Analyses of different parameters in a Multilayer perceptron model

Arnav Sundan
2019TT10954

1 Introduction

Implemented a model to recognize handwritten digits using neural networks built from scratch.

Dataset used: https://web.iitd.ac.in/~seshan/a1/handwritten_image_data.rar.

All the specifications given in the assignment like initialising of weights, regularisation, 3 layer model, gradient descent for backpropagation were followed.

For testing the data was split into a 20:80 ratio with 20% kept for testing.

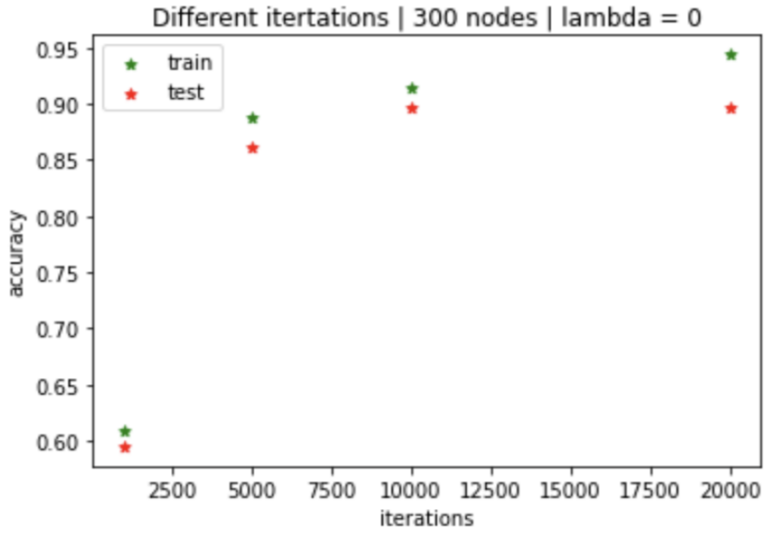
2 Results and observations

2.1 number of iterations

Number of nodes in the hidden layer was kept constant(300) and no regularisation was used.

Results of the experiment are given in the table and a graph is plotted using it.

No. of iterations	1000 iterations	5000 iterations	10,000 iterations	20,000 iterations
Training accuracy	60.98%	80.85%	91.38%	94.37%
Testing accuracy	59.50%	86.12%	89.71%	89.68%

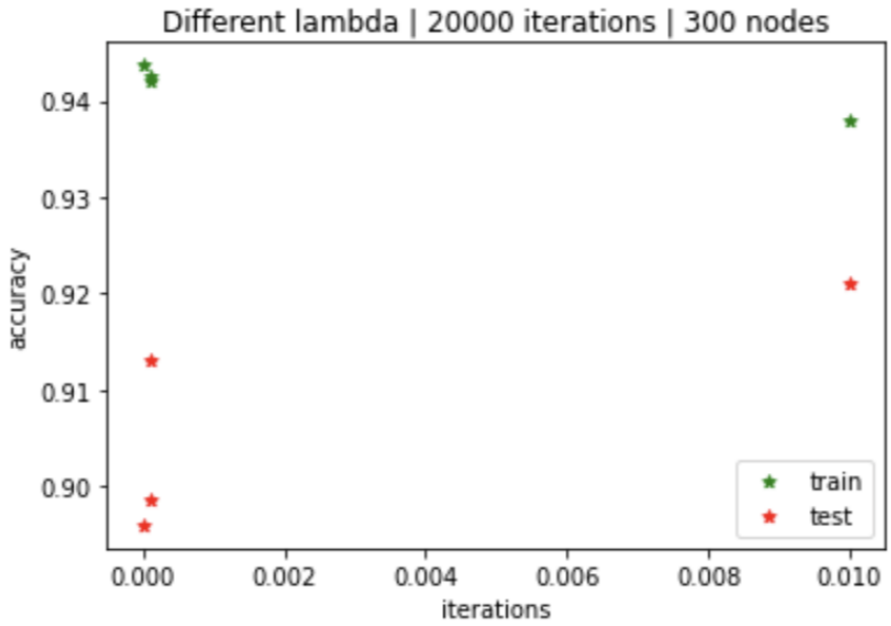


From the above results it is evident that increasing the number of iterations leads to increase of training accuracy but as we can see the testing accuracy for 20,000 iterations < testing accuracy for 10,000 iterations by a very small amount. This may be due to high accuracy of 20,000 that the model started overfitting to even the noise.

2.2 L2 - Regularisation(lambda)

Number of iterations was kept constant(20000) with 300 nodes in the hidden layer.

Lamda	0	0.0001	0.001	0.01
Training accuracy	94.37	94.25	94.22	93.82
Testing accuracy	89.68	89.87	91.31	92.14

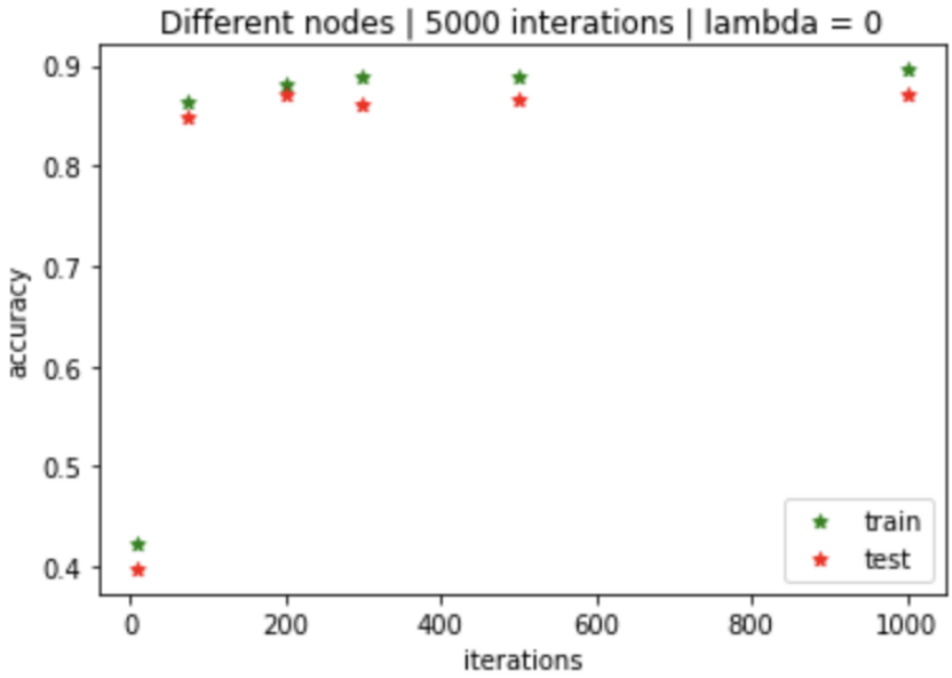


From the above results it is evident that increasing the lambda reduces the training accuracy and increases the testing accuracy in a given range. From extrapolating the graph we can see that both the training and testing accuracy may decrease if lamdba is increased beyond a limit.

2.3 number of nodes in the hidden layer

Number of iterations was kept constant(5000) and no regularisation was used.

No. of nodes	10	75	200	500	1000
Training acc.	42.37%	86.45%	88.22%	88.95%	89.61%
Testing acc.	39.85%	85.02%	87.15%	86.89%	87.11%



From the above results it is evident that increasing the number of nodes in the hidden layer leads to increase in both testing and training accuracy till a limit. After that training accuracy increases but testing accuracy decreases as the model starts overfitting.

3 Future-work

In the above experiment the activation function was kept constant(Sigmoid). Due to shortage of time I wasn't able to study the different activation functions like Softmax, Rectified Linear Unit, Leaky Rectified Linear Unit, Hyperbolic Tangent function.