

**The Experiment Report of**

***Machine Learning***

**College Software College**

**Subject Software Engineering**

**Members**   **周宇琛**

**Student ID 201530613887**

**E-mail 1057433699@qq.com**

**Tutor**   **谭明奎**

**Date submitted** **2017. 12 .14**

**1. Topic:** Logistic regression, linear classification and stochastic gradient descent

**2. Time:** 2017/12/9

**3. Reporter:**周宇琛

**4. Purposes:**

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

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

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

**5. Data sets and data analysis:**

Experiment uses a9a of LIBSVM Data, including 32561/16281(testing) samples and each sample has 123/123 (testing) features. We need divide the training set and validation set.

**6. Experimental steps:**

**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 G 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.

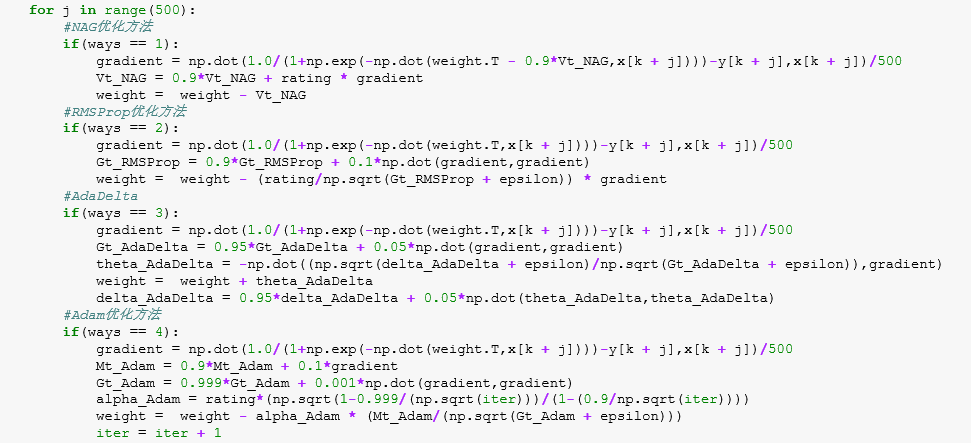
(5)、Update model parameters using different optimized methods(NAG，RMSProp，AdaDelta and Adam).

(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.

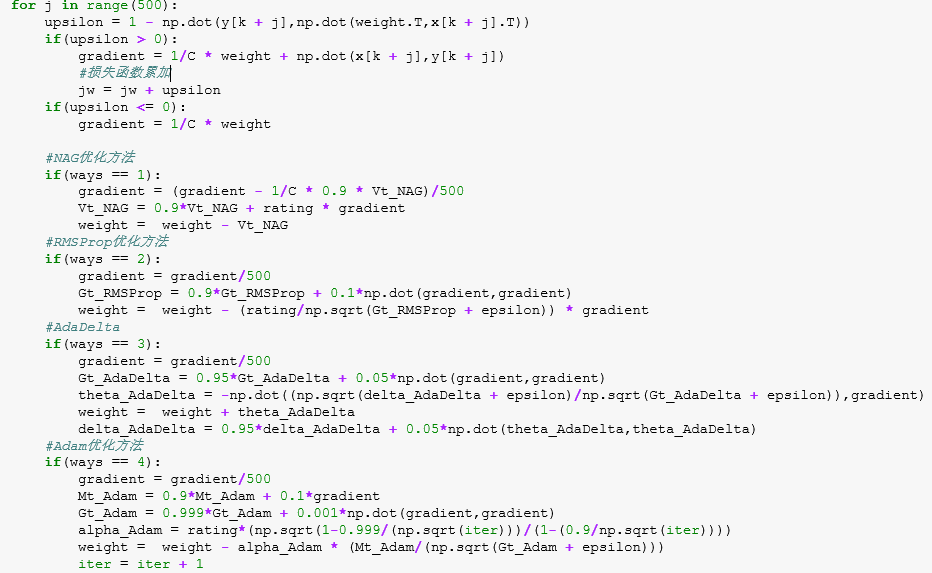
(7)、Repeate step 4 to 6 for several times, and drawing graph of L\_NAG，L\_RMSProp，L\_AdaDelta and L\_Adam with the number of iterations.

**7. Code:**

**Logistic Regression and Stochastic Gradient Descent:**



**Linear Classification and Stochastic Gradient Descent:**



**8. The initialization method of model parameters:**

**Logistic Regression and Stochastic Gradient Descent:**

All parameter initialization to 0 or 1

**Linear Classification and Stochastic Gradient Descent:**

All parameter initialization to 0 or 1

**9. The selected loss function and its derivatives:**

**Logistic Regression and Stochastic Gradient Descent:**

**Loss function:**

**Derivative:**

**Linear Classification and Stochastic Gradient Descent:**

**Loss function:**

**Derivative:**

**If :**

**Else:**

**10. Experimental results and curve:**(Fill in this content for various methods of gradient descent respectively)

**Logistic Regression and Stochastic Gradient Descent:**

## Hyper-parameter selection:

rating = 0.02 学习率

四种优化方法的参数初始化（NAG、RMSProp、AdaDelta、Adam）

jw = 0 梯度

Vt\_NAG = 0

Gt\_RMSProp = 0

Gt\_AdaDelta = 0

theta\_AdaDelta = 0

delta\_AdaDelta = 0

Gt\_Adam = 0

Mt\_Adam = 0

alpha\_Adam = 0

epsilon = 0.00001 防止分母为零的参数

## Predicted Results (Best Results):

准确率：

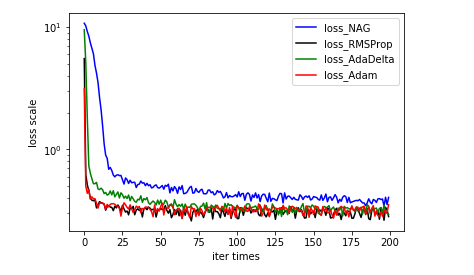
NAG:87%

RMSProp:88%

AdaDelta:87%

Adam:90%

## Loss curve:



**Linear Classification and Stochastic Gradient Descent:**

## Hyper-parameter selection:

rating = 0.03 学习率

C = 0.9 惩罚系数

四种优化方法的参数初始化（NAG、RMSProp、AdaDelta、Adam）

jw = 0 梯度

Vt\_NAG = 0

Gt\_RMSProp = 0

Gt\_AdaDelta = 0

theta\_AdaDelta = 0

delta\_AdaDelta = 0

Gt\_Adam = 0

Mt\_Adam = 0

alpha\_Adam = 0

epsilon = 0.00001 防止分母为零的参数

## Predicted Results (Best Results):

准确率：(当预测值符合设定阈值时)

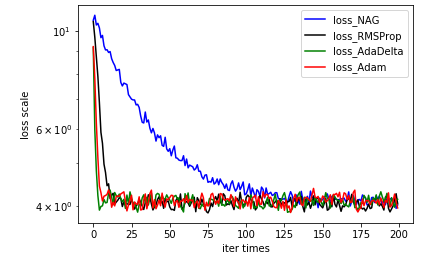
NAG:77%

RMSProp:78%

AdaDelta:78.4%

Adam:79%

## Loss curve:



**11. Results analysis:**

**Logistic Regression and Stochastic Gradient Descent:**

**We use four ways to**

NAG：

RMSProp：

AdaDelta：

Adam：

**Linear Classification and Stochastic Gradient Descent:**

**We use four ways to**

NAG：

RMSProp：

AdaDelta：

Adam：

**12. Similarities and differences between logistic regression and linear classification：**

**13. Summary:**