DS 222
Assignment 2

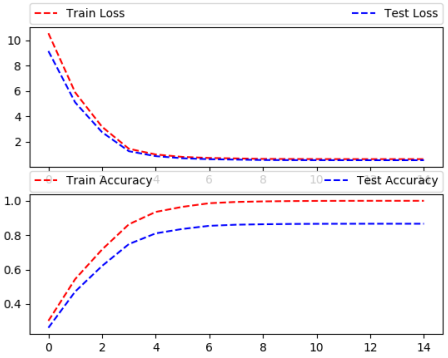
1. Local Implementation

Implemented logistic regression for text classification using the DBPedia dataset.

Details of the implementation are:
1.Load the text and split it into tokens.(Remove stop words, urls and clean the text)
2.Prepare the dictonary(vocabulary) of words in all the documents.
3.Using the dictonary created above convert each document into one hot encoded input vector.
4.Create a weight vector of size (number-of-classes x number-of-words)
5.Calculate the probabilities by taking the dot product of the weight vector and the input vector
6.Calculate the cross-entropy loss using the probabilities and ground truth.
7.Update the weights with gradient descent scheme.

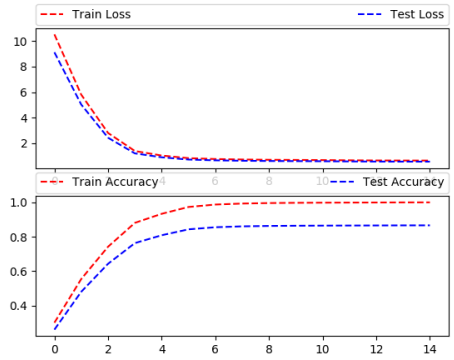
2. Experiments

1.Graph of Training and Test Loss/Accuracy with a Constant learning rate of 1.



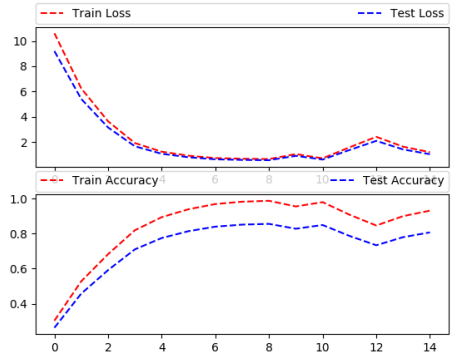
Training loss=0.629 Test loss=0.5458 Training Accuracy=99Testing Accuracy=86.66

2.Graph of Training and Test Loss/Accuracy with a Decreasing learning rate starting with 1 and decreasing/decay by 1.1 every epoch.



Training loss=0.645 Test loss=0.5590 Training Accuracy=99.9Testing Accuracy=86.59

3.Graph of Training and Test Loss/Accuracy with a Increasing learning rate of 1 starting with 1 and increasing by factor of 1.1 every epoch.



Training loss=1.2 Test loss=1.03 Training Accuracy=93Testing Accuracy=80

3. Parallel Implementation

Chose tensorflow as parameter server.

Due to lack of quality results, have included the working code in the repoisitory and have excluded the results from the report. Collaborated with Ramabhadhra to setup the parameter server using tensorflow.