

DS 222 Assignment-2

Abhinav Anand, SR - 13954

1 Part a

Number of epochs - 20

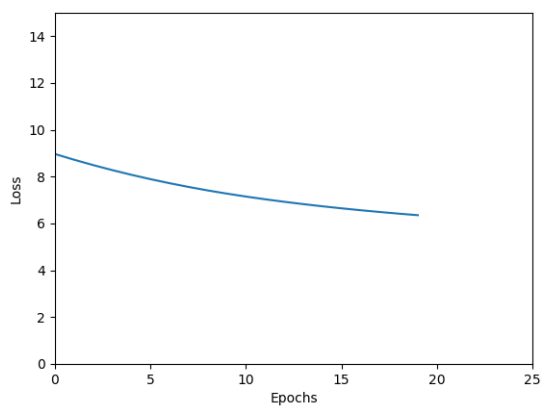
1.1 Decreasing Learning Rate

Initial Learning Rate = 0.3

Regularization = 0.003

Feature Vector Size - 5000

- Decreasing Factor : 0.95
- Train Time : 524.92 secs
- Test Time : 1.53 secs
- Train Accuracy : 56.56 %
- Test Accuracy : 51.82 %



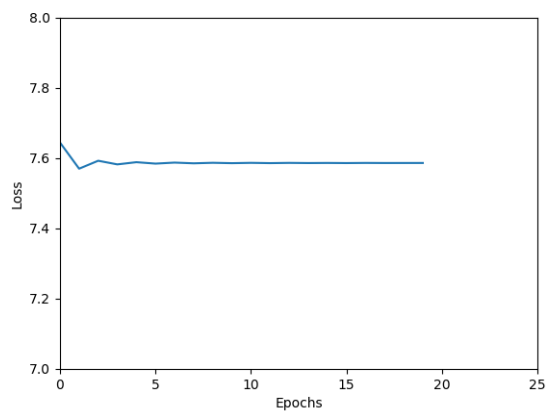
1.2 Constant Learning Rate

Initial Learning Rate = 0.3

Regularization = 0.001

Feature Vector Size - 250000

- Train Time : 715.033 secs
- Test Time : 1.928 secs
- Train Accuracy : 57.9 %
- Test Accuracy : 52.1 %



Increased accuracy is due to increase in number of feature vectors.

1.3 Increasing Learning Rate

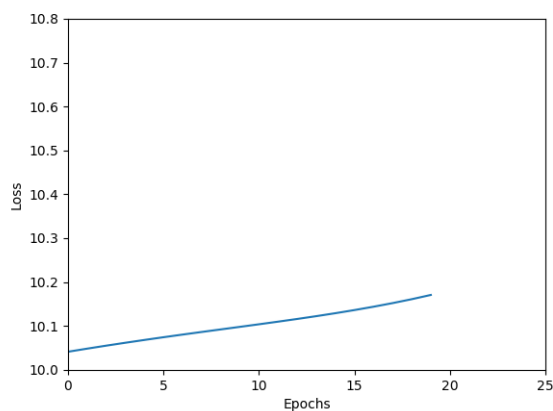
Initial Learning Rate = 0.1

Regularization = 0.01

Feature Vector Size - 5000

- Increasing Factor : 1.05
- Train Time : 484.40 secs
- Test Time : 1.34 secs
- Train Accuracy : 40.1 %

- Test Accuracy : 38.4 %



2 Part b

2.1 Parameter Server choice

I used JBossen parameter server framework. Hadoop does not support parameter server framework and parameter server integrated with Spark is not stable.

2.2 BSP Settings

Initial Learning Rate = 0.1

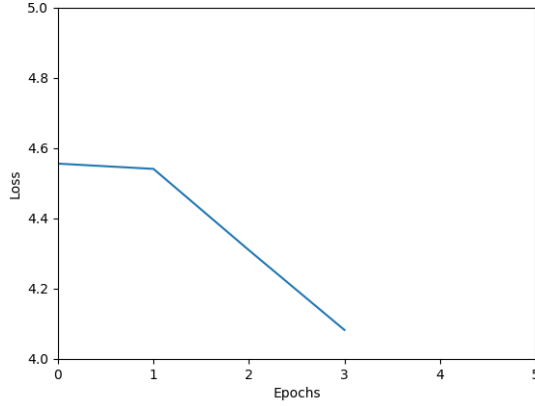
Regularization = 0.01

staleness-parameter = 5

Number of workers = 6 (3 host and 2 threads per client)

Epochs = 4 per worker

- Decreasing Factor : 0.95
- Test Time : 1.4 secs
- Train Accuracy : 47.34 %
- Test Accuracy : 42.69 %



2.3 SSP Setting

The program failed to terminate when staleness was set to 0. I was not able to find the root cause of why `PSTableGroup.globalBarrier()` failed to terminate.

2.4 Bounded Asynchronous Settings

Initial Learning Rate = 0.1

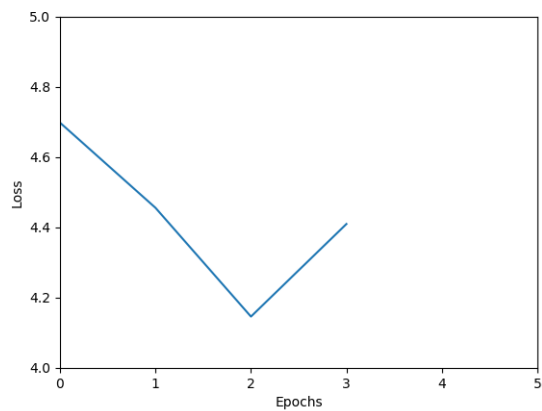
Regularization = 0.01

staleness-parameter = 10000

Number of workers = 6 (3 host and 2 threads per client)

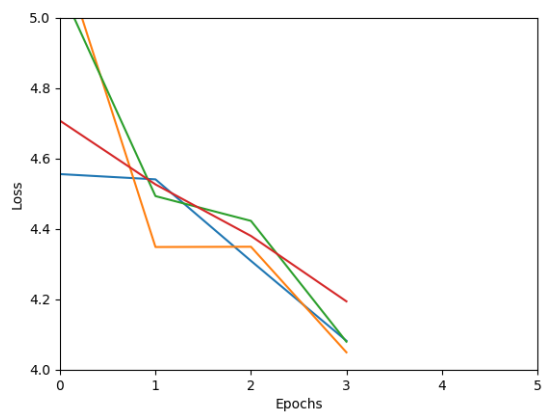
Epochs = 4 per worker

- Decreasing Factor : 0.95
- Test Time : 1.4 secs
- Train Accuracy : 39.56 %
- Test Accuracy : 34.27 %



2.5 Plot SSP for different delay parameters

Blue - Staleness = 4
 Orange - Staleness = 10
 Green - Staleness = 15
 Red - Staleness = 20



2.6 Plot Asynchronous for different worker parameters

Blue - Workers = 6
 Orange - Workers = 9
 Green - Workers = 12
 Red - Workers = 15

