The Experiment Report of Machine Learning



December 9, 2017

Grade:

Undergraduate

Student ID：

201530612767

Supervisor:

Qingyao Wu

Author:

Wanting Su

**SUBJECT:**SOFTWARE ENGINEERING

**SCHOOL:** SCHOOL OF SOFTWARE ENGINEERING

[[1]](#footnote-1)

Linear Regression, Linear Classiﬁcation and Gradient Descent

Abstract—

Abstract——In the experiment we compare the difference between gradient descent and stochastic gradient descent. We also compare the differences and relationships between Logistic regression and linear classification. From the experiment we got a further understand the principles of SVM and practice on larger data.

# INTRODUCTION

Gradient descent is a first-order iterative optimization algorithm for finding the minimum of a function. To find a local minimum of a function using gradient descent, one takes steps proportional to the negative of the gradient (or of the approximate gradient) of the function at the current point. If instead one takes steps proportional to the positive of the gradient, one approaches a local maximum of that function; the procedure is then known as gradient ascent.

Support vector machines are supervised learning models with associated learning algorithms that analyze data used for classification and regression analysis. Given a set of training examples, each marked as belonging to one or the other of two categories, an SVM training algorithm builds a model that assigns new examples to one category or the other, making it a non-probabilistic binary linear classifier.

Linear Classification based on SVM and logic regression are the two models we use to apply gradient decent. There are many ways to update parameters. In this experiment, we compare four different ways of stochastic gradient decent, called NAG，RMSProp，AdaDelta and Adam.

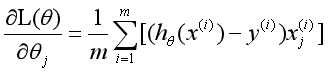
In the experiment we implements the two models mentioned above and use different gradient decent technique to compare the performance of them.

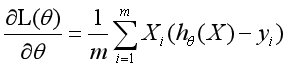
# METHODS AND THEORY

## For logistic regression:

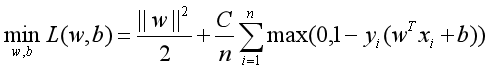
## Loss function is: C:\Users\siye\AppData\Local\Temp\ksohtml\wpsC045.tmp.png

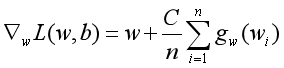
In which C:\Users\siye\AppData\Local\Temp\ksohtml\wps85D7.tmp.jpg

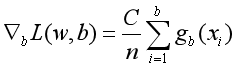
Gradient: 



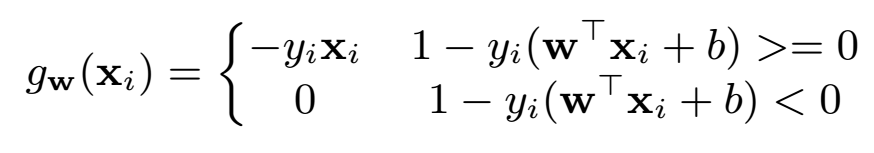
## For linear classification:

Loss function is: 

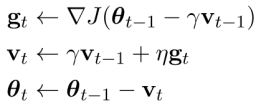
Gradient: 



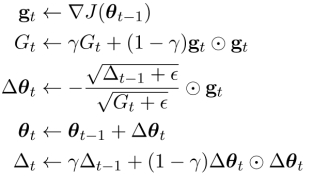
In which we have:



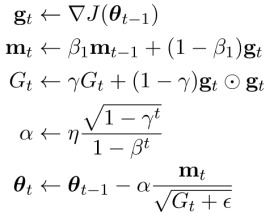
*C . Batch Gradient decent and Stochastic Gradient Decent*



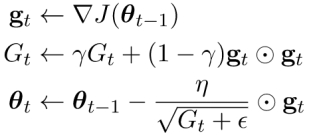
*Figure 1:NAG update rules*



*Figure 2:AdaDelta update rules*



*Figure 3:Adam update rules*



*Figure 4:RMSProp update rules*

# Experiment

## Hyper Parameters Selection

For logistic regression:

（1）NAG：η=0.005 =0.9 epoch = 300

（2）RMSProp:η=0.005 = 0.9 = 1e-8 epoch =300

（3）AdaDelta: = 0.9 = 1e-8 epoch=100

（4）Adam:η=0.005 = 0.9 = 0.999 = 1e-8 epoch=300

For linear regression:

(1)Learning rate:0.07 Iteration number:500

C:5 Batch size:100

(2)NAG: r=0.9

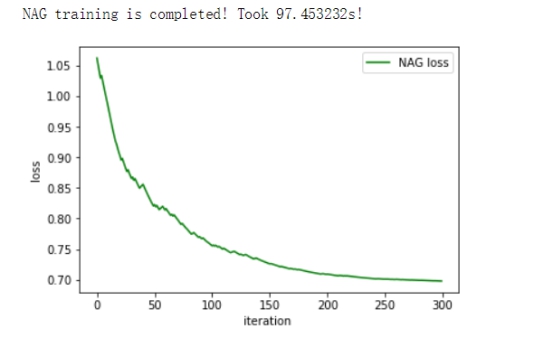
(3)RMSProp: r= 0.9 ε=1e-8

(4)AdaDelta: r=0.95ε=1e-4

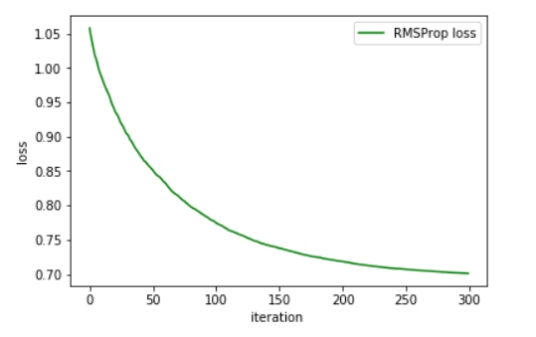
(5)Adam r=0.9ε=1e-6 β=0.9

## Experience Result

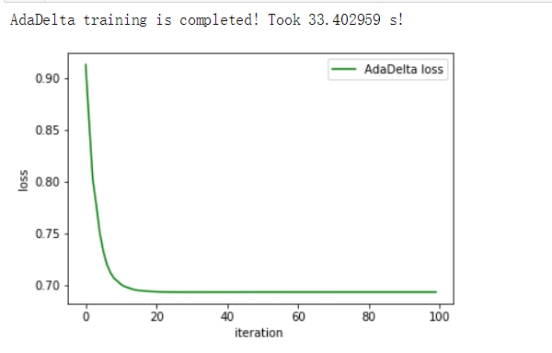
For Logistic Regression:



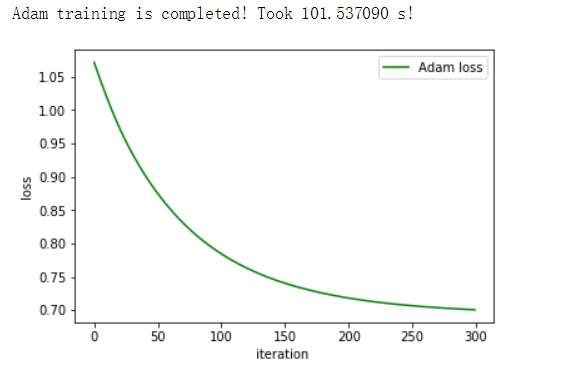
*Logistic Regression Loss figure of NAG*



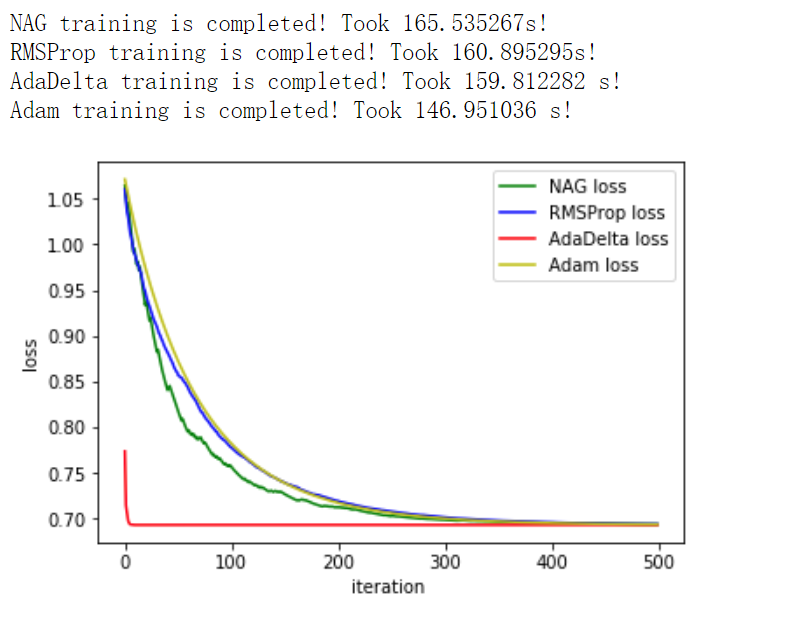
*Logistic Regression Loss figure of*  *RMSProp*



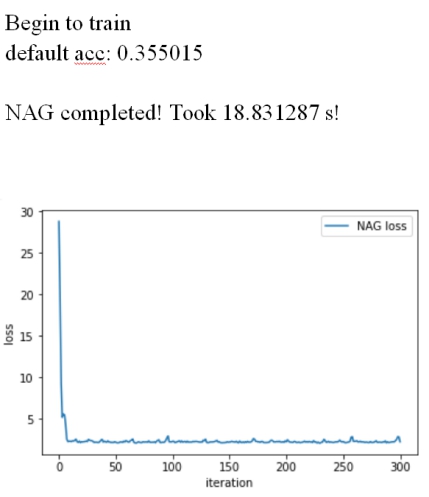
*Logistic Regression Loss figure of AdaDelta*



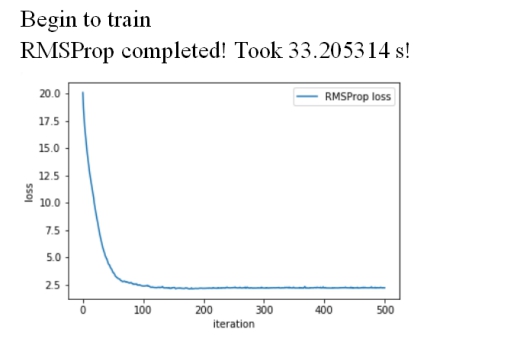
*Logistic Regression Loss figure of*  *Adam*



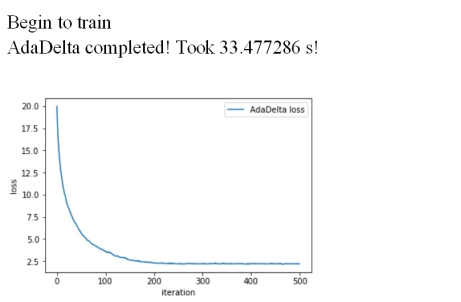
·For Linear Regression

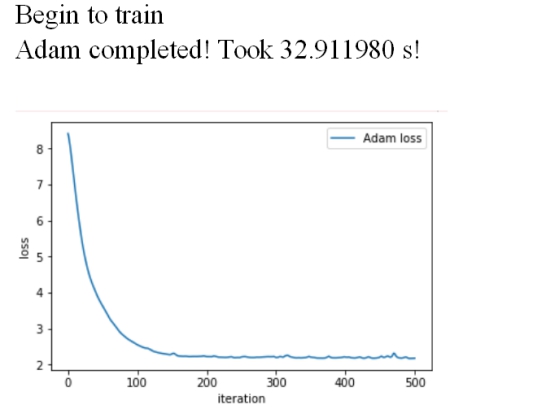


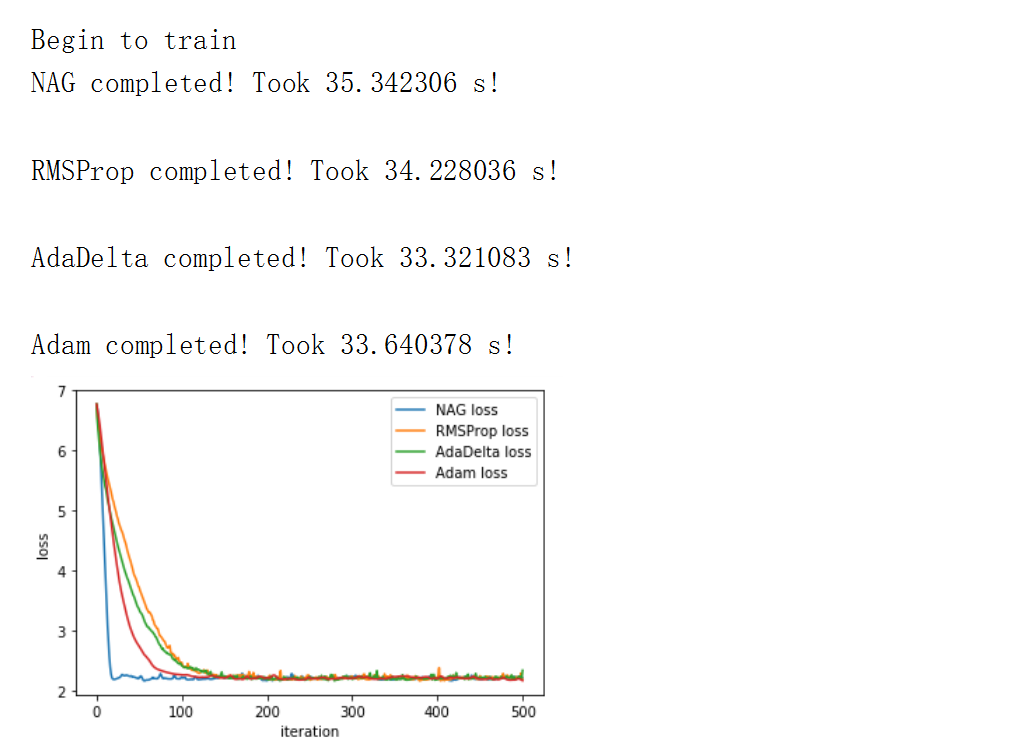
*Linear Classification Loss figure of*  *NAG*



*Linear Classification Loss figure of RMSProp*

 *Linear Classification Loss figure of AdaDelta*

 *Linear Classification Loss figure of Adam*



# conclusion

In this experiment we implement two model and four different algorithm, compare the difference between gradient descent and stochastic gradient descent. From the experiment we got a further understand the principles of SVM and practice on larger data.The detail of different algorithm does benefit the study in the future research .

1. [↑](#footnote-ref-1)