South China University of Technology

The Experiment Report of Machine Learning

SCHOOL: SCHOOL OF SOFTWARE ENGINEERING

SUBJECT: SOFTWARE ENGINEERING

Author:
DingKun Song

Supervisor:
Mingkui Tan or Qingyao Wu

Student ID：
201720145006

Grade:
Undergraduate

December 12, 2017

Linear Regression, Linear Classification and Gradient Descent

Abstract—The experiment is to use Linear Regression to analysis Housing Data and use Linear Classification to analysis Australian Data.

I. INTRODUCTION

A. Linear Regression

Linear regression is a statistical analysis method using regression analysis in mathematical statistics to determine the quantitative relationship between two or more variables and more variables interdependent.

For example,given a dataset as:

$$D = \{(\mathbf{x}_i, y_i)\}_{i=1}^{n}$$

We should learn the "best" model f(x,w) to fix D.And this is call Linear regression.

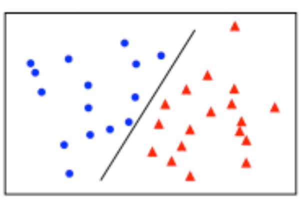
B. Linear Classifi cation

Linear classification uses a linear function to partition the data set.

Given training data (xi,yi) for i=1...n,with xi R^m and yi $\{ \; 1,1\}$,learn a classfier f(x)such that

$$f(\mathbf{x}_i) \begin{cases} \geq 0 & y_i = +1 \\ < 0 & y_i = -1 \end{cases}$$

A linear classifier function can be $f(x) = w^T x + b$ .We can see the picture,the f(x) we learn can seperate the dataset into two cluster.



C. Gradient Descent

Identify a set of hypotheses f(x;w) and define a loss function l(w),then we pick the best w^* by minimizing the loss function.This is call Gradient Decent.

D. Experiment

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

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

II. METHODS AND THEORY

A. *Linear Regression And Gradient Descent*

First,defined the model function as

$$f(x) = w_0 + w_1x_1 + w_2x_2 + ... + w_m x_m = \mathrm{w}^T X$$

Second,find the loss function.

$$\mathrm{L}(w) = \frac{1}{2n} \; {}_{i=1}^{n} (y_i \quad w^T x_i)^2 = \frac{1}{2}(y \quad Xw)^T (y \quad Xw)$$

Third,minimizing the loss function use Gradient Descent.

$$\frac{\mathrm{L}(w)}{w} = \frac{1}{n} X^T (Xw \quad y)$$

$$\mathrm{w}_{\mathrm{t}} = w_{t \;\; 1} \qquad \frac{L(w)}{w}$$

We use these two formula to update w and find the best w to minimize loss function.

B. Linear Classification And Gradient Descent

First,defined the model function as

$$f(x) = w_0 + w_1x_1 + w_2x_2 + ... + w_m x_m = \mathrm{w}^T X$$

Second,find the loss function.

$$L(w) = \frac{\|\mathrm{w}\|^2}{2} + \frac{\mathrm{C}}{n} \; {}_{i=1}^{n} \max(0,1 \quad y_i(w^T x_i))$$

Third,minimizing the loss function use Gradient Descent.

$$g_{\mathbf{w}}(\mathbf{x}_i) = \begin{cases} -y_i \mathbf{x}_i & 1 - y_i(\mathbf{w}^\top \mathbf{x}_i + b) >= 0 \\ 0 & 1 - y_i(\mathbf{w}^\top \mathbf{x}_i + b) < 0 \end{cases}$$

$$\frac{\mathrm{L}(w)}{w} = \mathrm{w} + \frac{C}{n} \; {}_{i=1}^{n} g_w(x_i)$$

$$\mathrm{w}_{\mathrm{t}} = w_{t \;\; 1} \qquad \frac{L(w)}{w}$$

We use these two formula to update w and find the best w to minimize loss function.

III. Experiment

A. Linear Regression And Gradient Descent

A. Dataset

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

I get 2/3 of data as training set and 1/3 of data as validation set.

B Implementation

The Initialization value is showned ad the table.

Rate	0.01
Traning time	1000

Then I use the formula above to caculate loss function and update the w.

I use two array,loss_train and loss_test, to save the loss of training set and validation set.

Finally,I use matplotlib to show the loss_train and loss_test.

B. Linear Classification And Gradient Descent

A. Dataset

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

I get 2/3 of data as training set and 1/3 of data as validation set.

B. Implement

The Initialization value is showned ad the table.

Rate	0.01
C	0.1
Traning time	1000

Then I use the formula above to caculate loss function and update the w.

I use two array,loss_train and loss_test, to save the loss of training set and validation set.

Finally,I use matplotlib to show the loss_train and loss_test.

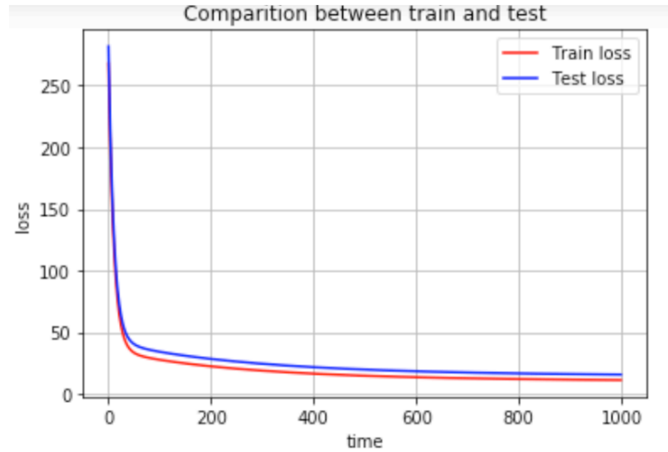
IV. Conclusion

A, Linear Regression And Gradient Descent

With the appeals method, as the number of training increases, the loss is getting smaller and smaller, finally tends to be smooth.

This is mean we find the best w.

The result is showned as follow.



B. *Linear Classification And Gradient Descent*

With the appeals method, as the number of training increases, the loss is getting smaller and smaller, finally tends to be smooth.

This is mean we find the best w.

The result is showned as follow.