

**The Experiment Report of**

***Machine Learning***

**College Software College**

**Subject Software Engineering**

**Members**   **Canguang Li**

**Student ID 201530611937**

**E-mail 1065561759@qq.com**

**Tutor**   **Mingkui Tang**

**Date submitted** **2017.12 .8**

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

**2. Time:** 2017.12.02

**3. Reporter:** 李灿光

**4. Purposes:**

1). Further understand of linear regression and gradient descent.

2). Conduct some experiments under small scale dataset.

3). Realize the process of optimization and adjusting parameters.

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

Linear Regression uses Housing(scaled edition) in LIBSVM Data, including 506 samples and each sample has 13 features.

Linear classification uses australian(scaled edition) in LIBSVM Data, including 690 samples and each sample has 14 features.

**6. Experimental steps:**

1).Linear Regression and Gradient Descent

①Load the experiment data. You can use load\_svmlight\_file function in sklearn library.

②Devide dataset. You should divide dataset into training set and validation set using train\_test\_split function. Test set is not required in this experiment.

③Initialize linear model parameters. You can choose to set all parameter into zero, initialize it randomly or with normal distribution.

④Choose loss function and derivation: Find more detail in PPT.

⑤Calculate gradient G toward loss function from all samples.

⑥Denote the opposite direction of gradient G as D.

⑦Update model: .  is learning rate, a hyper-parameter that we can adjust.

⑧Get the loss  under the training set and  by validating under validation set.

⑨Repeate step 5 to 8 for several times, and drawing graph of  as well as  with the number of iterations.

2). Linear Classification and Gradient Descent

①Load the experiment data.

②Divide dataset into training set and validation set.

③Initialize SVM model parameters. You can choose to set all parameter into zero, initialize it randomly or with normal distribution.

④Choose loss function and derivation: Find more detail in PPT.

⑤Calculate gradient G toward loss function from all samples.

⑥Denote the opposite direction of gradient G as D.

⑦Update model: .  is learning rate, a hyper-parameter that we can adjust.

⑧Select the appropriate threshold, mark the sample whose predict scores greater than the threshold as positive, on the contrary as negative. Get the loss  under the trainin set and  by validating under validation set.

⑨Repeate step 5 to 8 for several times, and drawing graph of  as well as  with the number of iterations.

**7. Code:**

1). Linear Regression and Gradient Descent

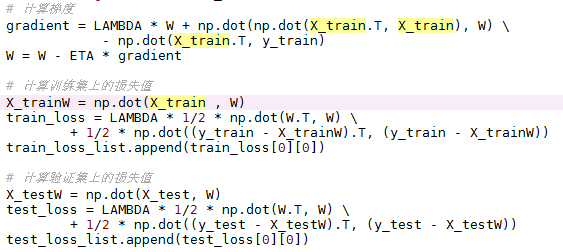


figure 1 code of linear regression

2). Linear Classification and Gradient Descent

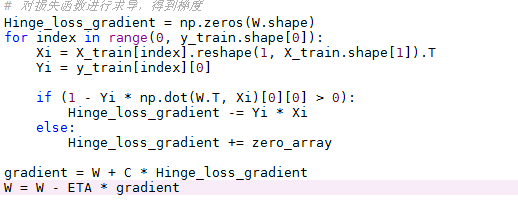


figure 2 code of linear classification

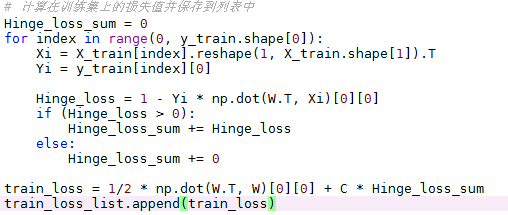


figure 3 code of linear classification

**8. Selection of validation (hold-out, cross-validation, k-folds cross-validation, etc.):**

Hold-out- a quarter of the samples were randomly divided from the data set as the verification set, and the rest as the training set.

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

1). Linear Regression and Gradient Descent

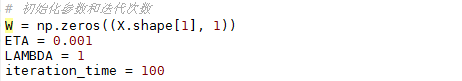


figure 4

2). Linear Classification and Gradient Descent

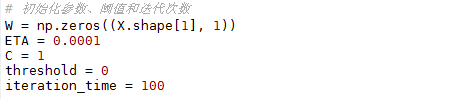


figure 5

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

1). Linear Regression and Gradient Descent

Loss function：

Gradient：

2). Linear Classification and Gradient Descent

Loss function：

Denote ,

,

then the gradient :

**11. Experimental results and curve:**

## Hyper-parameter selection (η, epoch, etc.):

1). Linear Regression and Gradient Descent

iteration\_time = 100

2). Linear Classification and Gradient Descent

C = 1

iteration\_time = 100

threshold = 0

## Assessment Results (based on selected validation):

1). Linear Regression and Gradient Descent

The minimum of loss under the training set is 4012.0963673122073.

The minimum of loss under the validation set is 2058.3742636076668

2). Linear Classification and Gradient Descent

The minimum of loss under the training set is 160.0298752798023

The minimum of loss under the validation set is 45.982449217649219

## Predicted Results (Best Results):

1). Linear Regression and Gradient Descent

The minimum of loss under the validation set is 2058.37426360

2). Linear Classification and Gradient Descent

The highest right rate under the validation set is 0.9017341040462428

## Loss curve:

1). Linear Regression and Gradient Descent

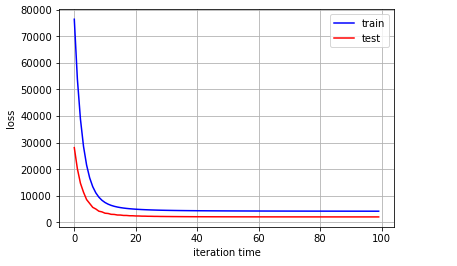


figure 6 Linear Regression

2). Linear Classification and Gradient Descent

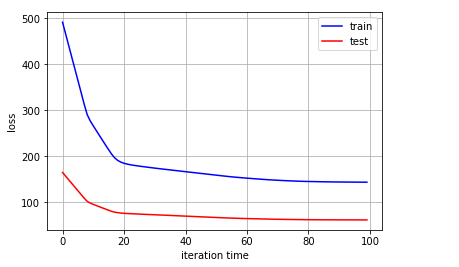


figure 7 Linear Classification

**12. Results analysis:**

1). Linear Regression and Gradient Descent

I assigned 0.1, 0.01, 0.001, 0.0001 to the learning rate respectively, and found that when or 0.01, the curve would never converge. When , the curve would converge. When , the curve converged slowly.

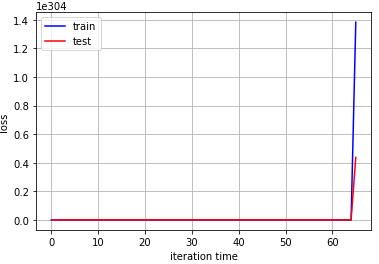


figure 8

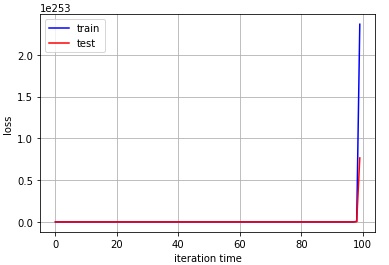


figure 9

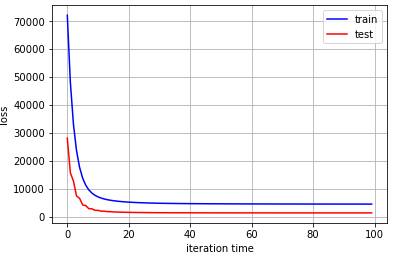


figure 10

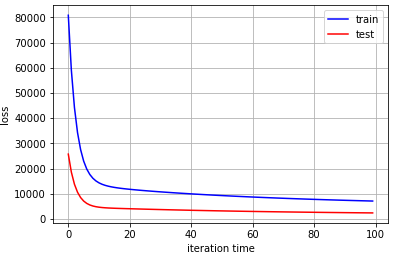


figure 11

2). Linear Classification and Gradient Descent

I assigned 0.1, 0.01, 0.001, 0.0001, 0.00001 to the learning rate respectively, and found that when , the curve would oscillate seriously. When , the curve would oscillate, too. When , the curve started to converge.When , the curve converged well. When, the curve converged slowly.

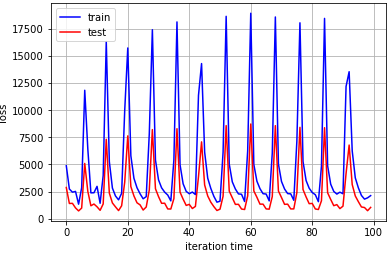


figure 12

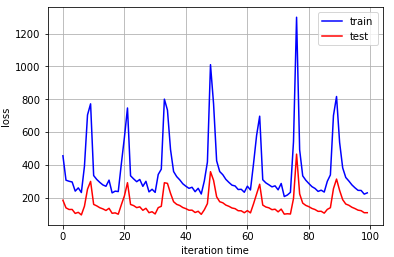


figure 13

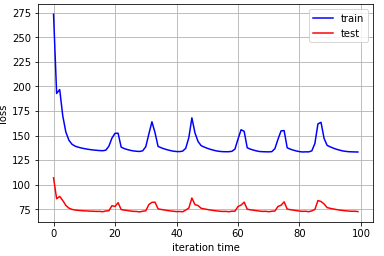


figure 14

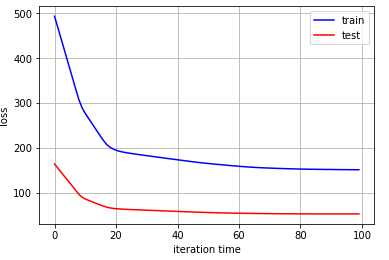


figure 15

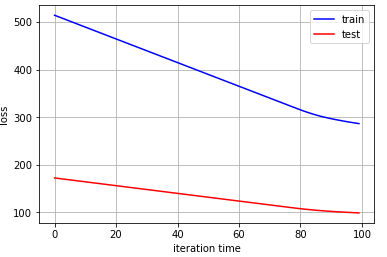


figure 16

**13. Similarities and differences between linear regression and linear classification:**

Similarities: The linear functions used for linear regression and linear classification are the same.

The difference is: the way to predict is not the same - linear regression put the sample into the linear function of the calculation, the results obtained as a predictive value; linear classification need to set a threshold value, the sample is put into the linear function, the result is compared with the threshold, according to the comparison classifies the sample as positive or negative.

**14. Summary:**

By implementing linear regression and linear classification algorithms by myself, I understand the principle of matrix derivation and the gradient descent more. At the same time through the practice of adjusting parameters, I also received some experience to adjusting parameters . In addition, I also understand the similarities and differences between linear classification and linear regression.