The Experiment Report of Machine Learning



December 9, 2017

Grade:Undergraduate

Undergraduate or Graduate

Student ID：201530613078 201530611111 and 201530600000

Supervisor:

Mingkui Tan

Author:Xianzhe Wu

Shoukai Xu and Yaofu Chen

**SUBJECT:**SOFTWARE ENGINEERING

**SCHOOL:** SCHOOL OF SOFTWARE ENGINEERING

[[1]](#footnote-0)

Experimental Study on Stochastic Gradient Descent for Solving Classification Problems

Abstract—

We conducted two experiments on stochastic gradient descent , using logistic regression and linear classification . We used four methods to optimize the process of gradient descent in each experiment . We wanted to compare the efficiency and the results in four cases of each experiment .

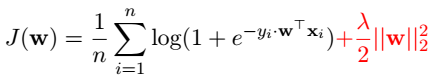
# INTRODUCTION

In experiments , our main idea was to use two models to solve classification problems . They were logistic regression model and support vector machine model . As for stochastic gradient descent in each model , we used four methods to optimize . The methods were respectively NAG , RMSProp , AdaDelta and Adam .

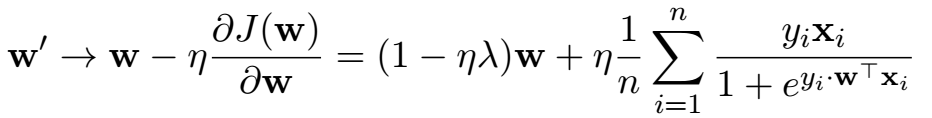
We wanted to find out the influence of adjusting parameters to different optimizing process and compared the efficiency between optimizing methods . And we would figure out the difference between models .

# METHODS AND THEORY

In logistic regression :

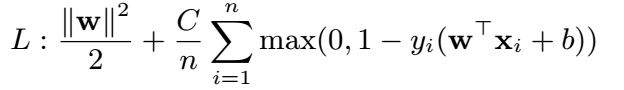
Our loss function 

The update of w



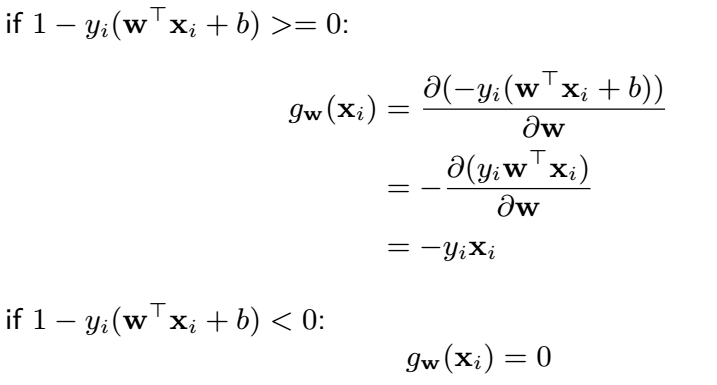
In linear classification :

Our loss function



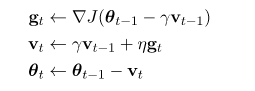
The update of w

W’ -> W - η(W+gw(Xi))



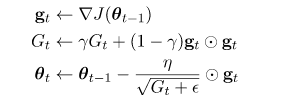
Four optimizing methods :

NAG:



Here we used a new variable v to predict the next position which gt would reach . And v was used to get the weighted average direction from the direction now and the directions before .(We set γ as 0.9 )

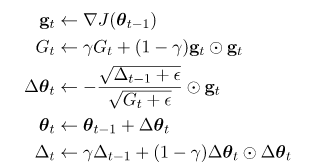
RMSProp:



Gt used the past gradient information to judge which features were often updated .

(We set γ as 0.9 and ε as 1e - 8)

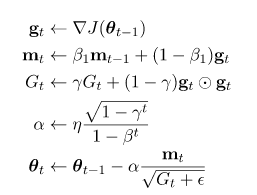
AdaDelta:



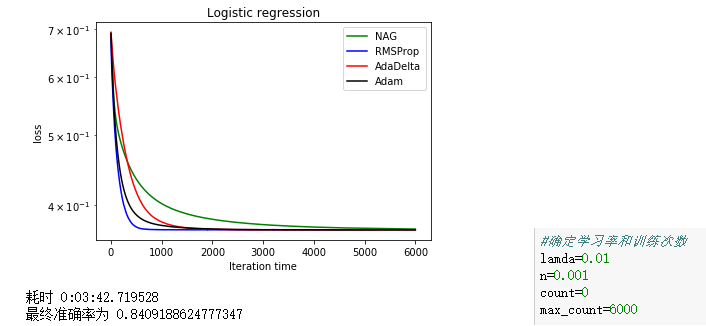
In this method , was used to estimate the learning rate . In other words , this method estimated next step size by the past step size information .

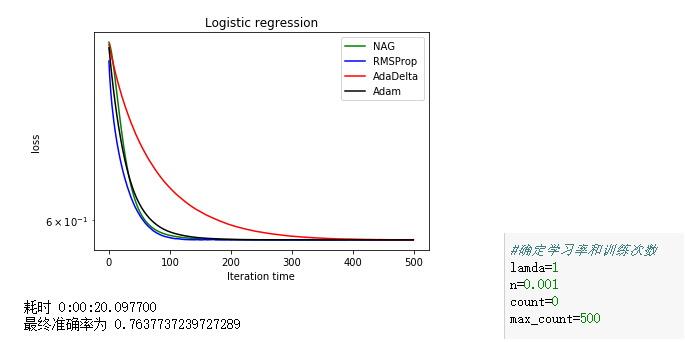
(We set γ as 0.95)

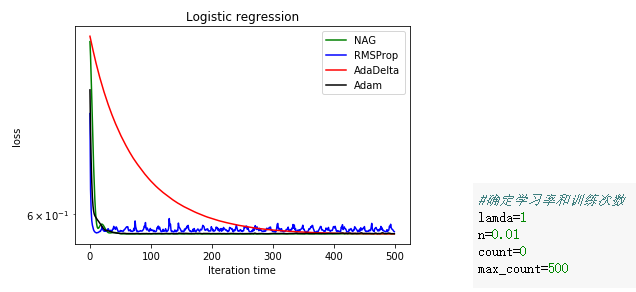
Adam:

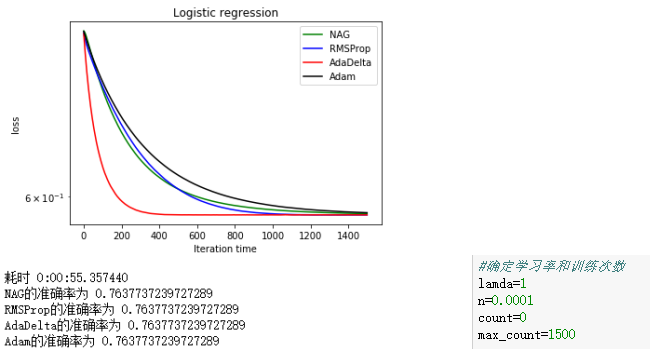


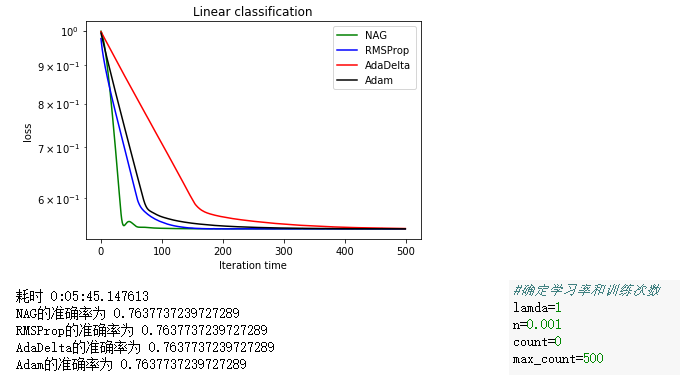
# Experiment

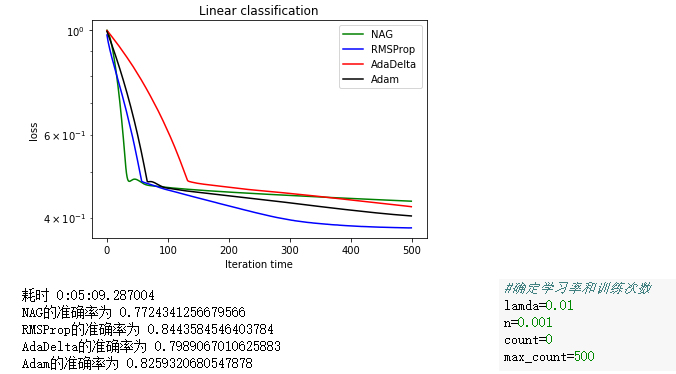


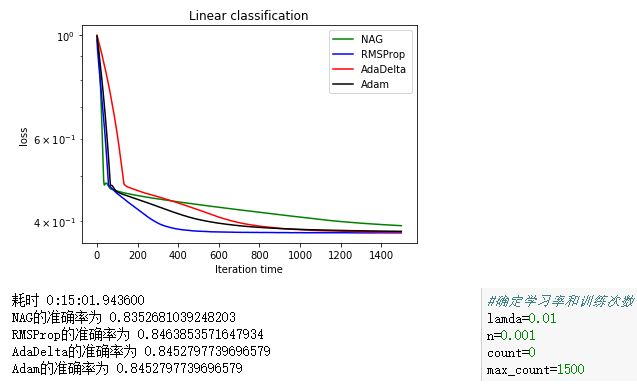












From the picture 1 and 2 , we can see that the increase of lamda speed up the learning process and the convergence . But we can also notice the accuracy rate decreases . And in the four optimizing methods , the efficiency of NAG is influenced to a great extent .

From the picture 2 , 3 and 4. We can see AdaDelta is the most stable curve and when learning rate was set smaller , it became the most efficient . When n increase from 0.001 to 0.01 ,RMSProp can’t converge .

From the picture 2 and 5 . We can see the results in both model were similar . But from the picture 1, 6 and 7, logistic regression model costed less study rounds than linear classification .

# conclusion

The influence of adjusting parameters to different optimizing process and the comparison of efficiency between optimizing methods :

Increasing lamda will speed up whole learning and may decrease the accuracy rate .It influence NAG most in logistic regression .

Increasing learning rate , RMSProp firstly become hard to converge . NAG and Adam converge quickly . Decreasing learning rate , AdaDelta is the most efficient .

In general , logistic regression’s learning process is shorter than linear classification .

se

1. [↑](#footnote-ref-0)