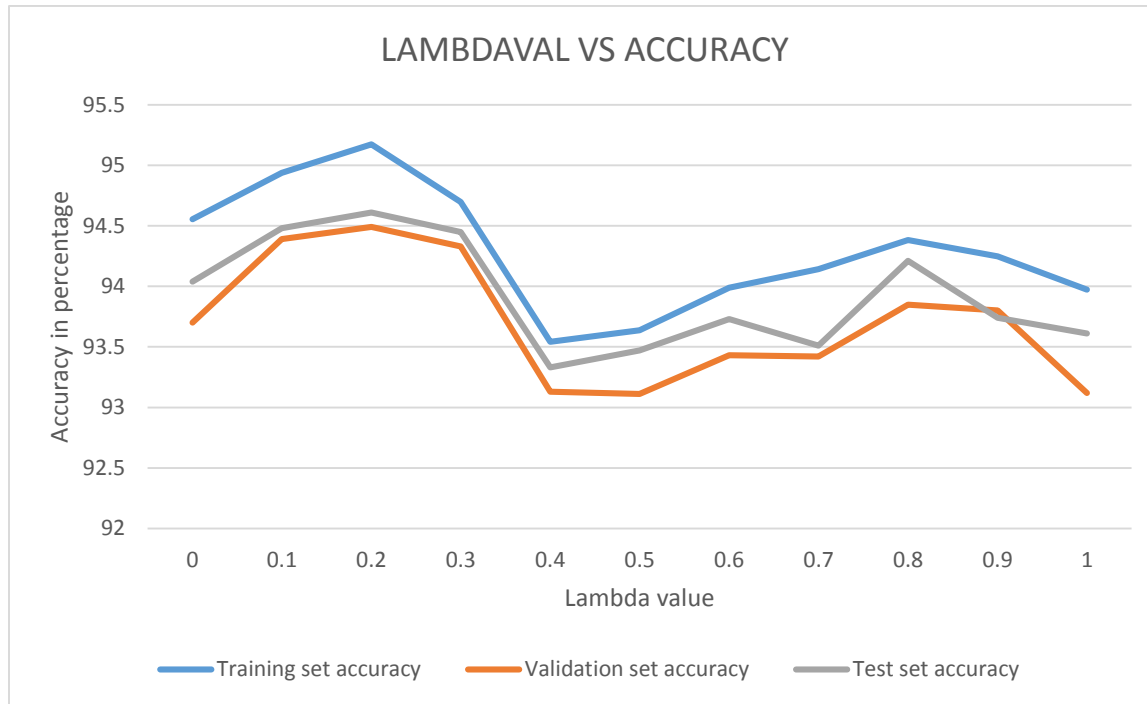
CSE 474/574
INTRODUCTION TO MACHINE LEARNING
PROGRAMMING ASSIGNMENT 1
REPORT

Submitted by

KARTHIK KAKARLA (50134820)
CHETAN PULAGAM (50133793)
PRATHYUSHA BOBBA (50134694)
(Team 57)

EXPERIMENTAL RESULTS:

- Lambda value vs Accuracy for test data:



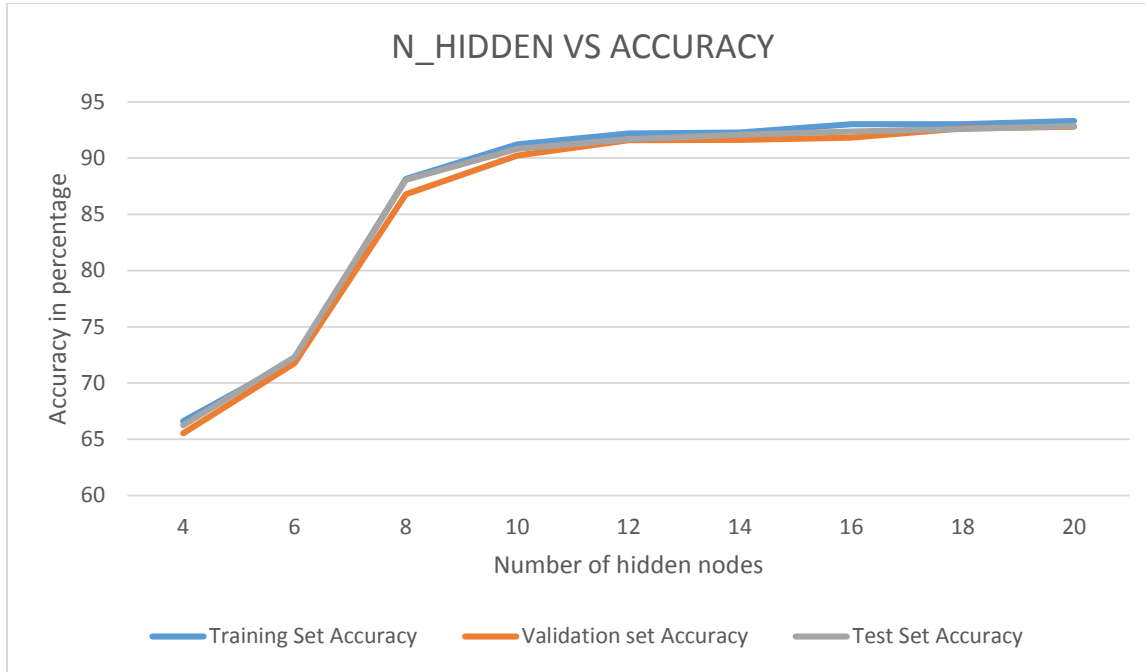
Lambda value	Training Set Accuracy	Validation Set Accuracy	Test Set Accuracy
0	94.554	93.7	94.04
0.1	94.938	94.39	94.48
0.2	95.174	94.49	94.61
0.3	94.696	94.33	94.45
0.4	93.542	93.13	93.33
0.5	93.638	93.11	93.47
0.6	93.988	93.43	93.73
0.7	94.142	93.42	93.51
0.8	94.384	93.85	94.21
0.9	94.248	93.8	93.74
1	93.972	93.12	93.61

TABLE (a): LAMBDA_VALUE vs ACCURACY when N_HIDDEN=50, MAXITER=50

INFERENCE:

As evident from the graph, the **optimal value of lambda is 0.2** in our experiments. For this particular value of lambda we see that the neural network does a better job. With higher values or lower values of lambda will result in lesser accuracies.

- Hidden nodes vs Accuracy for test data:



N_hidden	Training Set Accuracy	Validation Set Accuracy	Test Set Accuracy
4	66.616	65.53	66.23
6	72.034	71.75	72.3
8	88.156	86.77	88.06
10	91.214	90.22	90.82
12	92.194	91.6	91.67
14	92.284	91.63	92.06
16	93.022	91.82	92.35
18	93.02	92.64	92.59
20	93.302	92.78	92.83

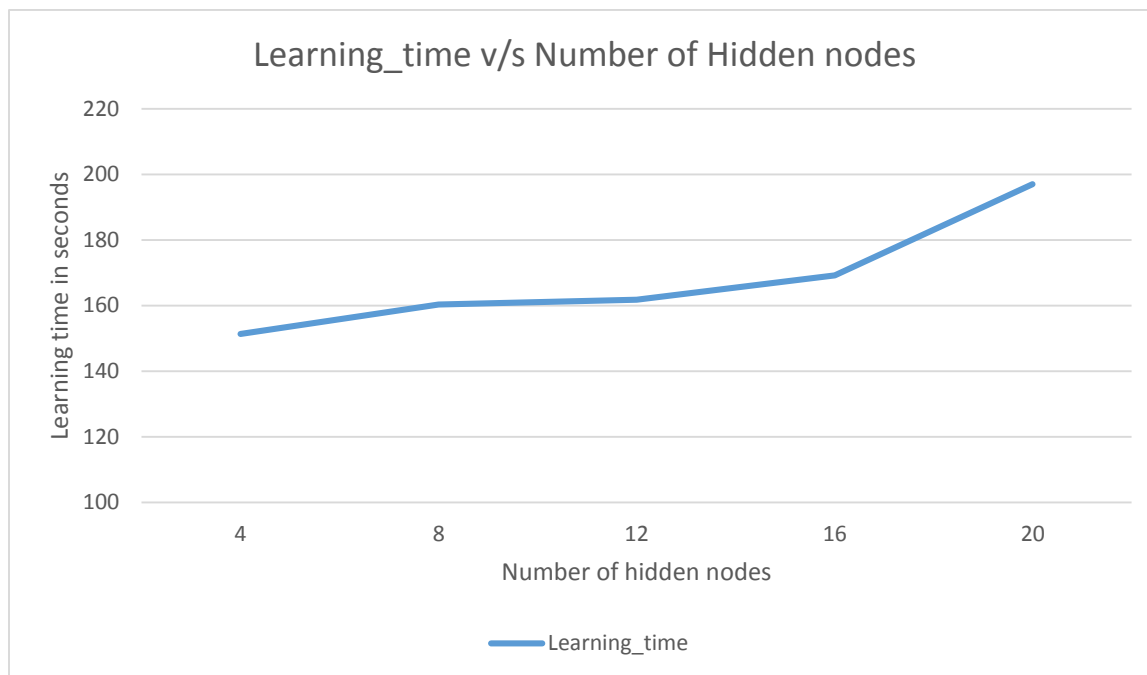
TABLE (b): N_HIDDEN vs ACCURACY when LAMBDA VALUE=0.4, MAXITER=50

INFERENCE:

As evident from the graph, the accuracy increases with the increase in the number of nodes. However, the rate of increase in the accuracy reduces once it reaches an optimal value. We

obtained an **optimal value of 16 hidden nodes** in our experiment. With very less number of hidden nodes, the accuracy decreases because the features learnt will be very less and hence the accuracy will be affected. After reaching an optimal value, the rate of increase in accuracy decreases because the neural network has already learnt the required features to accurately predict the test data. Further increase in the number of features will not increase the accuracy by significant amount. Also, having too many hidden nodes will increase the computational intensity and also it might lead to over-fitting some times

- Learning time vs Number of hidden nodes



Hidden nodes	Learning time(in secs)
4	151.366
8	160.361
12	161.807
16	169.184
20	197.019

INFERENCE:

As evident from the graph, the learning time increases with the increasing number of hidden nodes. This is because, as the number of hidden nodes increases, the complexity of the neural network increases and it becomes more computationally intensive. This increases the learning time required to learn the data.

HOW DO YOU CHOOSE THE HYPER-PARAMETERS OF A NEURAL NETWORK?

We choose the value of hyper-parameters λ and the number of hidden nodes after considering the factors like accuracy and run-time.

The hyper-parameter λ also known as the regularization coefficient is used to avoid the problem of over-fitting. Over-fitting results in learning the training data so perfectly that it fails to generalize and as a result will fail to predict the test data accurately. Thus we use a regularization parameter λ to avoid this problem. However, this value must not be too large. A very large value will not be able to solve this problem. As evident from the graph, the **optimal value of lambda is 0.2** in our experiments. For this particular value of lambda we see that the neural network does a better job. With higher values or lower values of lambda will result in lesser accuracies.

Increase in the number of hidden nodes will increase the complexity of the network and thus the accuracy will be increased. However, computational complexity also increases and hence the learning time for the algorithm increases. Thus it has to be chosen optimally such that the learning time is less and also the accuracy is as required.