

**The Experiment Report of**

***Machine Learning***

**College Software College**

**Subject Software Engineering**

**Members**   **林杨**

**Student ID 201530612279**

**E-mail 291192462@qq.com**

**Tutor**   **谭明奎**

**Date submitted** **2017.Dec.12th**

**1. Topic:Logistic Regression, Linear Classification and Stochastic Gradient Descent**

**2. Time: 2017-12-02 2:00-5:00 PM**

**3. Reporter:林杨**

**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](https://www.csie.ntu.edu.tw/~cjlin/libsvmtools/datasets/binary.html" \l "a9a" \t "https://www.zybuluo.com/chenyaofo/note/_blank) of [LIBSVM Data](https://www.csie.ntu.edu.tw/~cjlin/libsvmtools/datasets/" \t "https://www.zybuluo.com/chenyaofo/note/_blank), including 32561/16281(testing) samples and each sample has 123/123 (testing) features

1. **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 LNAG ，LRMSProp，LAdaDelta and LAdam.

7.Repeate step 4 to 6 for several times, and drawing graph of ，， and 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 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 LNAG ，LRMSProp，LAdaDelta and LAdam.

7.Repeate step 4 to 6 for several times, and drawing graph of ，， and with the number of iterations.

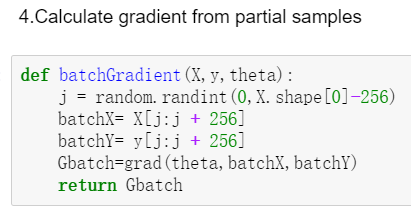
**7. Code:**

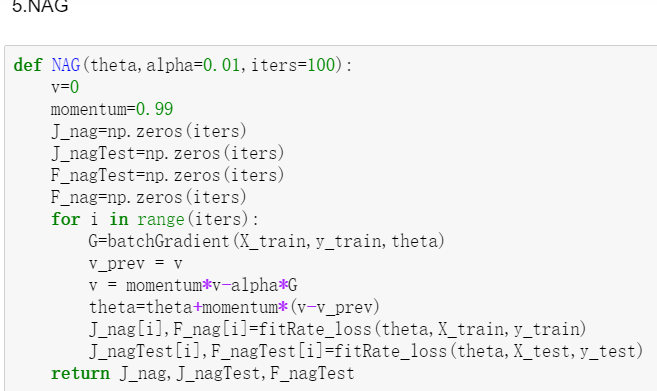
**Logistic Regression and Stochastic Gradient Descent**

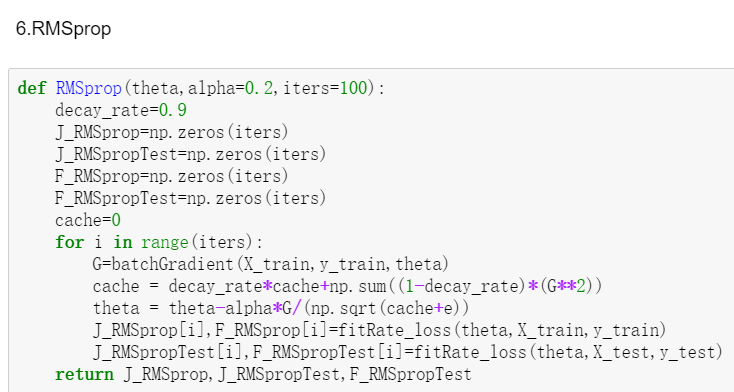
****

**Linear Classification and Stochastic Gradient Descent**

**Common Part（NAG,RMSProp,Adadelta,Adam)**

****

****

****

****

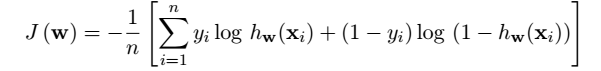
****

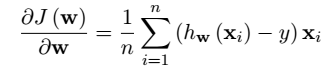
1. **The initialization method of model parameters:**

Initialize theta to ones, iterations to 1500 and learning rate to 0.05

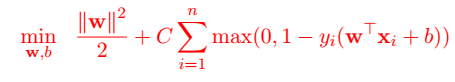
1. **The selected loss function and its derivatives:**

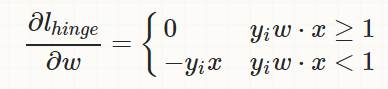
**Logistic Regression and Stochastic Gradient Descent**

****

****

**Linear Classification and Stochastic Gradient Descent**

****

****

**10. Experimental results and curve:**

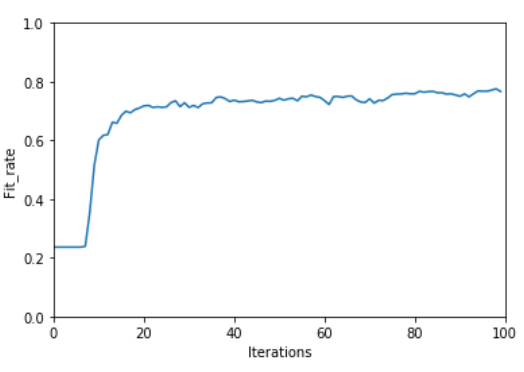
**Logistic Regression and Stochastic Gradient Descent**

Batch size =32

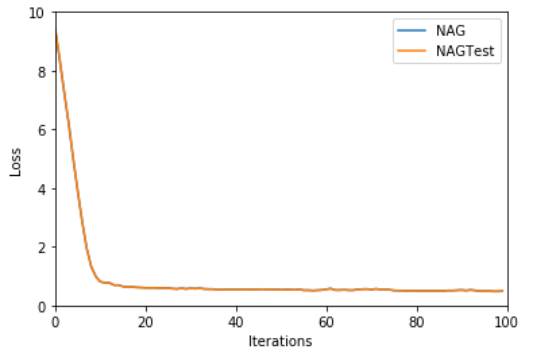
NAG:

## Hyper-parameter selection:momentum=0.99,alpha=0.01

## Predicted Results (Best Results):



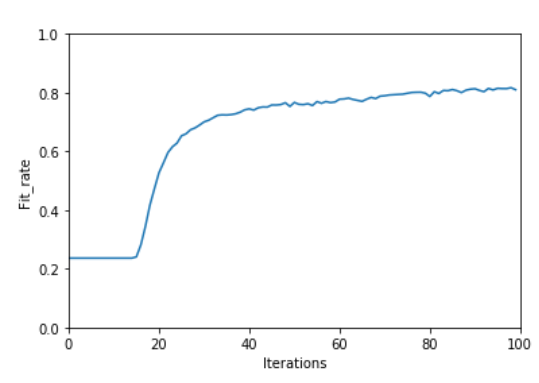
## Loss curve:



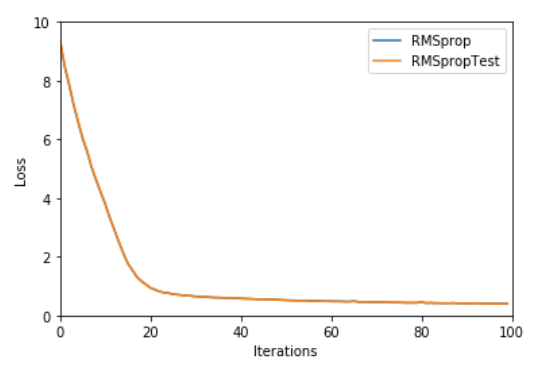
RMSProp:

## Hyper-parameter selection:decay\_rate=0.9,alpha=0.2

## Predicted Results (Best Results):



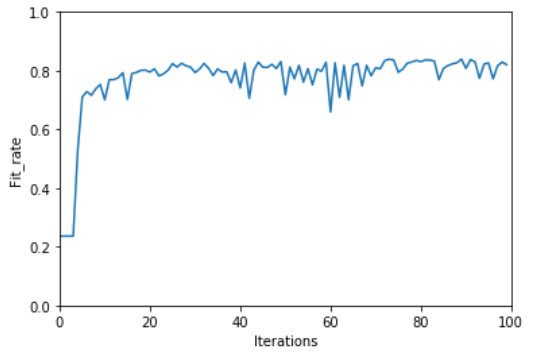
## Loss curve:



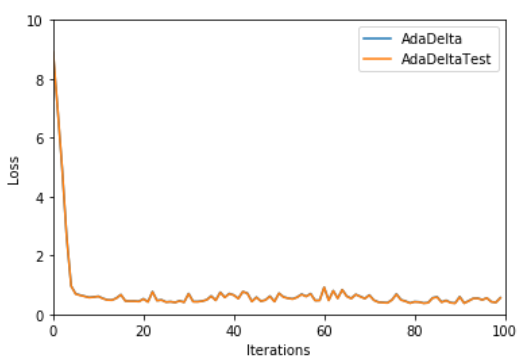
AdaDelta:

## Hyper-parameter selection:delta=0.5,alpha=2000

## Predicted Results (Best Results):



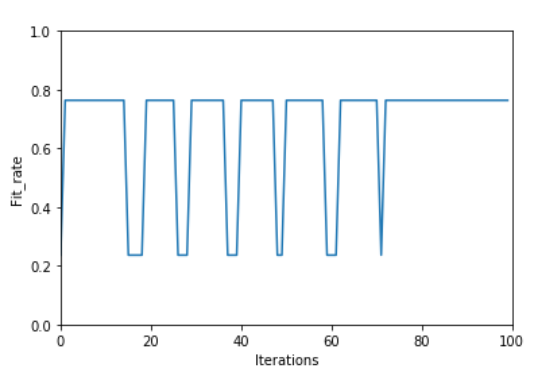
## Loss curve:



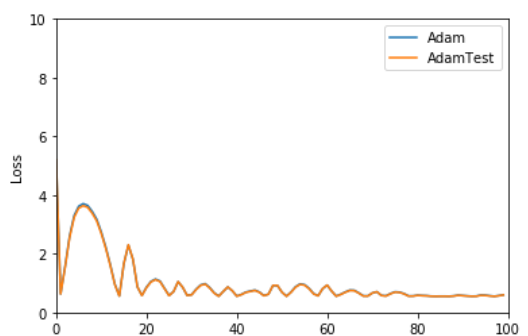
Adam:

## Hyper-parameter selection:beta1=0.9 beta2=0.99,alpha=0.1

## Predicted Results (Best Results):



## Loss curve:



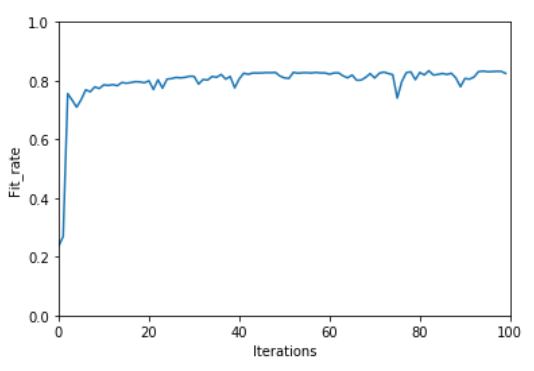
**Linear Classification and Stochastic Gradient Descent**

Batch size =256

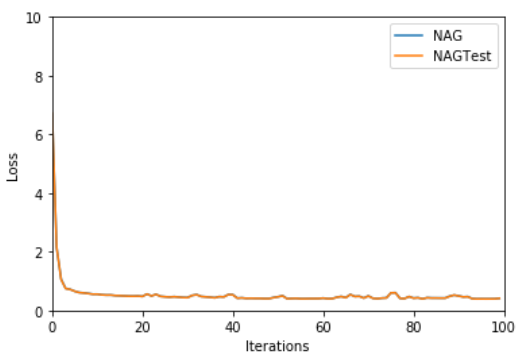
NAG:

## Hyper-parameter selection:momentum=0.99,alpha=0.01

## Predicted Results (Best Results):



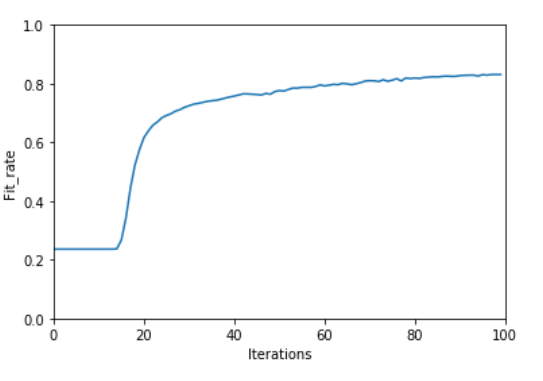
## Loss curve:



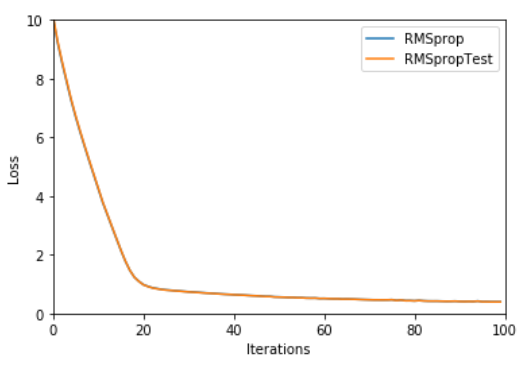
RMSProp:

## Hyper-parameter selection:decay\_rate=0.9,alpha=0.2

## Predicted Results (Best Results):



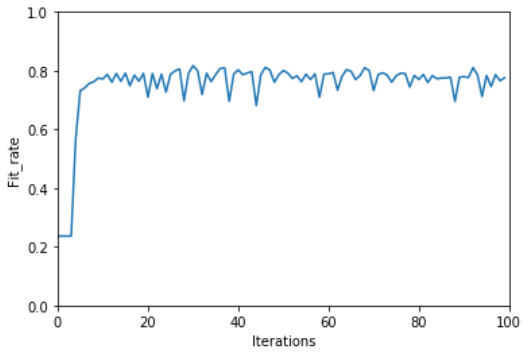
## Loss curve:



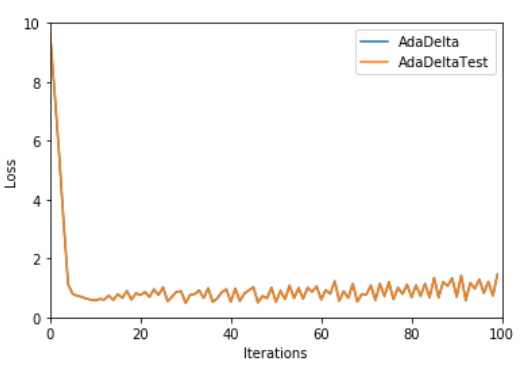
AdaDelta:

## Hyper-parameter selection:delta=0.5,alpha=2000

## Predicted Results (Best Results):



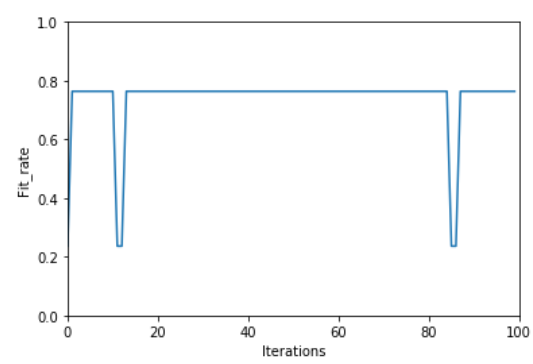
## Loss curve:



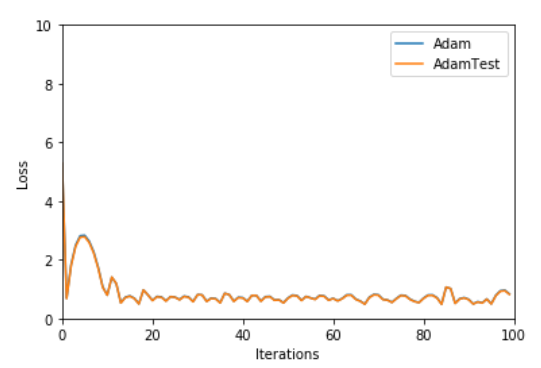
Adam:

## Hyper-parameter selection:beta1=0.9 beta2=0.99,,alpha=0.1

## Predicted Results (Best Results):

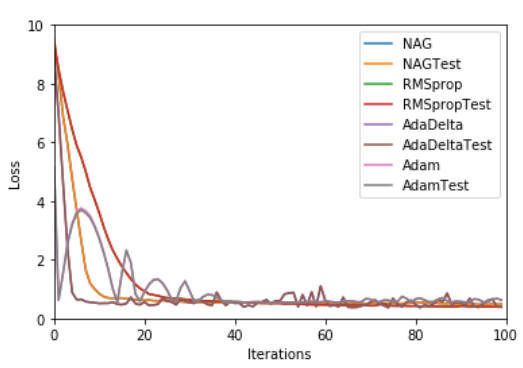


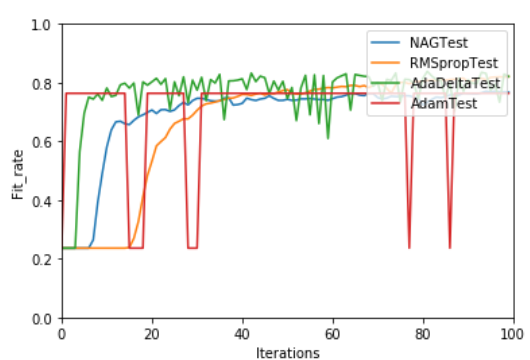
## Loss curve:



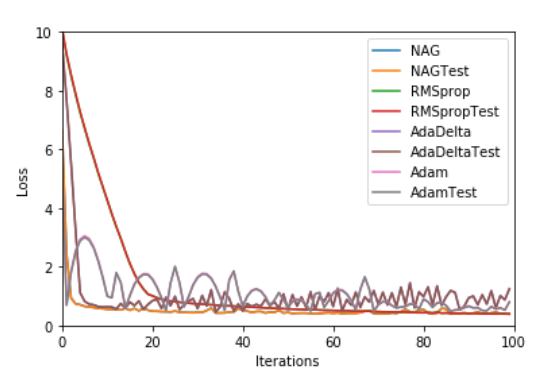
1. **Results analysis:**

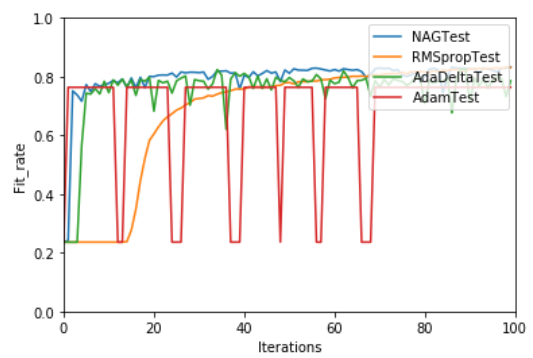
**Logistic Regression and Stochastic Gradient Descent**

****

****

**Linear Classification and Stochastic Gradient Descent**

****

****

The loss of four optimized methods all come to less than 2 and become stable after 40 iterations. Among these optimized methods, RMSProp is the slowest but the smoothest. And though Adam is the fastest that arrives the bottom, it shows the biggest undulation.

The fit rate of four optimized methods will eventually get to about 80%.Adam is the most unstable. RMSprop performs the best among these methods.AdaDelta also shows some undulation.

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

Simmilarities:logistic regression and linear classification both need a threshold to divide datas into multiple classes.

Differences:the loss function of logistic regression is logistic loss and the loss function of linear classification is Hinge loss,which means logistic regression’s target is to minimize logistic loss and linear classification’s target is to maximize the space between the classification planes and datas.

1. **Summary:**

In this experiment, I have understood the difference between gradient descent and stochastic gradient descent, that is SGD uses mini batch to reduce calculation and speed up algorithm.I have compared and understood the differences and relationships between Logistic regression and linear classification, and the principles of SVM and practice on larger data with SGD to reduce cost of time .