

The Experiment Report of Machine Learning

SCHOOL: SCHOOL OF SOFTWARE ENGINEERING

SUBJECT: SOFTWARE ENGINEERING

|  |  |
| --- | --- |
| Author: | Supervisor: |
| Zhifani Xu | Qingyao Wu |

|  |  |
| --- | --- |
| Student ID: | Grade: |
| 201530613269 | Undergraduate |

December 14, 2017

1

Logistic Regression, Linear Classification and Stochastic Gradient Descent

***Abstract—Implementing question of logistic regression and liner classification by stochastic gradient descent(SGD). And updating model parameters using different optimized methods(NAG, RM-***

***Sprop, AdaDelta and Adam***

1. INTRODUCTION

Logistic regression and linear classification are basic knowledge of the machine learning. And in this experiment I used stochastic gradient descent(SGD) to implement the previous two problems.At the same time I used four optimized methods to optimize both of them. This experiment helps me understanding the logistic regression and linear classification. What’s more, the four methods works well.

ǁ. METHODS AND THEORY

*A. Logistic Regression*

The logistic regression function is a common equation as follows:

 (1)

Where , we can obtain the continuous function:

 (2)

The loss function is:

 (3)

The derivation of this loss function is:

 (4)

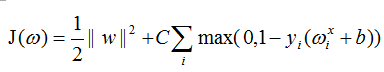
*B. Linear Classification*

I implemented linear classification using support vector ma-

chine(SVM). The margin is defined as follows:

 (5)

We have the loss function of SVM:

 (6)

and its derivation:

2017-12-14_235807 (7)

*C. Optimization Methods*

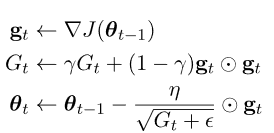
Some optimization methods had been used in this e- xperiment such as NAG, RMSProp, AdaDelta and Adam.

*a. NAG:* NAG is a way to give our momentum term this kind of prescience that slowing down before the loss rise again. When calculating 2017-12-15_005011, we use instead of C:\Users\del l\Pictures\2017-12-15_005736.png2017-12-15_005736, which is equivalent to using momentum to predict where to go next.

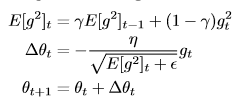


 (8)

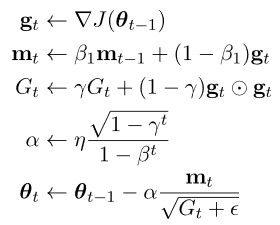
*b. RMSProp:* RMSprop and Adadelta have both been developed independently around the same time stemming from the need to resolve Adagrad's radically diminishing learning rates. RMSprop in fact is identical to the ﬁrst update vector of AdaDelta.

 (9)

*c. AdaDelta:* Adadelta restricts the window of accumulated past gradients to some fixed size ω, instead of accumulating all past squared gradients. Moreover, instead of inefficiently storing ω previous squared gradients, the sum of gradients is recursively deﬁned as a decaying average of all past squared gradients. The running average  at time step t then depends (as a fraction γ similarly to the Momentum term) only on the previous average and the current gradient:

 (10)

*d. Adam:* Adam is another method that computes adaptive learning rates for each parameter. In addition to storing an exponentially decaying average of past squared gradients  like Adadelta and RMSprop, Adam also keeps an exponentially decaying average of past gradients mt, similar to momentum:

 (11)

Ш.EXPERIMENTS

A. Data set

This experiment uses a9a which provided by LIBSVM Data, including 32561 training, 16281 testing records and each record in training or testing set has 123 features.

B. Implementation

In both logistic regression and linear classification, the parameter loops is limited to 666, for comparing the performance of four optimization methods. At the same time, other parameters like ε in those methods are adjusted respectively for better accuracy.

The graph of loss descent using NAG, RMSProp, AdaDelta and Adam respectively are shown as Fig.1 amd Fig.2:

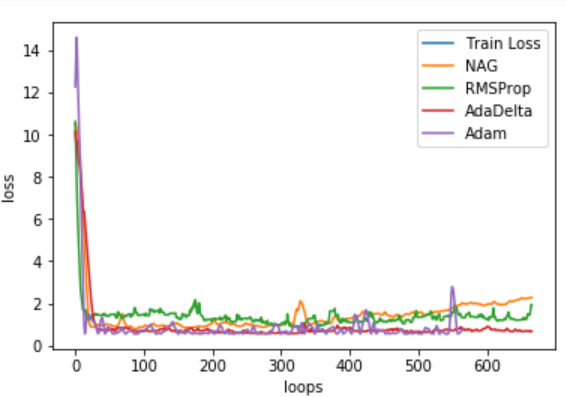


Fig. 1. Loss of va set on logistic regression

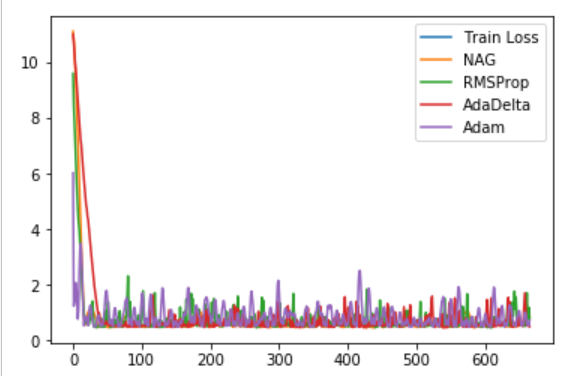


Fig. 2. Loss of training set on linear classification

IV. CONCLUSION

I have a deeper understanding of the logistic regression and the linear classification after this experiment. What’s more, I have known that the principle of stochastic gradient descent(SGD). Writing the code of optimization methods such as NAG, PMSProp, AdaDelta and Adam makes me know more about optimization in machine learning, which may help me improve the efficiency of my code in the future.