Machine Learning with Large Datasets
04-04-00-10-42-16-1-13300
Assignment 2

November 11, 2017

1. **Results:**

Table 1: Accuracies of In Multiclass Logistic Regression using Stochastic Gradient Descent in %

Train	Test	Development
98.7	91.8	90.9

Table 2: Training Time And Testing Time in Seconds
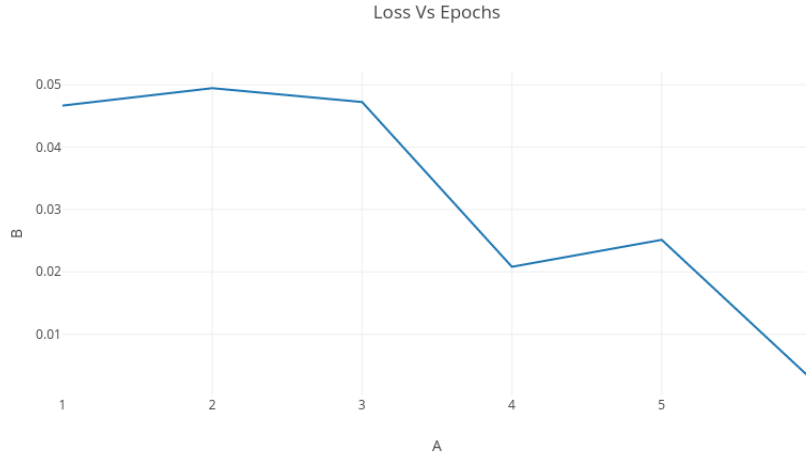
Train	Test
30219	60

Figure 1: Plot of Loss vs Epoch

Reason For Numbers: Since the learning method is a stochastic approach, we are approximating the gradient by taking one row for its calculation. Hence the loss increased slightly first and then decreased.

Table 3: Accuracies of In Multiclass Logistic Regression using Stochastic Gradient Descent in % with increasing learning rate

Train	Test	Development
98.7	83.3	82.7

Table 4: Training Time And Testing Time in Seconds

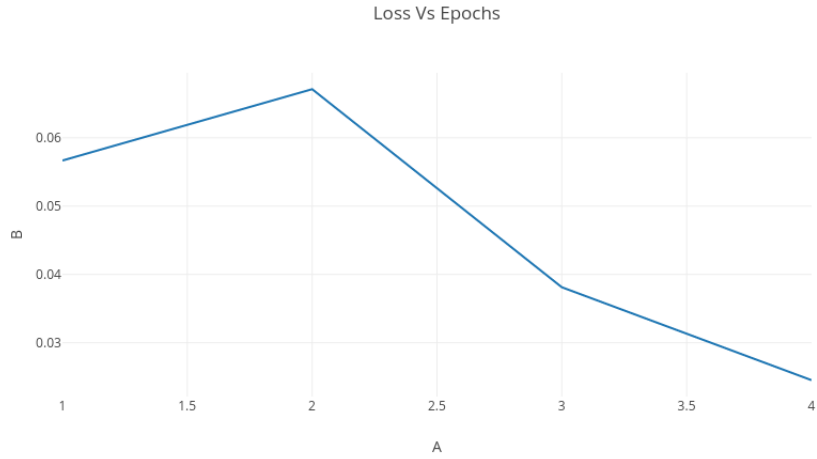| Train	Test
18590	59

Loss Vs Epochs

Figure 2: Plot of Loss vs Epoch

Reason For Numbers: Since the learning method is a stochastic approach, we are approximating the gradient by taking one row for its calculation. Hence the loss increased slightly first and then decreased. Since the learning rate is increasing, the rate of convergence is high because it is taking huge steps at a time. How ever, this huge steps means values can go high sometimes by skipping the minima. Hence the alternating increased in the graph

Table 5: Accuracies of In Multiclass Logistic Regression using Stochastic Gradient Descent in % with decreasing learning rate

Train	Test	Development
98.9	87.9	88.9

Table 6: Training Time And Testing Time in Seconds
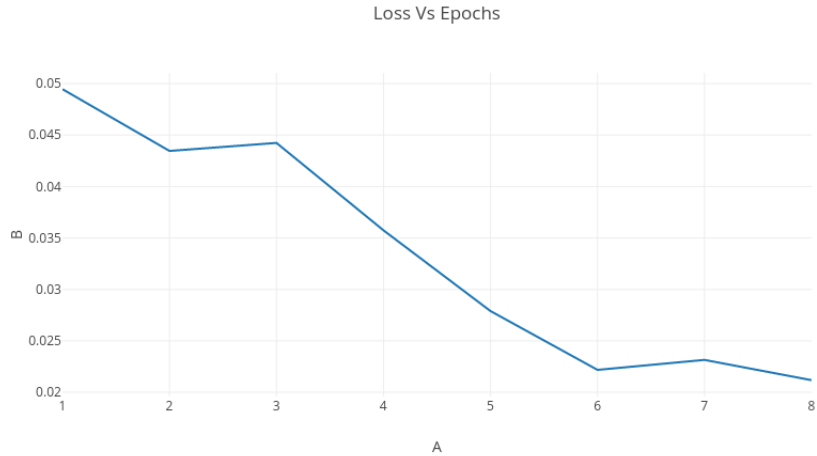
Train	Test
37289	59

Figure 3: Plot of Loss vs Epoch

Reason For Numbers: Since the learning method is a stochastic approach, we are approximating the gradient by taking one row for its calculation. Hence the loss increased slightly first and then decreased. Since the learning rate is decreasing, the training will take longer time to converge as it is takes small steps to converge.

2. The freezed hyperparameter alpha was 0.01 and lambda was 0.001 It is an API to be used in spark. For the second part, I have tried to implement the parameter server using DistML API. I didnt get enough time to implement the parameter server since there was no support for the same API available in the internet. I did run a Logistic regression on the spark machine once,using 1 parameter server, however it was holding all the resources and took around 8-9 hours to finish for training. It was an BSP implementation with maximum lag size of 2. I got a training accuracy of 96 % and a test accuracy of 71 %. I didn't get time to implement the rest.