MACHINE LEARNING
Assignment 3

Surabhi S Nath
2016271

QUESTION 1 - Train Neural Network from scratch

a) Implementation using Sigmoid

1 hidden layer - 100 units

- Initialized using random weights drawn out from a normal distribution with:
    - Mean = 0
    - Standard deviation = 0.001
- Alpha = 0.01
- Batch gradient descent (weights and biases are updated only after a batch of images)
- Batch size = 20 images
- Number of epochs = 100

3 hidden layers - 100, 50, 50 units

- Initialized using random weights drawn out from a normal distribution with:
    - Mean = 0
    - Standard deviation = 0.1
- Alpha = 0.01
- Batch gradient descent (weights and biases are updated only after a batch of images)
- Batch size = 20 images
- Number of epochs = 100


b)

Since the testing error is very high, the model could not have overfitted. If the model would have overfitted, it could not have generalized to new unseen test samples. If the model were underfitting, the training and testing accuracies would have been very low, however in this case, they are both very high hence we can conclude that the model is not overfitting or underfitting and is a good fit on the data.

1 hidden layer - 100 units
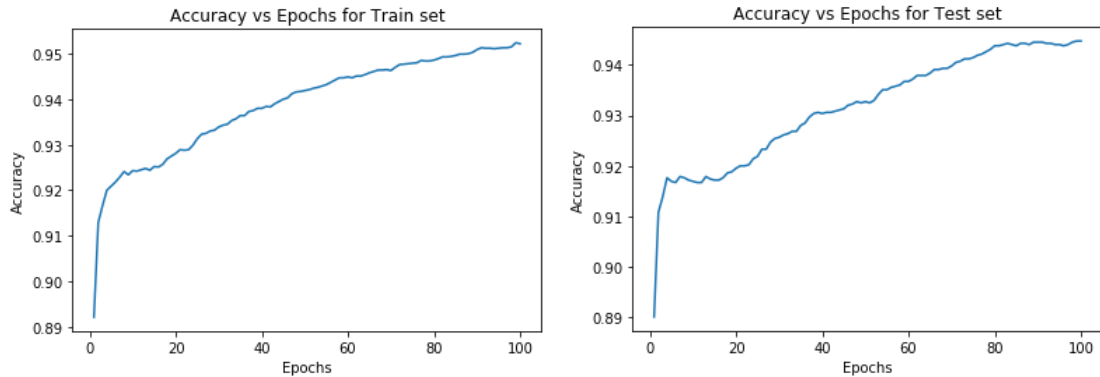
3 hidden layers - 50, 50, 100 units
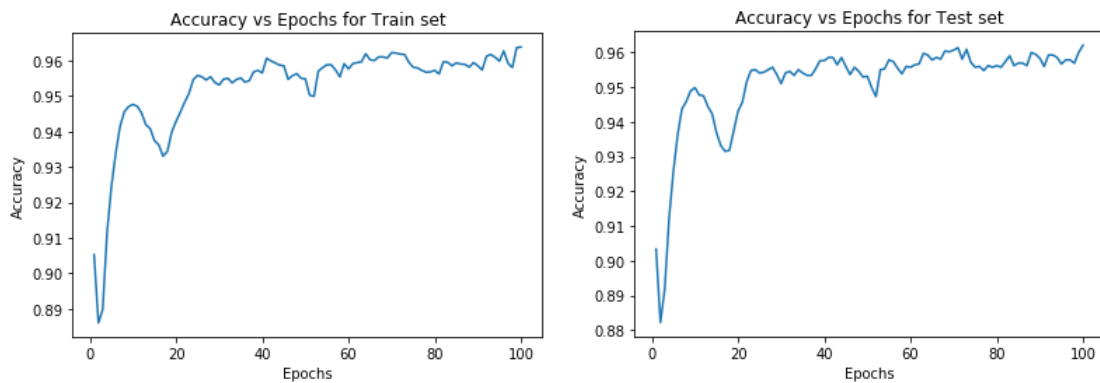
The accuracies for 3 hidden layers is higher than 1 hidden layer because 3 hidden layer neural network can learn better through more complicated decision boundaries.

It is also observed that the 3 hidden layer neural network reaches the stable value in almost as less as 20 iterations after which it keeps fluctuating very little increase. On the other hand, the single hidden layer model accuracy keeps increasing gradually with each epoch.

Some challenges faced during neural network coding include dimension matching, multiplying correct shape of vectors, accuracies not increasing, nan and 0 values in Relu and tuning of hyperparameters. To counter these, batch gradient descent was performed, leaky Relu was used, code was run over multiple epochs and values of the parameters - alpha, and normal distributions (using which weights were initialized randomly) parameters mean and standard deviation were varied to check the best accuracies for model.

----

c) Implementation using Relu

The parameters used were the same as in case of Sigmoid activation, except number of epochs was reduced to 20 because accuracy had reached its optimal value within 20 epochs and then remained there and kept fluctuating.
It was observed that Relu activation worked better for for the data and resulted in higher accuracies. This could be due to 2 main disadvantages faced by the sigmoid activation - vanishing gradient and high density - which are overcome by using Relu.
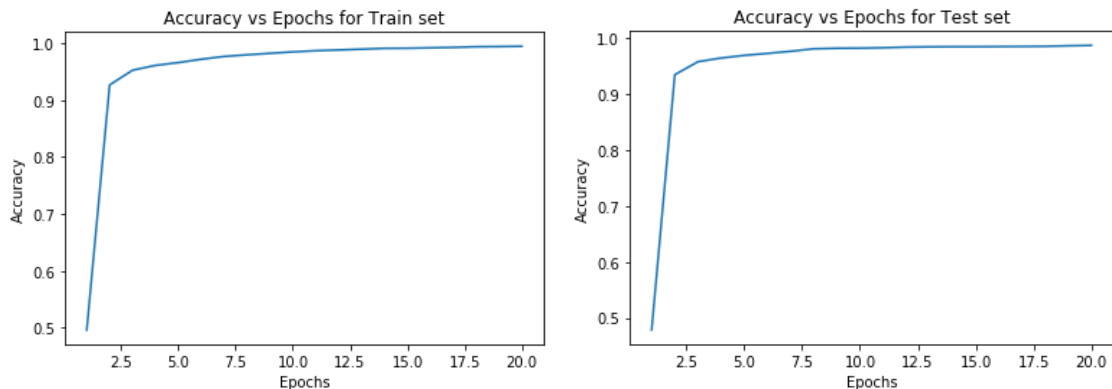
1 hidden layer - 100 units
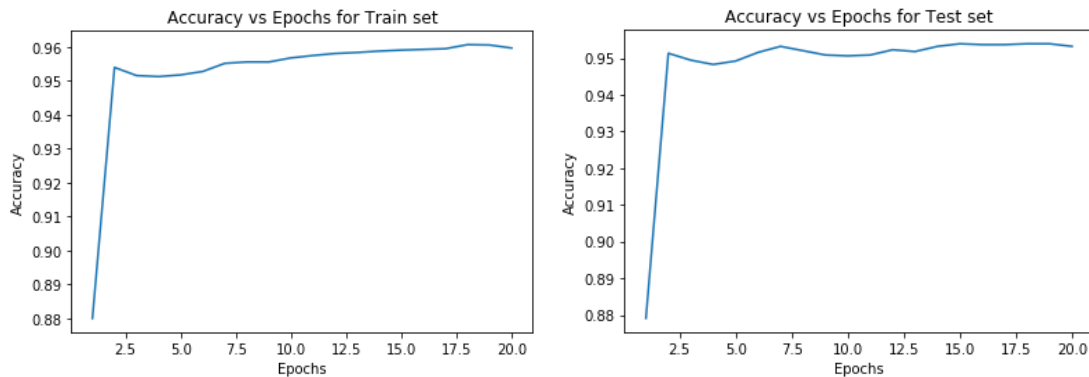
3 hidden layers - 100, 50, 50 units

Alpha value = 0.001
Standard deviation = 0.01
Leaky Relu was used. Very high accuracies were obtained.

----

d) Using sklearn

The parameters provided the the MLPClassifier were matched with the values provided in part a) for our own neural network.

1 hidden layer - 100 units

Accuracy = 0.9856

3 hidden layers - 100, 50, 50 units

Accuracy = 0.9814

The 2 accuracies for 1 hidden layer and 3 hidden layers are nearly the same. The accuracy here is only slightly greater than the one achieved using the self coded neural network. This could be because of the optimal weight initialization that the inbuilt neural network performs.

QUESTION 2 - CIFAR 10 Alexnet SVM

- Accuracy = 0.9165
- Confusion matrix =

Actual\Predicted	0	1
0	900	67
1	100	933

- AUC score = 0.9739
- ROC curve