Neural Networks and Deep Learning

CSCI 5922 – Assignment 2 (Fall 2017) by Akshit Arora (108631342)

Part 1:

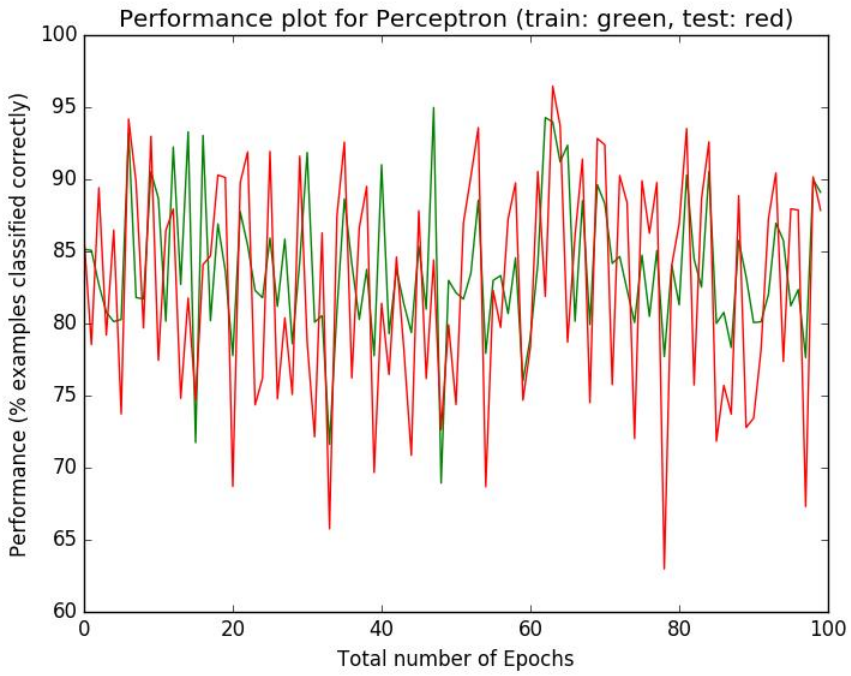
Question 1a

One way to go about this problem is to do it with a fixed learning rate. Because we know that in perceptron, learning rate only affects the scale of the weights, therefore, we can take it as 1.0. Here is the performance (% correctly classified examples) I get after 100 epochs and weights initialized to 0.0

Final Training Set performance: **89.1072**

Final Test Set performance: **87.8589**

Here is the performance over test (in red) and training (in green) datasets that I get for fixed learning rate over 100 epochs:

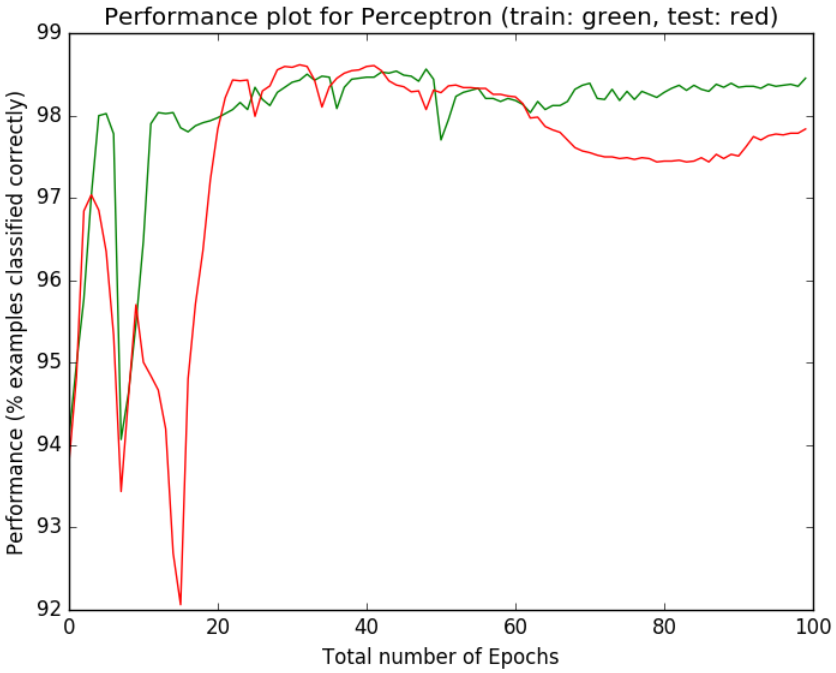


Since this is a very noisy dataset, the accuracy jitters a lot. In order to smooth it, I used time-decay learning rate. Here are the results I got after using a time-decayed learning rate:

Final Training Set performance: **98.4526**

Final Test Set performance: **97.8363**

Here is the performance over test (in red) and training (in green) datasets that I get for time-decayed learning rate over 100 epochs:



Part 2: (Assume 20 epochs, 5 hidden units, 0.1 learning rate; unless specified otherwise)

Question 2a

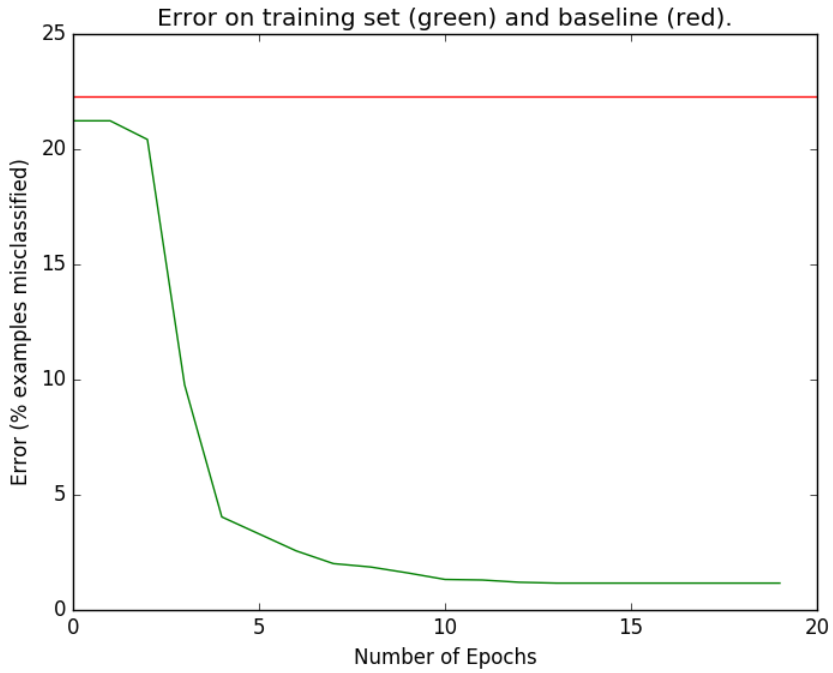
Error function used: squared error. Although, it is better to use cross-entropy for classification problems. I tried cross entropy too and there wasn't much difference for this particular data-set.

Question 2b

Baseline: 0.0. Since its frequency is: 6414 in a training set with: 8143 examples. Here is the baseline error (percentage) after 20 epochs when the network is trained in a mini-batch setting of batch size 100: **22.2768**

Question 2c

With 5 hidden units, 20 epochs and 0.1 learning rate, the training error does not drop any further after roughly 15 epochs. Here is the error plot:



Question 2d

Learning rate used in Q2c) is **0.1**

Could have gone for time decayed learning rate too.

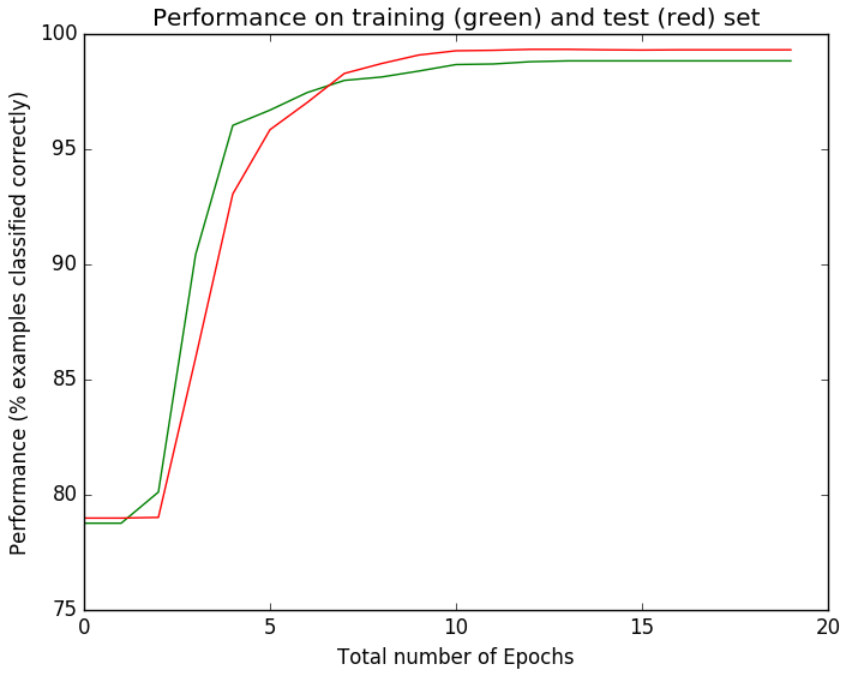
Question 2e

Performance has been measured in terms of percentage of examples classified correctly.

Training set performance: **98.8333**

Test set performance: **99.3335**

Here is the plot:



Question 2f

To plot the training and test set performance as a function of number of hidden units. I would keep the number of epochs to 20 and learning rate: 0.1

In the plot below:

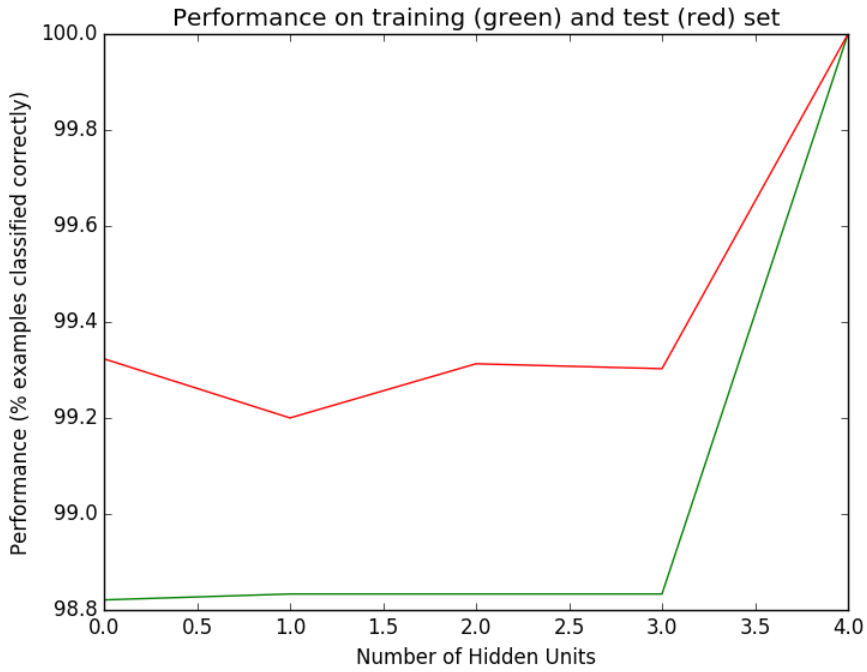
number of hidden units = 0.0 means 1 hidden unit

number of hidden units = 1.0 means 2 hidden units

number of hidden units = 2.0 means 5 hidden units

number of hidden units = 3.0 means 10 hidden units

number of hidden units = 4.0 means 20 hidden units


Performance on training (green) and test (red) set
Performance (% examples classified correctly)
Number of Hidden Units
100.0
99.8
99.6
99.4
99.2
99.0
98.8
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
4.0