South China University of Technology

The Experiment Report of Machine Learning

SCHOOL: SCHOOL OF SOFTWARE ENGINEERING

SUBJECT: SOFTWARE ENGINEERING

Author:
Zhaohui Huang

Supervisor:
Qingyao Wu

Student ID：201720145112

Grade: Graduate

December 15, 2017

Linear Regression, Linear Classification and Gradient Descent

Abstract—Linear regression is perhaps one of the most well known and well understood algorithms in statistics and machine learning. The main purposes of this report is further understand of linear regression and gradient descent and learn how to conduct some experiments under small scale dataset. Meanwhile , we have realized the process of optimization and adjusting parameters.

I. INTRODUCTION

Linear regression is a very simple approach for supervised learning. Though it may seem somewhat dull compared to some of the more modern algorithms, linear regression is still a useful and widely used statistical learning method. Linear regression is used to predict a quantitative response Y from the predictor variable X.

The main purposes of this report can be concluded as the following:

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.

II. METHODS AND THEORY

Linear regression

Learn f(x;w) with

Parameters:w $\in R^m$,,$w_0 \in R$

Input:x where $x_j \in R$ for j $\in$ 1,...m features

Model Function:

f(x;w 0 ,w) = $w_0 + w_1 x_1 + ... + w_m x_m$

= $w^T x + w_0$

The loss function of linear regression that we use is:

$$L = \frac{1}{2N}\sum_{i=1}^{n}(w^T x_i - y_i)^2$$

Gradient:

$$\frac{\partial L}{\partial w} = X^T(Xw - y)$$

Linear classification

The loss function of linear classofication that we use is:

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

Gradient:

$$sign_i = \begin{cases} 0 & 1 - y_i(w^T x_i + b) \le 0 \\ 1 & 1 - y_i(w^T x_i + b) > 0 \end{cases}$$

$$\frac{\partial L}{\partial w} = w + \mathrm{C}X^T(y \odot sign)$$

III. EXPERIMENT

Dataset

Linear Regression uses Housing in LIBSVM Data, including 506 samples and each sample has 13 features. You are expected to download scaled edition. After downloading, you are supposed to divide it into training set, validation set.

Linear classification uses australian in LIBSVM Data, including 690 samples and each sample has 14 features. You are expected to download scaled edition. After downloading, you are supposed to divide it into training set, validation set.

Experiment Step

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 toward loss function from all samples.

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

7.Update model: $.\omega_t = \omega_{t-1} + \eta D$.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.

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 toward loss function from all samples.

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

7.Update model: $.\omega_t = \omega_{t-1} + \eta D$.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.

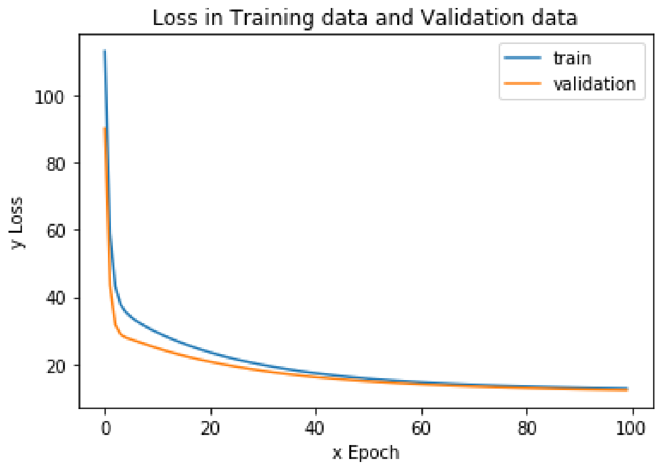
9.Repeate step 5 to 8 for several times, and drawing graph of L_{train} as well as $L_{validation}$ with the number of iterations.

Results

Linear regression

The chosen parameters are: $\eta = 0.1$, epoch = 100

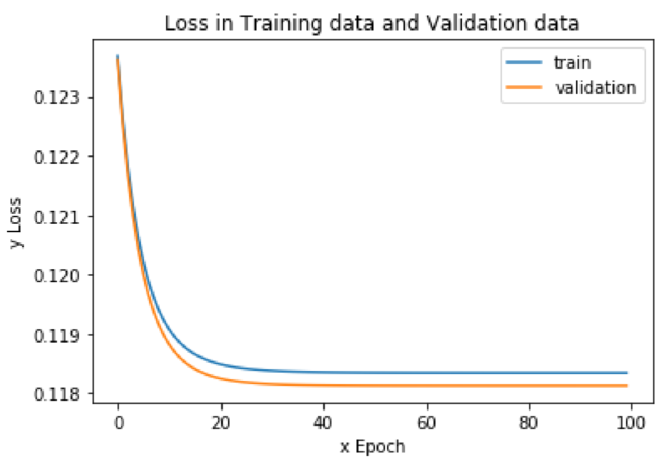
The graph of L_{train} as well as $L_{validation}$ with the number of iterations.



Linear classification

The chosen parameters are:$\eta = 0.1$, $C = 2^{-1}$, epoch = 100

The graph of L_{train} as well as $L_{validation}$ with the number of iterations.



IV. CONCLUSION

Linear regression and linear classification are both related to prediction. Regression involves estimating or predicting a response and the output variable takes continuous values. Classification is identifying group membership and the output variable takes class labels.