
DS222 Assignment 2

Bharat Kumar Padi¹

1. Part 1: Local Logistic Regression

Data set: DBpedia full

Input: Each sentence is one hot encoded and presented to the classifier. Maximum number of unique words is restricted to 10000.

Classifier: L2 regularized logistic regression

Optimizer: SGD

Hyper parameters: Learning rate: 0.01

Mini batch size: 50

L2 weight: 0.01

Accuracies:

On Train data: 78.55%

On Test data: 74.68%

Timings:

Training: 1226.47 s

Testing: 1.118 s

Training Loss:

Three learning rate strategies are used, constant, exponentially increasing, exponentially decreasing. Following function was used for exponentially varying the learning rate:

$\text{decayed_Learning_Rate} =$

$\text{learning_rate} * \text{decay_rate}^{(\text{global_step} / \text{decay_steps})}$

learning_rate is the initial learning rate, global_step increases once after processing of every mini batch. decay_steps controls how often we want to update the learning rate, decay_rate value more than 1 results in exponentially increasing learning rate and less than 1 results in exponentially decreasing learning rate.

learning_rate = 0.01

decay_rate = 0.9 and 1.5

decay_steps = 10000

2. Part 2:

I have chosen tensorflow to implement the distributed models both ASGD and SSP SGD. This is partly because I'm already using tensorflow as part of my MTech project and also because tensorflow uses GPUs to speed up the compu-

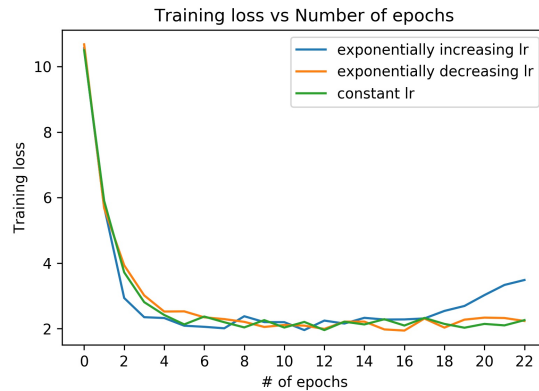


Figure 1. Training loss vs no. of epochs

tations.

2.1. Asynchronous SGD

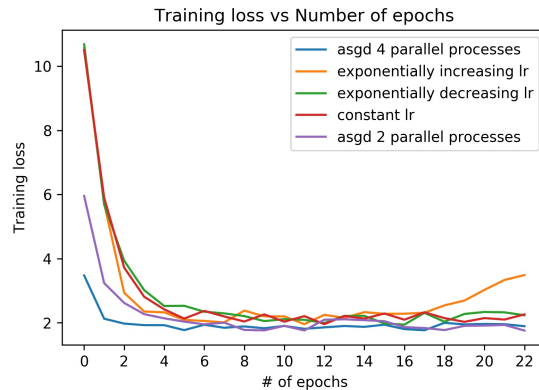


Figure 2. Training loss vs no. of epochs

| No of workers | Train time | Test time | Train Acc | Test Acc |
|---------------|------------|-----------|-----------|----------|
| 2 | 811.74 | 1.027 | 79.62 | 76.60 |
| 4 | 352.94 | 1.01 | 79.58 | 76.597 |

2.2. SSP SGD:

Train and Test accuracies are same as those of ASGD.

| Staleness | Train time | Test time |
|-----------|------------|-----------|
| 4 | 424.95 | 0.899 |
| 10 | 359.14 | 1.144 |

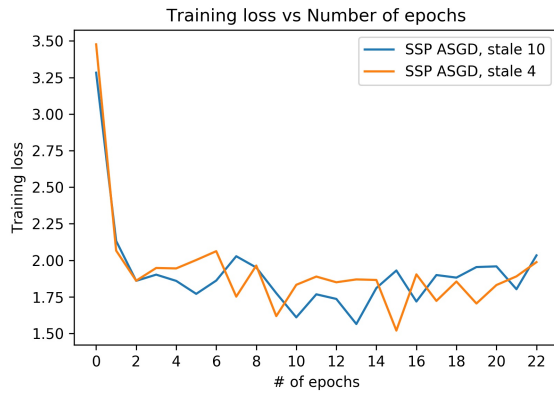


Figure 3. Training loss vs no. of epochs

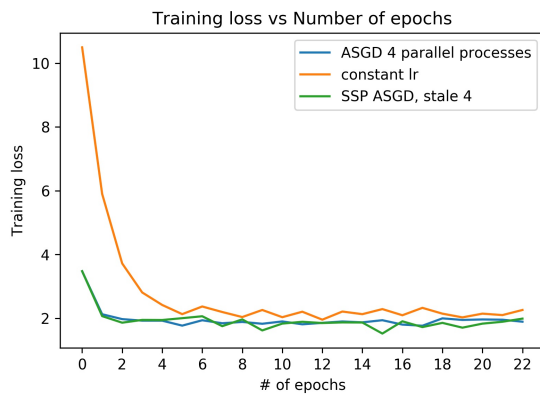


Figure 4. Training loss vs no. of epochs