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:
Fangyuan Ding

Supervisor:
Qingyao Wu

Student ID：201720144979

Grade:
Graduate

December 15, 2017

Linear Regression, Linear Classification and Gradient Descent

Abstract—
Further understand of linear regression and gradient descent. Realize and understand the process of optimization and adjusting parameters.

I. INTRODUCTION

In the experiment, we use linear regression model for the regression problem and linear classification model to solve the classification problem.

II. METHODS AND THEORY

A. Linear Regression and Gradient Descent

1) The loss function of linear regression

$$\text{Loss} = \frac{1}{2}(y - Xw)^T(y - Xw)$$

2) Derivatives

$$\frac{\partial Loss}{\partial w} = -X^T y + X^T X w$$

B. Linear Classfication and Gradient Descent

1)The loss function of linear classification

$$\text{Loss} = \frac{\|w\|^2}{2} + \frac{C}{n}\sum_{i=1}^{n} max(0, 1 - y_i(x_i w + b))$$

2) Derivatives

$$g_t = \begin{cases} w + C\sum_{i=1}^{n} -x_i^T y_i & 1 - y_i(x_i w + b) \geq 0 \\ w & 1 - y_i(x_i w + b) < 0 \end{cases}$$

III. EXPERIMENT

A. Dataset

In this experiment,Linear Regression uses Housing in LIBSVM Data, including 506 samples and each sample has 13 features. and linear classification uses australian in LIBSVM Data, including 690 samples and each sample has 14 features.

B. Experimental steps

1)Linear Regression and Gradient Descent

1.Load the experiment data. You can use load_svmlight_file function in sklearn library.

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

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

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

5.Calculate gradient G toward loss function from all samples.

6.Denote the opposite direction of gradient G as D.

7.Update model: $W_t = W_{t-1} + \eta D$. η is learning rate, a hyper-parameter that we can adjust.

8.Get the loss L_{train} under the training set and $L_{validation}$ by validating under validation set.

9.Repeate step 5 to 8 for several times, and drawing graph of L_{train} as well as $L_{validation}$ with the number of iterations.

2) Linear Classification and Gradient Descent

1.Load the experiment data.

2.Divide dataset into training set and validation set.

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

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

5.Calculate gradient G toward loss function from all samples.

6.Denote the opposite direction of gradient G as D.

7.Update model: $W_t = W_{t-1} + \eta D$. η is learning rate, a hyper-parameter that we can adjust.

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

9.Repeate step 5 to 8 for several times, and drawing graph of L_{train} as well as $L_{validation}$ with the number of iterations.

C .initialization model parameters

1)Linear Regression and Gradient Descent

All parameters are set into zero in the linear regression model.

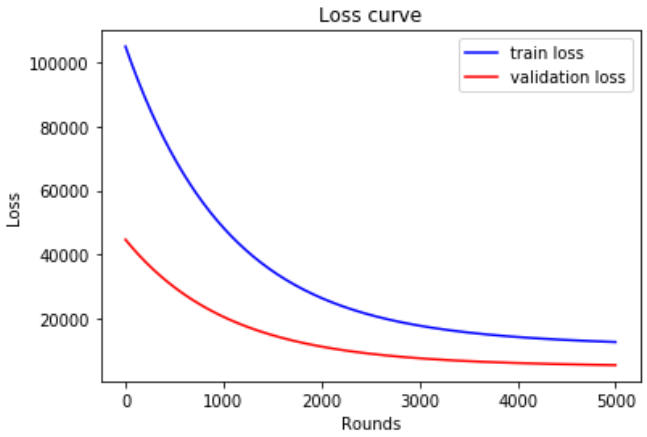
2)Linear Classification and Gradient Descent

All parameters are set into zero in the SVM model.

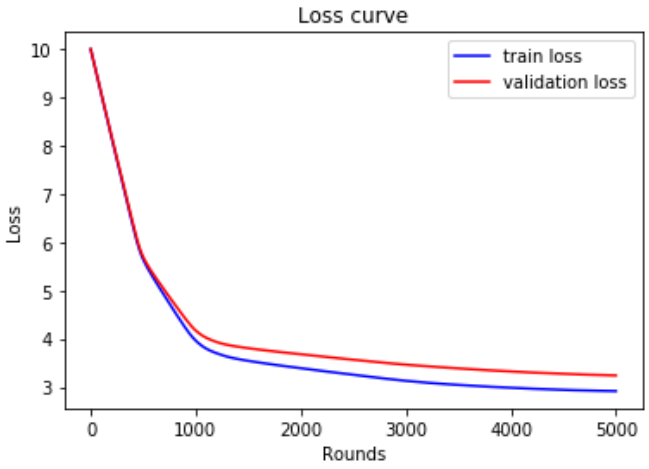
D. Experimental results and curve

In this Experiment, hyper-parameter were selected as follows: the learning rate is 0.001 and the iterations is 5000.

1)Linear Regression Loss:



2)Linear Classification Loss:



In the linear regression experiment, after thousands of iterations, the L_{train} and the $L_{\text{validation}}$ both descend, and they are relatively close. and in the linear classification experiment, the L_{train} and the $L_{\text{validation}}$ both descend, and will go to different convergence at the end.

IV. CONCLUSION

Finally, I have a further understand of linear regression and gradient descent .The linear regression is to be used to solved the regression problem, and linear classification is aimed to solved the classification problem.Moreover, the linear classification is need to select a threshold, but the other needn't. The division of the data set and parameter adjustment have aeffects on the final result.