

---

# Assignment 2 - Logistic Regression

---

Aditya Sharma <sup>1</sup>

## Abstract

Here, I present and contrast two different implementations of the Logistic Regression algorithm with L2 regularization for document classification - one on single threaded sequential environment and the other in a distributed manner using the Parameter Server (Li et al.) framework.

## 1. Preprocessing

I used code the following page <https://www.kdnuggets.com/2018/03/text-data-preprocessing-walkthrough-python.html> for preprocessing.

I removed stop words, converted to lower case and removed html code, among other things. Importantly, I only sample words that have a word count greater than 3 in the entire dataset.

## 2. Parameters

We use learning rate = 0.005 and regularization constant 50e-6.

Total number of parameters = (Number of words in the vocabulary \* Number of classes)

Number of classes = 50

Number of words in vocabulary = 74,033

Hence, number of parameters is 3,701,650.

## 3. Local

For constant learning rate (lr), training time per epoch was 9414.34 seconds on an average and test time was 64.04 seconds.

For increasing lr, we use the function  $k * \log(t + 1) * lr$ ,

---

<sup>\*</sup>Equal contribution <sup>1</sup>Department of Computational and Data Sciences, Indian Institute of Science, Bangalore. Correspondence to: Aditya Sharma <adityasharma@iisc.ac.in>.

*Proceedings of the 35<sup>th</sup> International Conference on Machine Learning*, Stockholm, Sweden, PMLR 80, 2018. Copyright 2018 by the author(s).

	Train	Dev	Test
Local Constant LR	79.01	80.46	84.03
Local Decreasing LR	71.35	70.78	73.08
Local Increasing LR	72.24	71.11	74.62

Table 1. Accuracy comparisons between local for different LR schemes

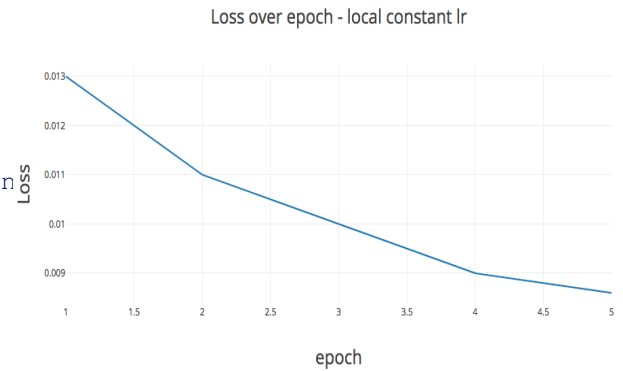


Figure 1. Training Loss per epoch for constant LR

where k is a constant and t is the time step. The training time is 12458.71 seconds per epoch and test time is 150.30 seconds.

For decreasing lr, we use the function  $k * \exp(-(t + 1)) * lr$ , where k is a constant and t is the time step. The training time is 9132.45 seconds per epoch and test time is 60.03 seconds.

## References

Li, M., Andersen, D. G., and Park, J. W. Scaling distributed machine learning with the parameter server.

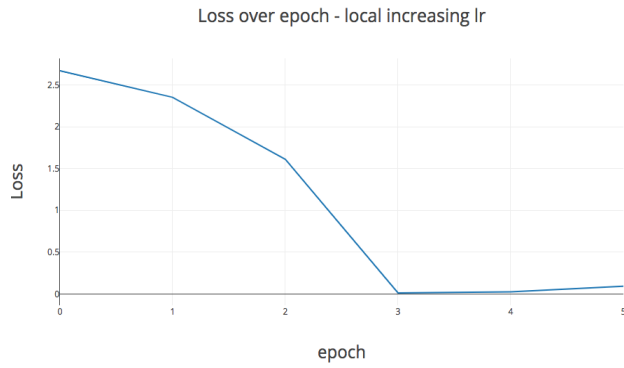


Figure 2. Training Loss per epoch for increasing LR

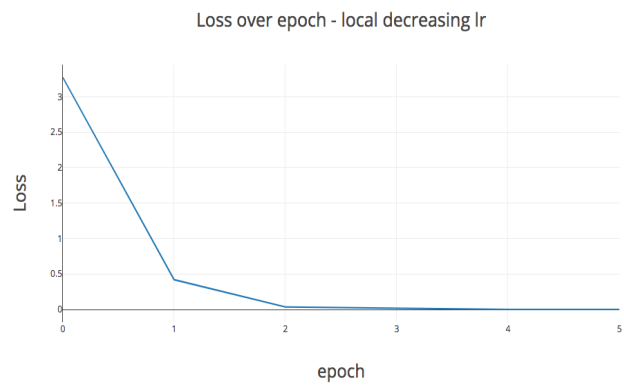


Figure 3. Training Loss per epoch for decreasing LR