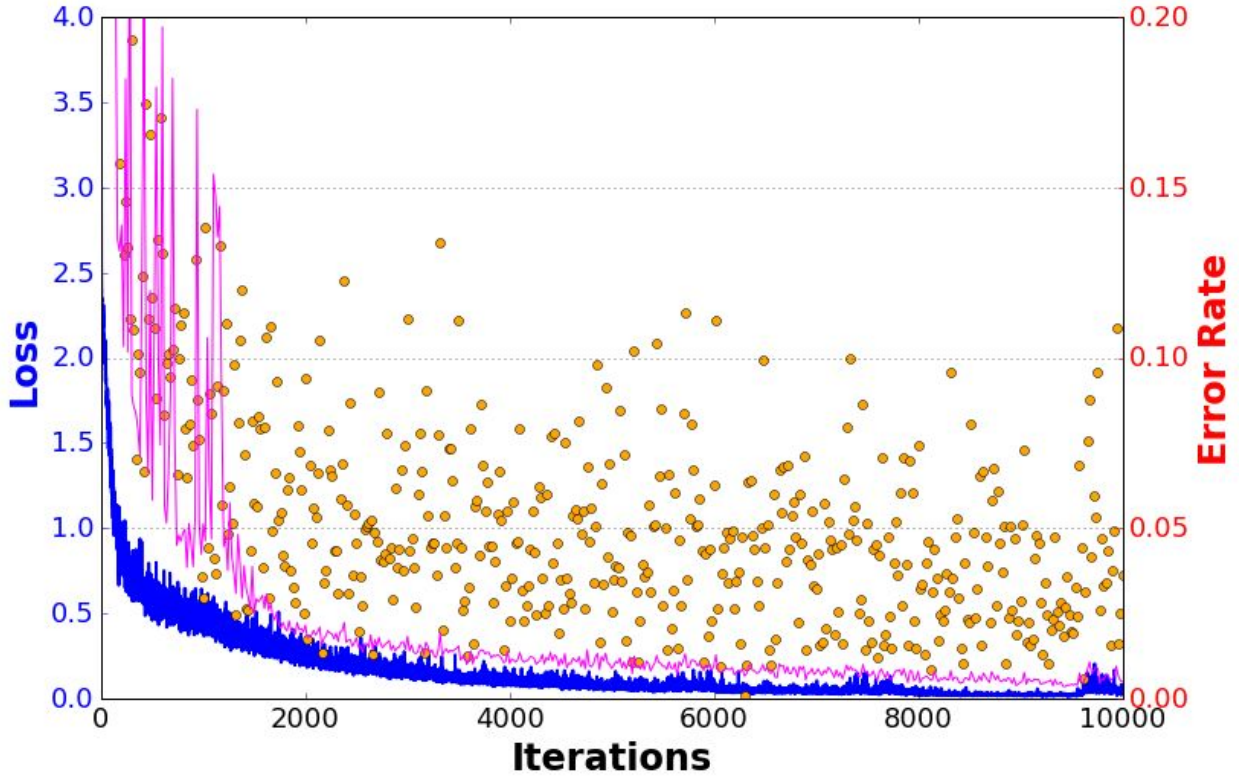
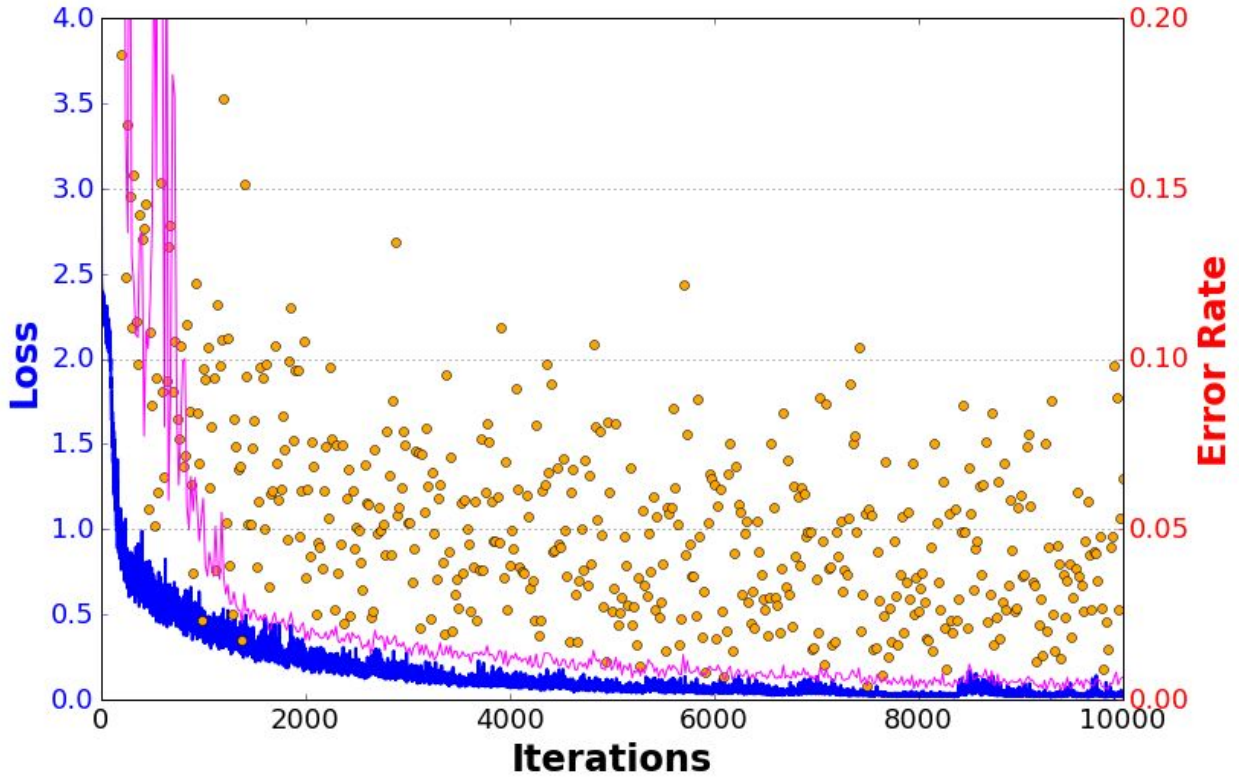
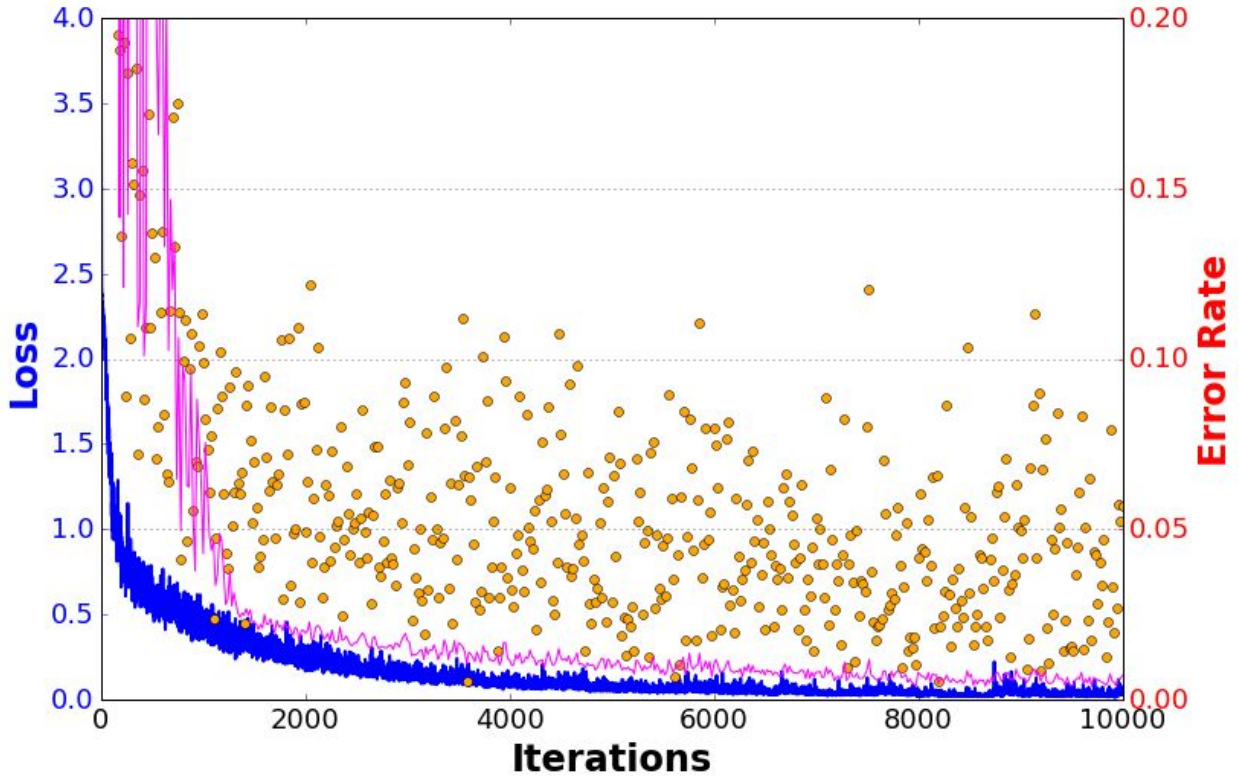
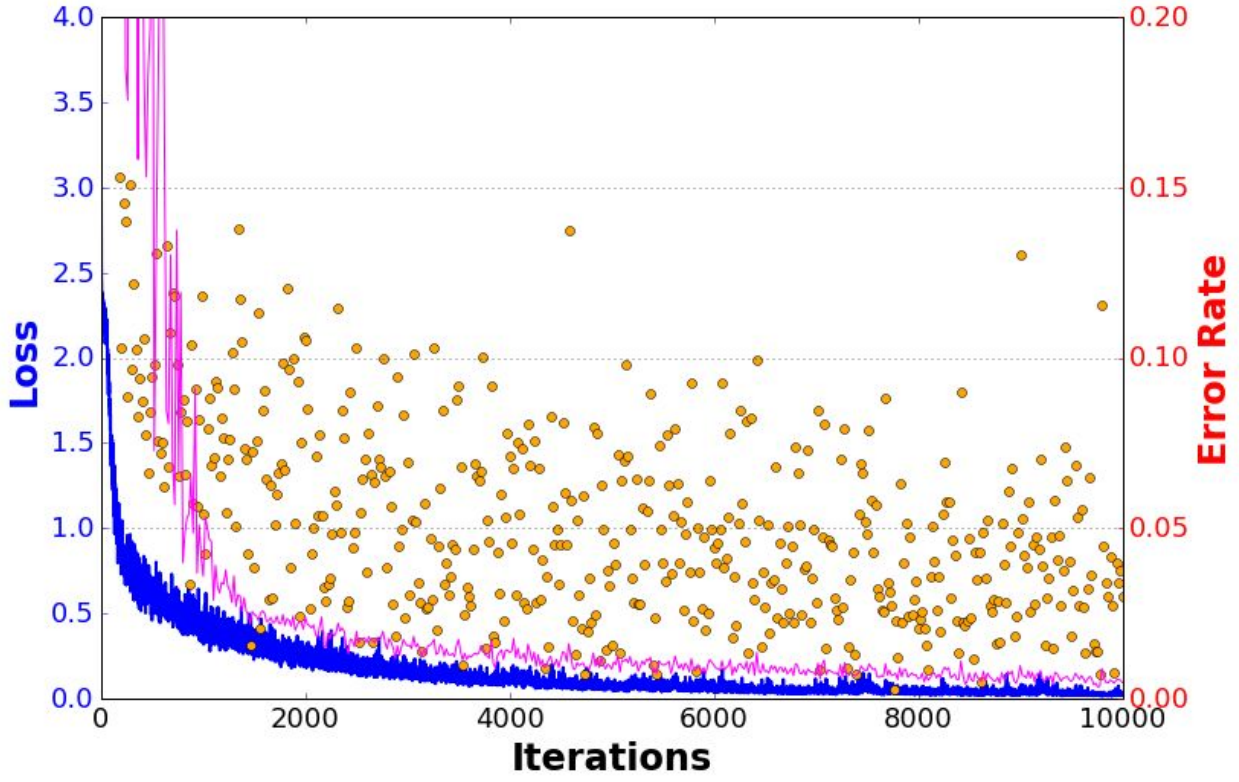
With BatchNorm:

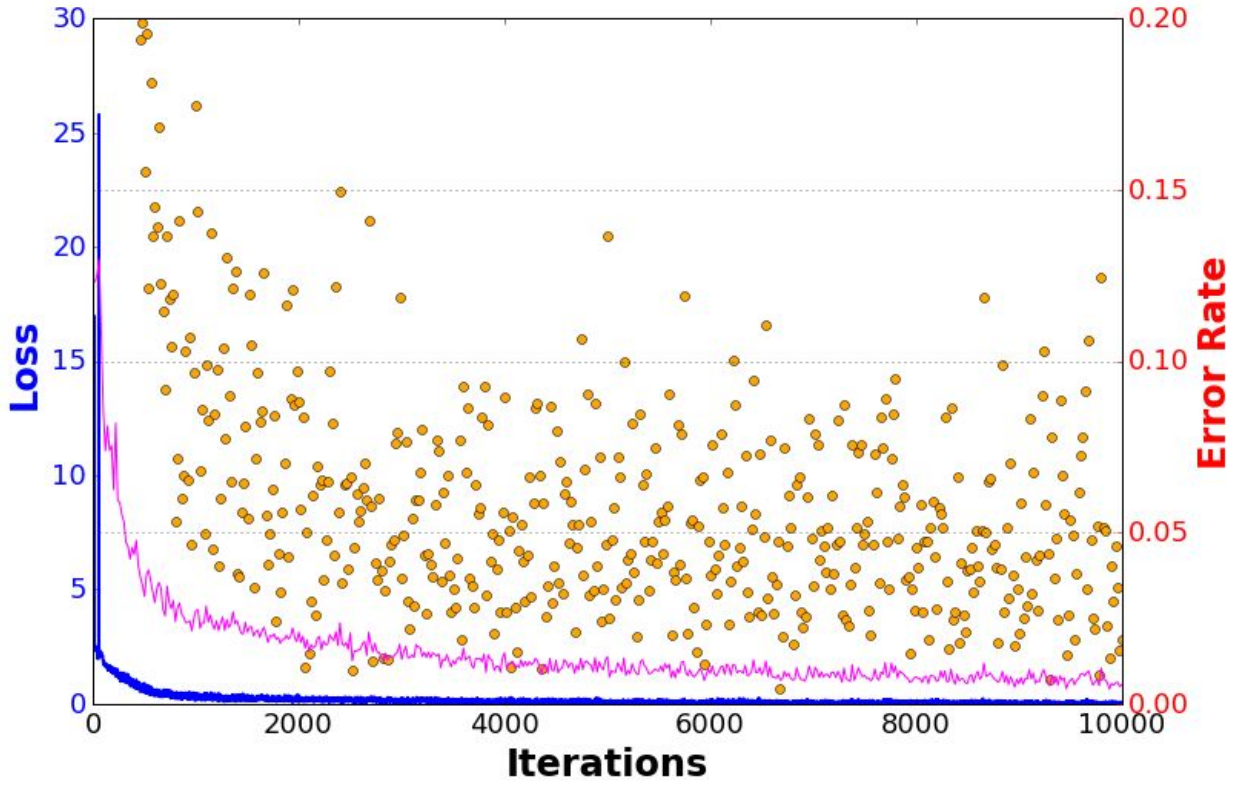


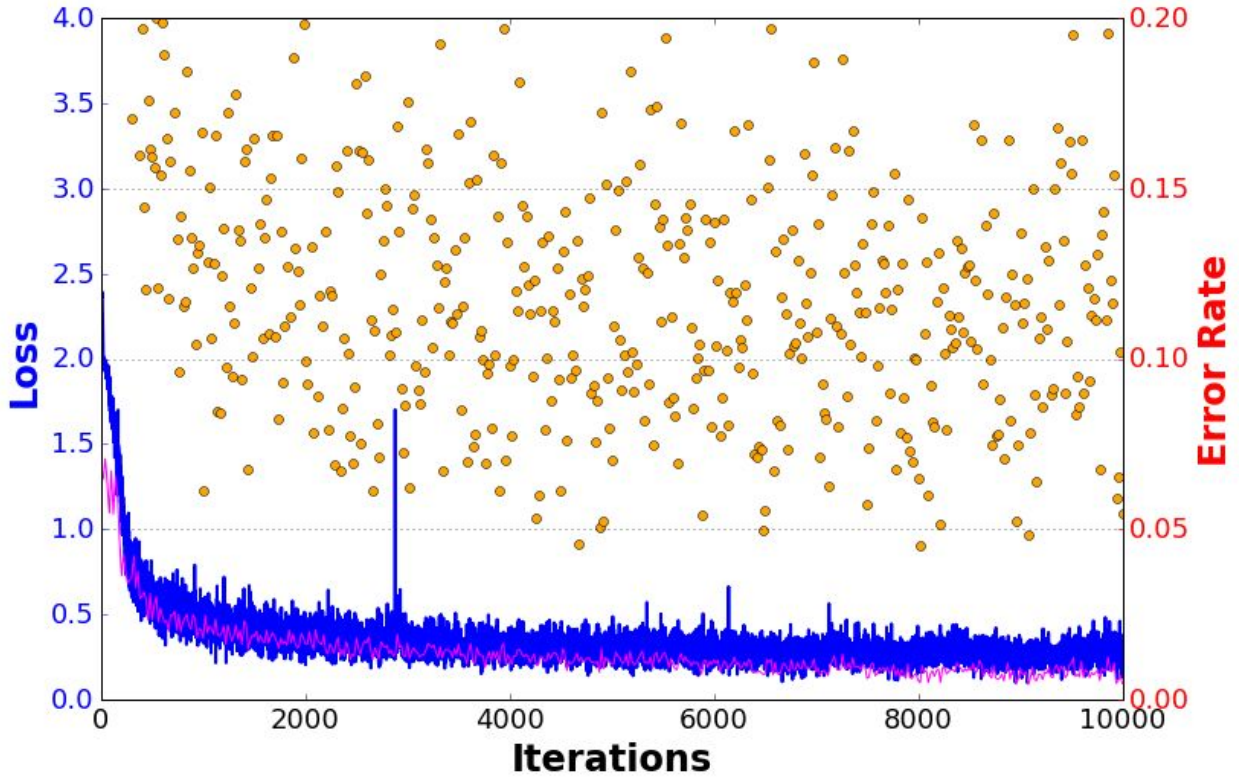



Loss
Error Rate
Iterations
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
4.0
0.00
0.05
0.10
0.15
0.20
0
2000
4000
6000
8000
10000

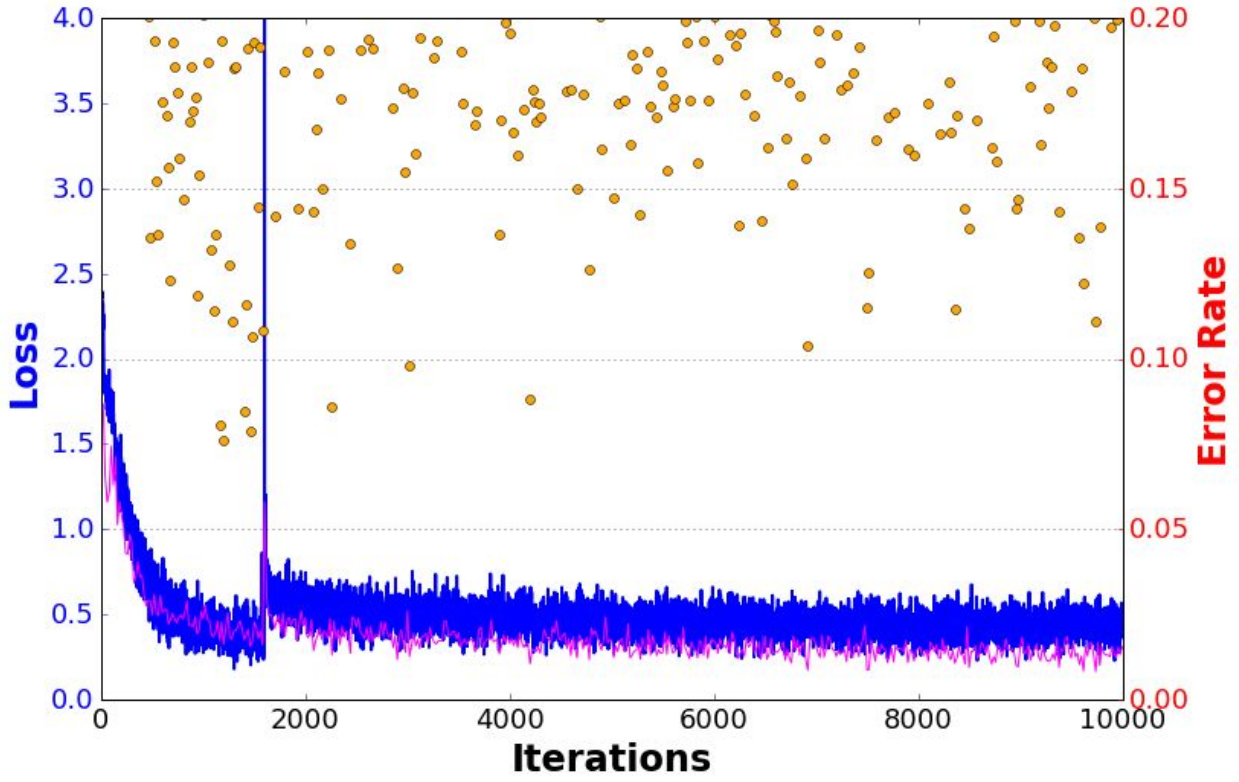

Loss
Error Rate
Iterations
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
4.0
0.00
0.05
0.10
0.15
0.20
0
2000
4000
6000
8000
10000

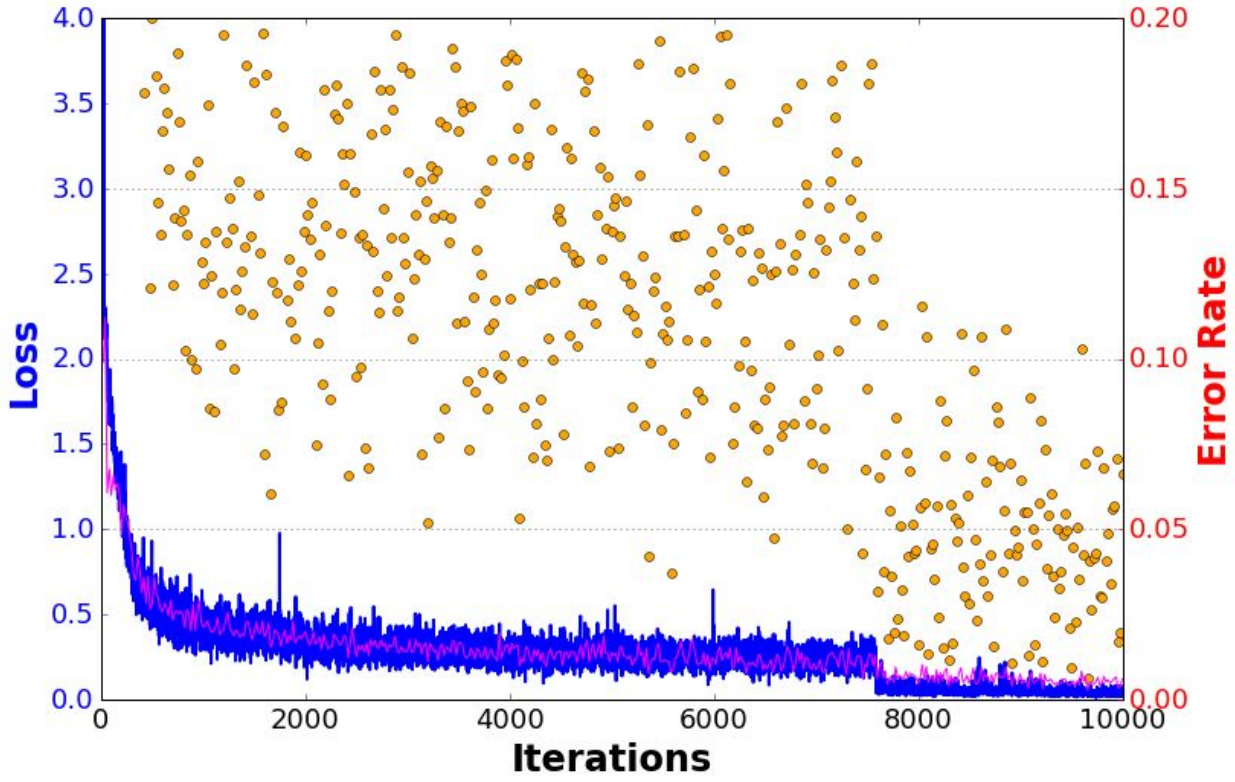
Without BatchNorm:
(note this example below's y-axis is not on the same scale as the BatchNorm trials, will fix in later diagrams)





On the example below, the spike is probably due to no batch norm and something like exploding gradients?




Loss
Error Rate
Iterations
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
0.20
0.15
0.10
0.05
0.00
0
2000
4000
6000
8000
10000