华南理工大学

South China University of Technology

The Experiment Report of Machine Learning

SCHOOL: SCHOOL OF SOFTWARE ENGINEERING

SUBJECT: SOFTWARE ENGINEERING

Author:
YangyangZhao

Supervisor:
Qingyao Wu

Student ID：
201710106598

Grade:
Graduate

December 13, 2017

Logistic Regression, Linear Classification and Stochastic Gradient Descent

Abstract—Logistic regression is usually used for the binary classification model. The objective function is the second class cross entropy, and the value of y is the probability of the first class. The user can set a classification threshold by himself.

Linear regression is used to fit data, and the objective function is plain and error.

I. Introduction

Purposes:

Compare and understand the difference between gradient descent and stochastic gradient descent.

Compare and understand the differences and relationships between Logistic regression and linear classification.

Further understand the principles of SVM and practice on larger data.

Data sets and data analysis:

Experiment uses a9a of LIBSVM Data, including 32561/16281(testing) samples and each sample has 123/123 (testing) features.

II. Methods and Theory

Logistic Regression and Stochastic Gradient Descent

1.Load the training set and validation set.

2.Initalize logistic regression model parameters, you can consider initalizing zeros, random numbers or normal distribution.

3.Select the loss function and calculate its derivation, find more detail in PPT.

4.Calculate gradient toward loss function from partial samples.

5.Update model parameters using different optimized methods(NAG，RMSProp，AdaDelta and Adam).

6.Select the appropriate threshold, mark the sample whose predict scores greater than the threshold as positive, on the contrary as negative. Predict under validation set and get the different optimized method loss L_{NAG}, $L_{RMSProp}$, $L_{AdaDelta}$ and L_{Adam}.

7.Repeate step 4 to 6 for several times, and drawing graph of L_{NAG}, $L_{RMSProp}$, $L_{AdaDelta}$ and L_{Adam} with the number of iterations.

Linear Classification and Stochastic Gradient Descent

1.Load the training set and validation set.

2.Initalize SVM model parameters, you can consider initalizing zeros, random numbers or normal distribution.

3.Select the loss function and calculate its derivation, find more detail in PPT.

4.Calculate gradient toward loss function from partial samples.

5.Update model parameters using different optimized methods(NAG，RMSProp，AdaDelta and Adam).

6.Select the appropriate threshold, mark the sample whose predict scores greater than the threshold as positive, on the contrary as negative. Predict under validation set and get

the different optimized method loss L_{NAG}, $L_{RMSProp}$, $L_{AdaDelta}$ and . L_{Adam}

7.Repeate step 4 to 6 for several times, and drawing graph of L_{NAG}, $L_{RMSProp}$, $L_{AdaDelta}$ and L_{Adam} with the number of iterations.

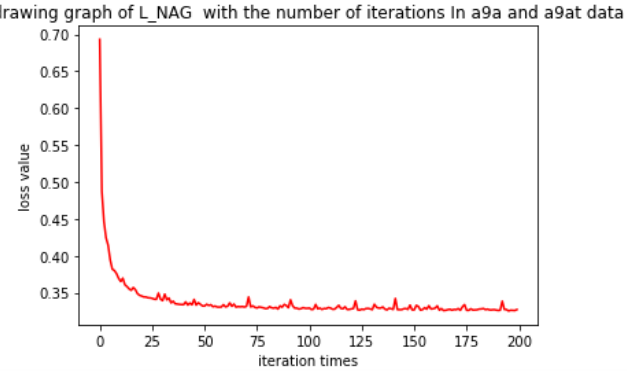
III. Experiment

The initialization method of model parameters:

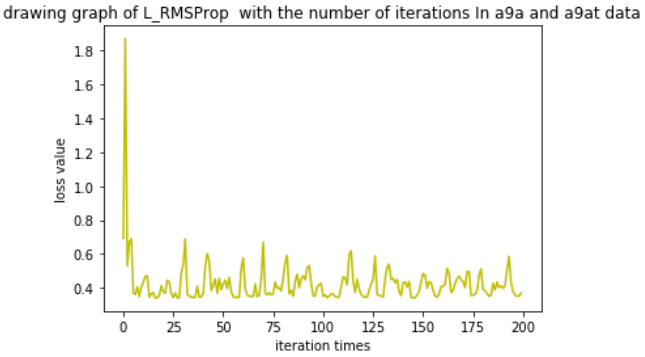
#Learning rate
learning_rate = 0.1
#the number of iteration
num_iter = 200
#Batch each training
k = 200
#Logistic gression parameter
eta = 0.001
gamma = 0.9
eps = 1e-8
beta=0.9

Experimental results and curve:
Logistic Regression and Stochastic Gradient Descent:

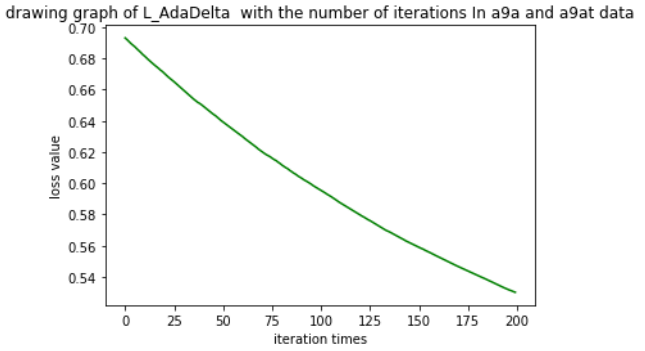
L_NAG



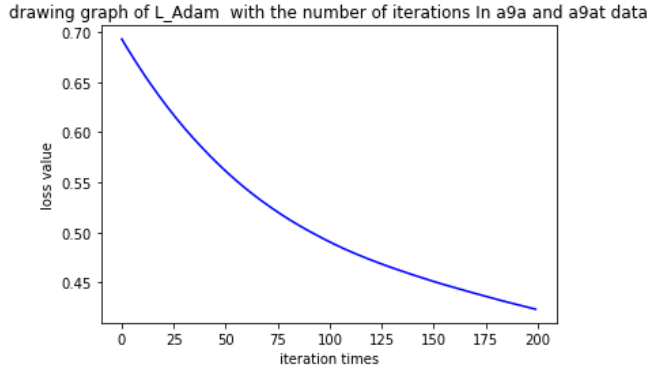
L_RMSProp



L_AdaDelta

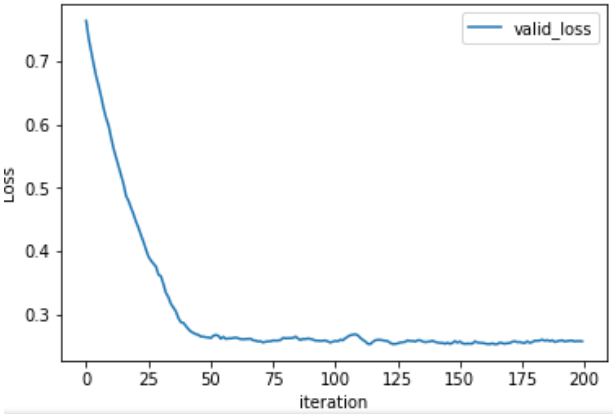


L_Adam

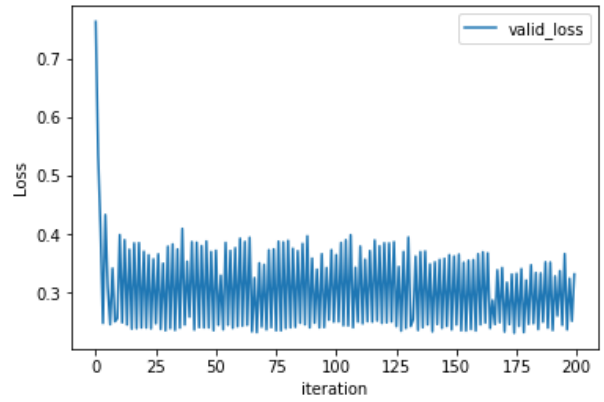


Linear Classification and Stochastic Gradient Descent:

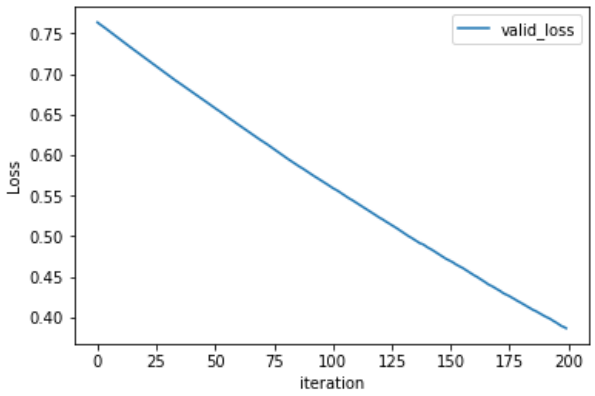
L_NAG



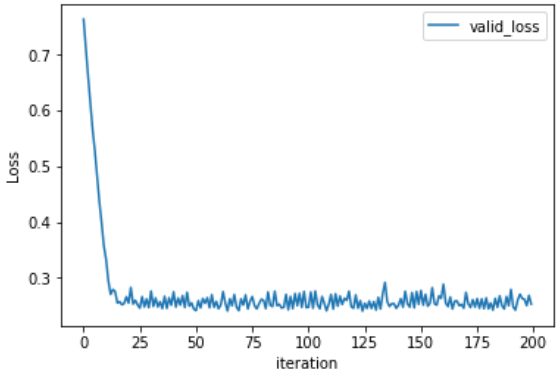
L_RMSProp



L_AdaDelta



L_Adam



IV. CONCLUSION

Logistic regression is usually used for the binary classification model. The objective function is the second class cross entropy, and the value of y is the probability of the first class. The user can set a classification threshold by himself.

Linear regression is used to fit data, and the objective function is plain and error.

Classification algorithm of two kinds of methods are common, from the point of the objective function, the difference is that logistic regression is the logistical loss, the SVM is the hinge loss. The purpose of these two loss function is to increase the weight of the data points for a greater influence on the classification, and classification relationship smaller proportion of data points. The processing of the SVM method is only considered the support vectors, is also the most relevant and classification of a few points, to learn a classifier, and logistic regression through nonlinear mapping, greatly reduced the weight of the plane distant point from the classification, promoted and classification of the relative weights of associated data point.