Shreya Roy
Credit
SR: 14996
Dept CDS

I solved the problem as a multi class classification problem where I considered 4 classes:
1. label 0 is given when the number itself is returned for neither being divisible by three nor by five.In the train_label.txt I named this class as 'num'.
2. label 1 when 'fizz'
3. label 2 when 'buzz'
4. label 3 when 'fizzbuzz'

Various hyperparameters (number of hidden layers , nodes per layer, learning rates)which I experimented with:

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)

Shreya Roy
Credit
SR: 14996
Dept CDS

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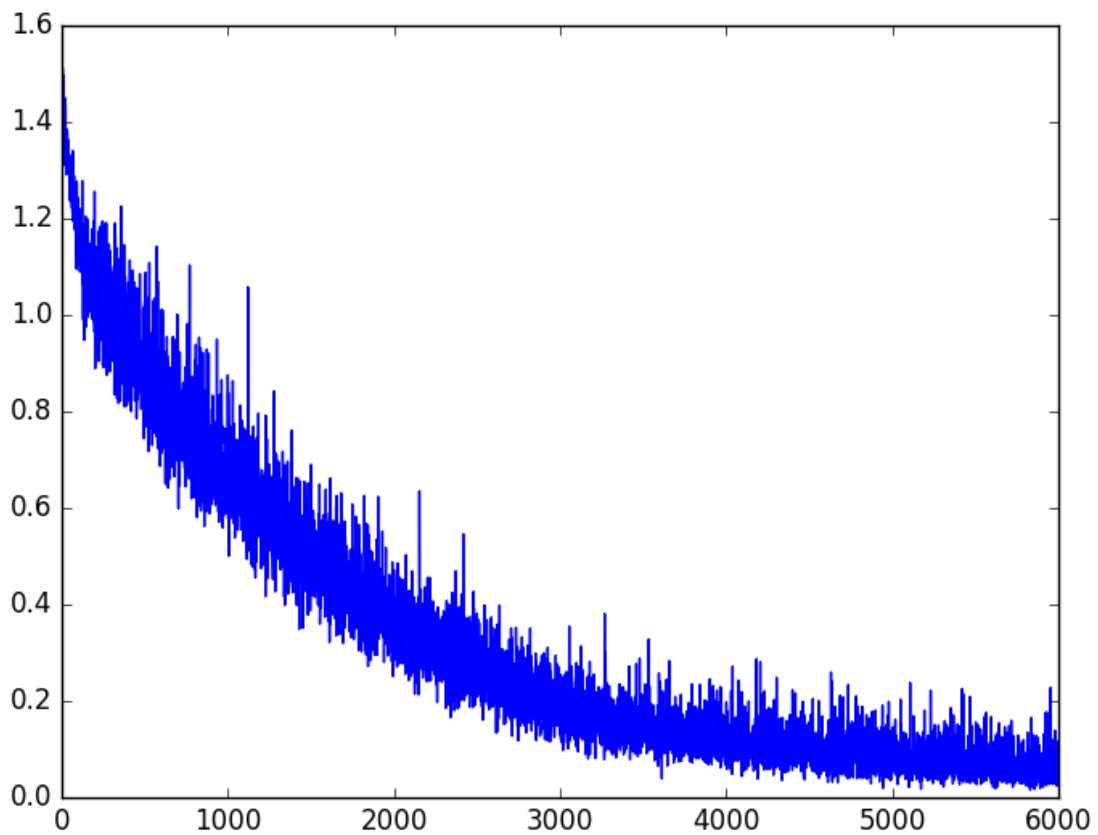I ran 500 iterations with initial learning rate=
0.001, reached 81%

Fig1: Train Loss in the First 500 iterations.
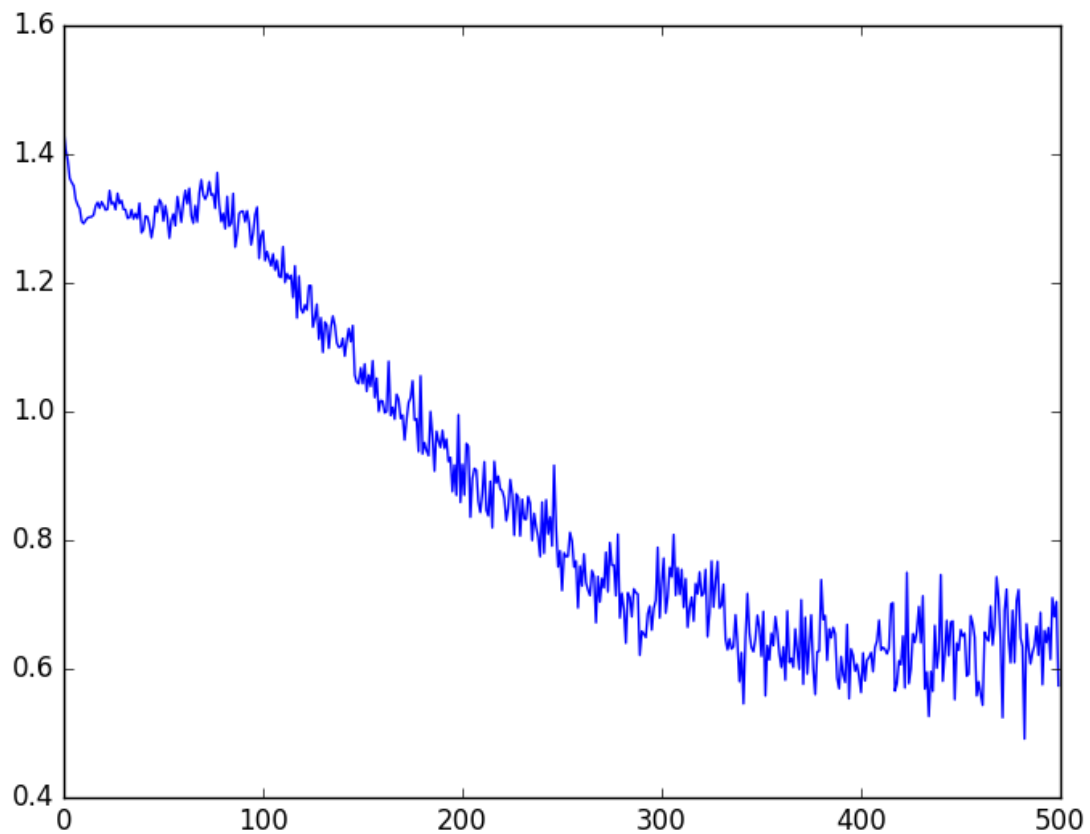
Shreya Roy
Credit
SR: 14996
Dept CDS

Fig1: Validation Loss in the Ist 500 iterations
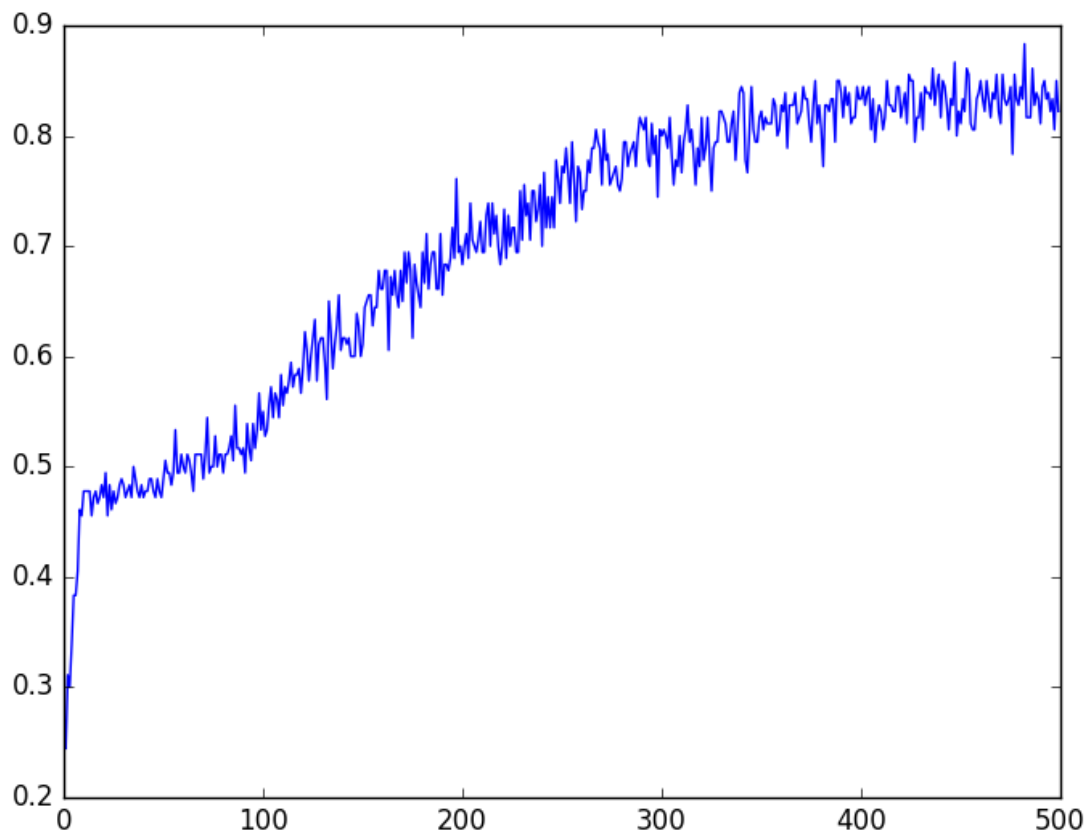
Shreya Roy
Credit
SR: 14996
Dept CDS

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.

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

Shreya Roy
Credit
SR: 14996
Dept CDS

optimizer as used with the model of 100 neurons in the hidden layer (Case1)

B.Experiments with two hidden layers:

Now, I used 2 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= .005 and using the same optimizer the validation accuracy reached 99 % and the accuracy of 98% was achieved on the test dataset.