Shreya Roy
Credit
SR: 14996
Dept CDS

A. Number of hidden layers are 1 and hidden neurons in the layer was 100.
Case1:
With initial learning rate .0001 after 1000 iterations 59% accuracy was achieved
Case2:
With initial learning rate .001 after 500 iterations 93% accuracy was achieved.
Case3
With initial learning rate .01 after 100 iterations 91.6% accuracy was achieved
Case4:
With initial learning rate .01 after 200 iterations training loss decreased but validation accuracy was decreased. This is the overfitting scenario.

Also,we can see that with lesser learning rate the model is underfitting (Case1)

Finally,
I ran 500 iterations with initial learning rate= .001, reached 81%

Shreya Roy
Credit
SR: 14996
Dept CDS

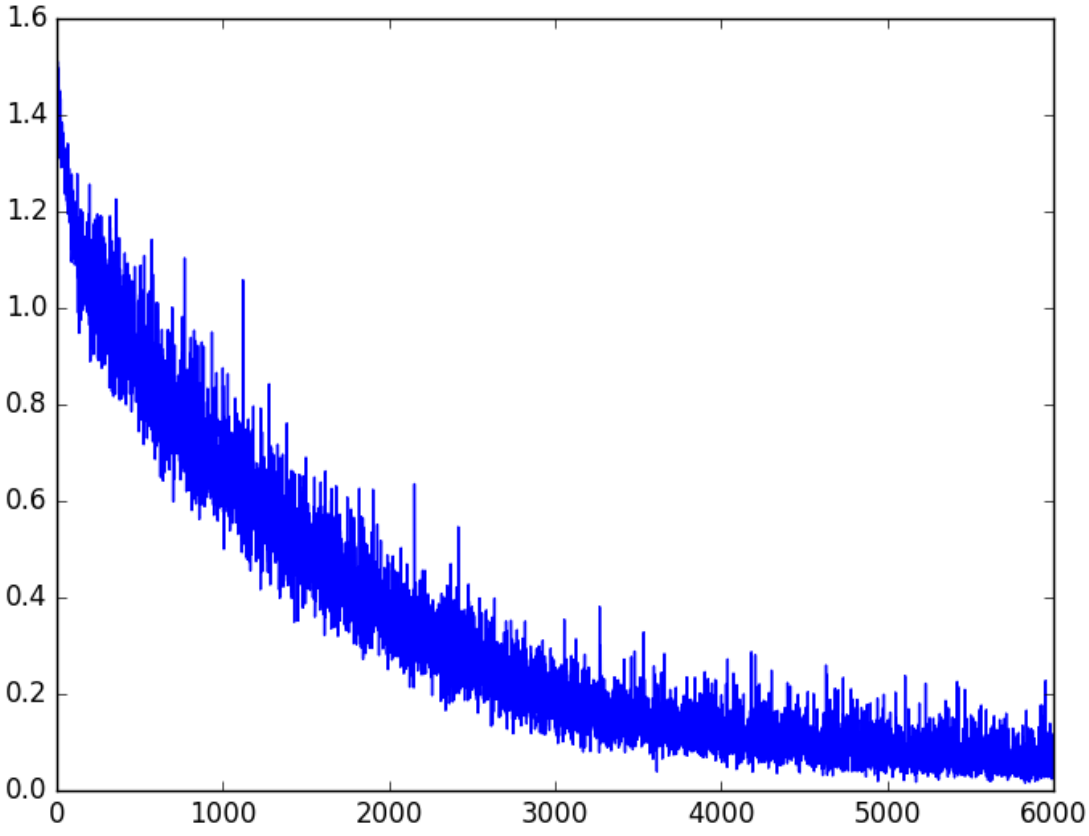

Fig1: Train Loss in the First 500 iterations.

Shreya Roy
Credit
SR: 14996
Dept CDS

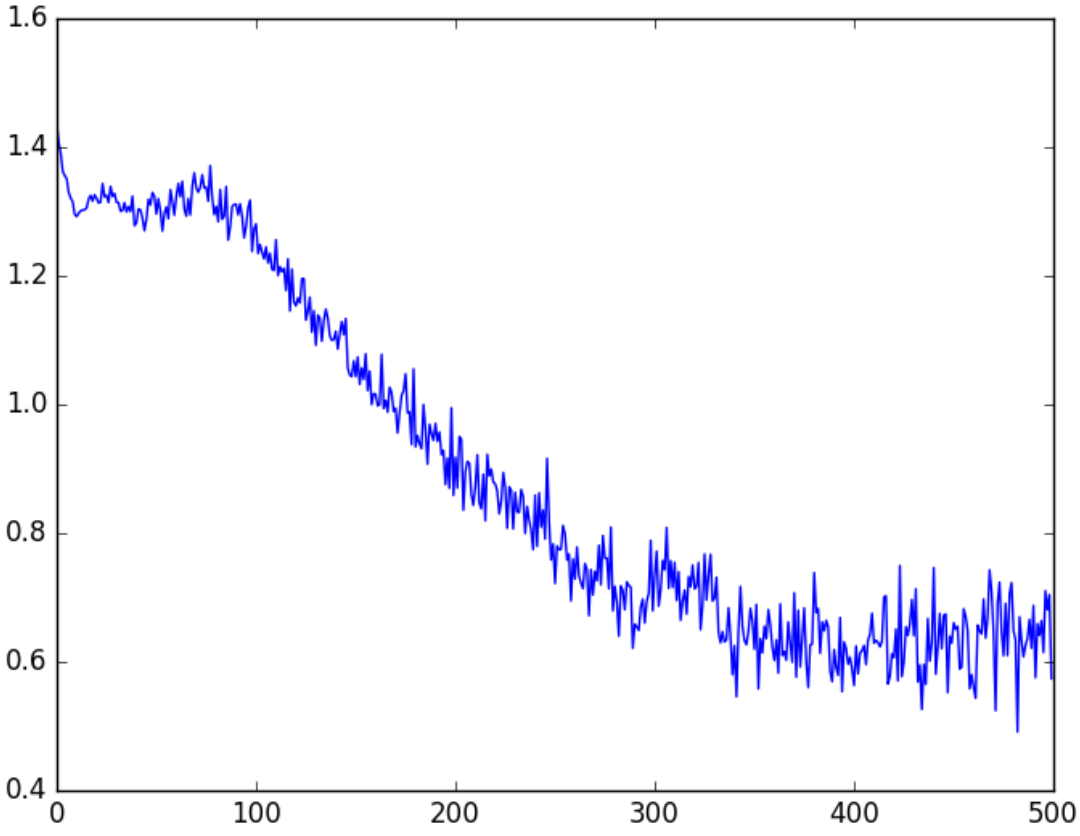

Fig1: Validation Loss in the Ist 500 iterations

Shreya Roy
Credit
SR: 14996
Dept CDS

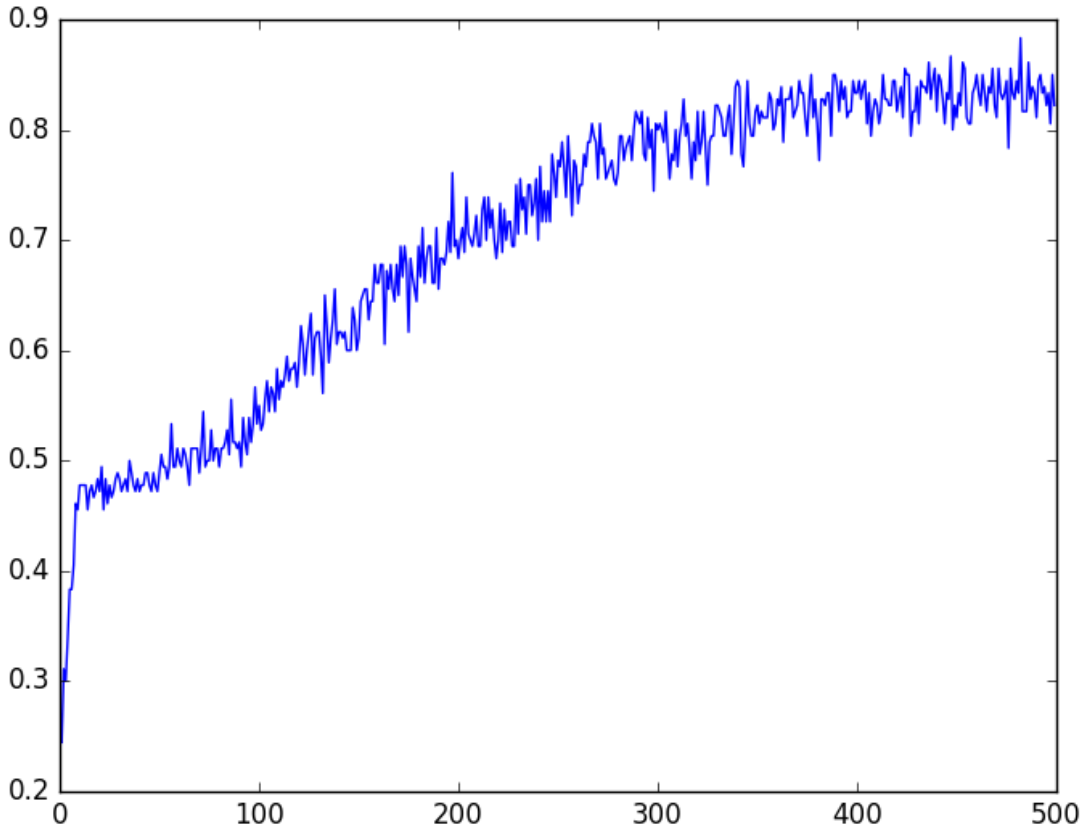

Fig3: Validation accuracy in the first 500 iterations

Again, I ran 50 iterations with initial learning rate= .0001, which led to 97.1 % accuracy
Although, in the test dataset it gave 89% accuracy.

Shreya Roy
Credit
SR: 14996
Dept CDS

Throughout, I used Adam optimizer with weight decay=1e-4 in the above 4 cases.

Case5:
With 10 neurons in the hidden layer I achieved only 61% accuracy in the first 500 iteration with same optimizer as used with the model of 100 neurons in the hidden layer (Case1)

Case6:
With 50 neurons in the hidden layer I achieved only 73.1% accuracy in the first 500 iteration with same optimizer as used with the model of 100 neurons in the hidden layer (Case1)

So, more number of neurons in the hidden layer is helping to converge the model faster.

Although,
I used only single hidden layer for all the above cases as according to the **universal approximation theorem** a feed-forward network with a single hidden layer containing a finite number of neurons can approximate continuous functions on compact subsets of $\mathbf{R}^n$, under mild assumptions on the activation function.

Shreya Roy
Credit
SR: 14996
Dept CDS

Experiments with varying hidden layer:
Now, I used 2 hidden layers each of which containing 100 neurons. But, in the performance I did not notice much differences. With the initial learning rate of 0.05 and using adam optimizerwith weight deacy=1e-4 after 300 iteration 90 % accuracy was reached on validation set and on the sample test data the model yield 86% accuracy.

Next, I used 3 hidden layers each of which containing 100 neurons per layer. After 100 iteration with initial learning rate of .05 and weight decay set as 1e-4 with adam optimizer 92% accuracy was achieved on validation set. Next I ran another 100 iterations with initial learning rate= .0005 and using the same optimizer the validation accuracy reached 99 % and the accuracy of 98% on the test dataset was achieved.