



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:  
Jing Liang

Supervisor:  
Mingkui Tan or Qingyao Wu

Student ID:  
201530741368

Grade:  
Undergraduate

December 13, 2017

# Logistic Regression, Linear Classification and Stochastic Gradient Descent

**Abstract—implication of logistic regression and linear classification with Stochastic Gradient Descent.**

## I. INTRODUCTION

In this experiment, we will do:

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.

Use different methods to optimize classification problems.

## II. METHODS AND THEORY

Logistic Regression:

Thought it calls Regression, it outputs possibility and these possibilities are used for classification problems.

Usually, we use logic loss function to estimate its loss:

The logistic function:

$$h_{\theta}(x) = g(\theta^T x) = \frac{1}{1 + e^{-\theta^T x}}$$

The loss function is:

$$\text{cost}(h_{\theta}(x), y) = -y_i \log(h_{\theta}(x)) - (1 - y_i) \log(1 - h_{\theta}(x))$$

and the derivatives is:

$$\text{gradient} = (h_{\theta}(x_i) - y_i) x_i$$

Linear Classification:

It is based on support vector machine. Its target is to minimize the support vector.

We usually use hinge loss to estimate the cost:

$$L(y) = C * \sum \max(0, 1 - y * w x)$$

the derivatives is:

$$-y_i x_i \quad 1 - y_i * w^T x_i \geq 0$$

$$0 \quad 1 - y_i * w^T x_i < 0$$

## III. EXPERIMENT

### A. Dataset

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

### B. Implication

Regression parameters: TABEL I

batch size	30
epsilon	1000

NAG:

$\eta$	0.003
$\gamma$	0.99

RMSProp:

$\eta$	0.003
$\epsilon$	$10^{-8}$
$\gamma$	0.9

AdaDelta:

$\epsilon$	$10^{-8}$
$\gamma$	0.95

Adam:

$\eta$	0.002
$\epsilon$	$10^{-8}$
$\beta_1$	0.9
$\beta_2$	0.99

Classification parameters: TableII

batch size	30
epsilon	250
C	0.9

NAG:

$\eta$	0.001
$\gamma$	0.8

RMSProp:

$\eta$	0.001
$\epsilon$	$10^{-8}$
$\gamma$	0.94

AdaDelta:

$\epsilon$	$10^{-6}$
$\gamma$	0.97

Adam:

$\eta$	0.0015
$\epsilon$	$10^{-6}$
$\beta_1$	0.9
$\beta_2$	0.999

Experimental steps:

Logistic Regression and Stochastic Gradient Descent

1. Load the training set and validation set.
2. Initialize logistic regression model parameters, you can consider initializing 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.
7. Repeat step 4 to 6 for several times, and drawing graph of losses with the number of iterations.

Linear Classification and Stochastic Gradient Descent

1. Load the training set and validation set.
2. Initialize SVM model parameters, you can consider initializing 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.
7. Repeat step 4 to 6 for several times, and drawing graph of losses with the number of iterations.

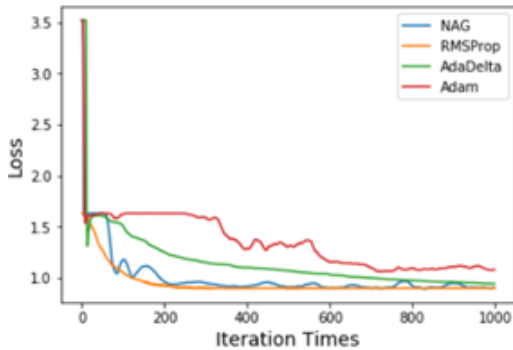


Figure I: Regression Results

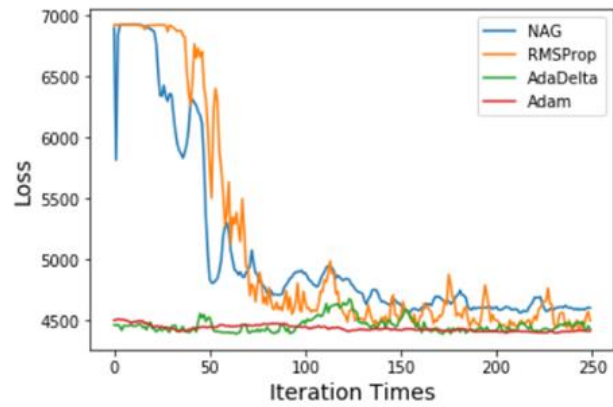


Figure II: Classification Results

#### IV. CONCLUSION

Stochastic Gradient Descent is a good way to optimize the parameters in our model for big data sets, each time, it use only one sample to update the gradient and weights. However, it cannot gain a good result as the batch gradient descent. To optimize this, we apply four algorithm to SGD. They usually use the history gradient information or make a prediction to the gradient or the learning rate to optimize the gradient and learning rate.

Logistic regression is a classical classification model, it classifies samples by output the possibility with its non-linear sigmoid function.

Similarities and differences between logistic regression and linear classification:

They both can solve the classification problem. They both can output continuous values.

However, the linear classification uses a linear model, but the logistic model is a non-linear model. The linear classification use 0 as threshold, by deciding the points below the line or not. The logistic model just output the possibility, so we usually select 0.5 as threshold.