Machine Leaning Assignment-3B

Lalith Satya Srinivas Kaladi
2017CS10340

April 2020

1 Neural Network Build

I have implemented Neural Network with back propagation algorithm to train the network and mini-batch stochastic gradient descent algorithm, Mean Squared Error as loss function having batch size, hidden layer sizes, learning rate, number of features, target classes as custom inputs. By default I have chosen number of attributes as 784, number of classes as 26, sigmoid is the default activation function.

2 Size of Hidden Layer Vs Accuracy

The plot below shows the accuracies obtained by training with different sizes of hidden layers. It can be observed from the plots that the accuracies of both test, training sets increases as we increase the size of hidden layers.

The stopping criterion for all the test cases is succesive change in error is less than 0.5.
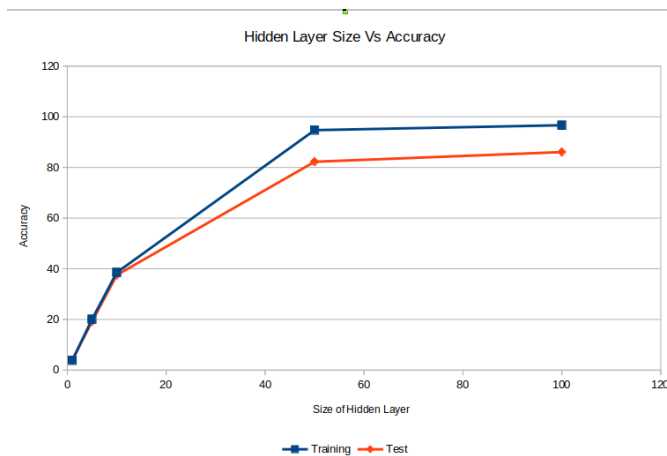
| Size	Training Acuuracy	Test Accuracy
1	3.86	3.91
5	20.06	19.14
10	38.54	37.52
50	94.72	82.25
100	96.67	86.06

3   Adaptive Learning Vs Constant Learning

Adaptive learning has been adopted and observed the accuracies, and time taken for execution. The stopping criteria is the difference between consecutive errors is less than 0.01.

Plot shown below shows the accuracies obtained with this model over test, training datasets. It can be observed from the plot that, as usual, the accuracies increase on increasing the layer size.
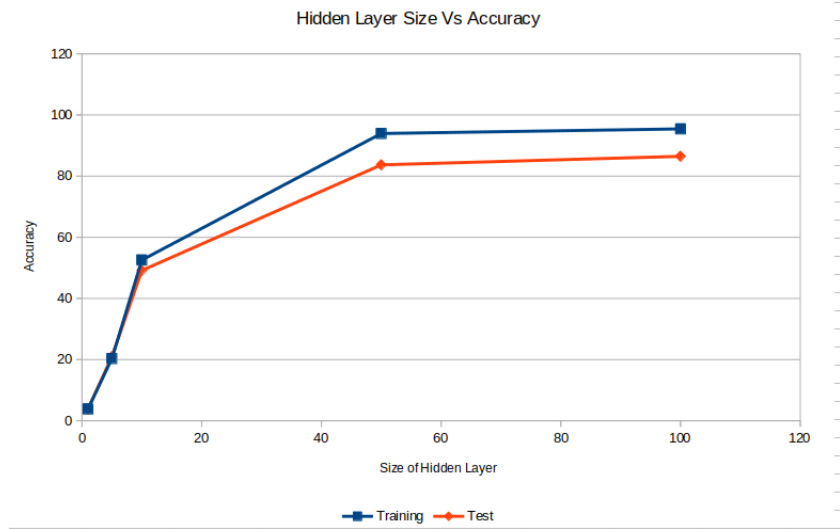
Size	Training Acuuracy	Test Accuracy
1	3.84	3.84
5	20.33	20.92
10	52.55	49.14
50	93.95	83.68
100	95.44	86.48

Plot shown below shows the execution times for adaptive learning, constant learning. It can be observed that adaptive learning has speeded up the process so much (almost 2.5 fold).
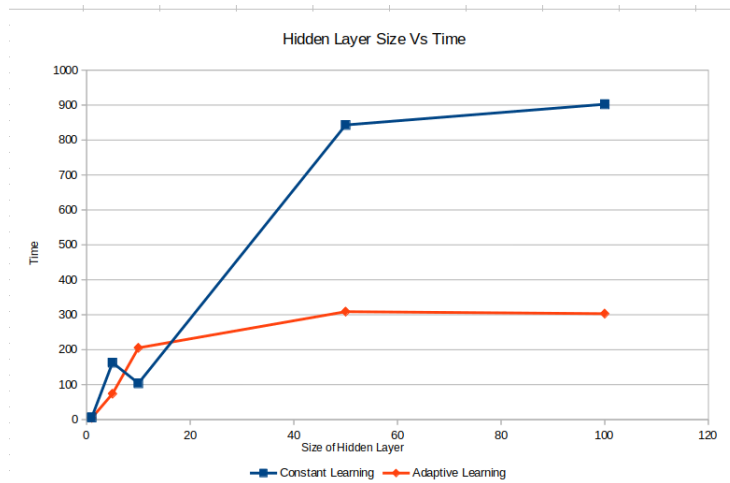
Size	Adaptive Learning	Const. Learning
1	6.17	4.80
5	162.95	73.99
10	103.90	205.31
50	843.14	308.84
100	902.45	303.11

Table 1: Execution times with different learning techniques

4   ReLU Vs Sigmoid

The activation function is now changed to ReLU and observed the accuracy obtained with two hidden layers of size 100 each.
The (96.21, 86.63) are the accuracies obtained with sigmoid as activation function on training, test datasets respectively where as (93.06, 83.53) are the accuracies observed with ReLU. It can be observed that sigmoid is producing better

results than ReLU. But, an obvious reason would be the difference in stopping criteria and time it takes to learn. But as of the results observed sigmoid is producing better result when compared to ReLU.

5 MLPClassifier

I have used MLPClassifier library from sklearn for learning the network, with hidden layer architecture same as above section and ReLU as activation function for hidden layers. The result observed was (87.62, 82.53) on training, test datasets respectively. The results from the above section are better than that from the library.