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:

Zhenyi Tang

Student ID：201630503026

Supervisor:

Qingyao Wu

Grade: Undergraduate

October 25, 2018

Linear Regression and Stochastic Gradient Descent

Abstract— By using closed-form solution and Stochastic Gradient Descent to work out the linear regression problem so that to further understand these two methods and realize the process of optimization and adjusting parameters.

I. INTRODUCTION

The experiment will first initialize linear model parameters and set all parameter into zero, then select the following loss function and calculate the value of the loss function of the training set denoted as Loss, then get the value of parameter W by the closed-form solution, and update the parameter W and get the Loss, loss_train under the training set, loss_val by validating under validation set.

The second part of the experiment will use the Stochastic Gradient Descent method to find the best W. Choosing the same loss function and derivation, then calculate G toward loss function from each sample. Denote the opposite direction of G as D and update model to find the best W.

II. METHODS AND THEORY

Lab 1.1 compare two different methods to calculate the loss of the dataset. The closed-form solution is a method which calculate the derivation and let the it equal to 0, however, this method needs to compute the inverse of the matrix which need huge memory. The loss function formula is showing following as

$$J(\theta) = \frac{1}{2m} \sum_{i=0}^{m} (h(x^i) - y^i)^2 \qquad (1.1)$$

Then vectorization this function and we get the

$$\mathcal{L}_D(\mathbf{w}) = \frac{1}{2} \sum_{i=1}^{n} (y_i - \mathbf{w}^\top \mathbf{x}_i)^2 = \frac{1}{2} (\mathbf{y} - \mathbf{X}\mathbf{w})^\top (\mathbf{y} - \mathbf{X}\mathbf{w}) \quad (1.2)$$

then calculate the derivation equal to zero then get the closed -form solution formula as following

$$\mathbf{w}^* = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{y} = \arg\min_{\mathbf{w}} \mathcal{L}_D(\mathbf{w}) \quad (1.3)$$

Lab 1.2 Stochastic Gradient Descent to update the W. This solution can not get a better loss than the closed-form solution however it is quicker that the former one and the output is not bad.

$$\mathbf{w}' \to \mathbf{w} - \eta \frac{\partial \mathcal{L}_D(\mathbf{w})}{\partial \mathbf{w}} \quad (1.4)$$

Finally, compute the G to update that right part of the formula below

$$G(\theta_j) = \frac{1}{m} \sum_{i=0}^{m} ((h(x^i) - y^i) * x_j^i) \quad (1.5)$$

III. EXPERIMENT

A. Dataset

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

B. Implementation

Using the load_svmlight_file function to load the dataset and using train_test_split function to divide dataset into training set and validation set then Initialize linear model parameters and Initialize linear model parameters, then using the Loss function (1.1) to calculate the value of the Loss function of the training set, denoted as Loss.

The second part of this experiment is to get the value of parameter W by the closed-form solution, and update the parameter W. Define the loss function as (1.2) then calculate the derivation equal to zero then get the closed -form solution formula as following (1.3)

The third part is to using SGD to find a W. First add one column into the matrix x_train and transform it into X_train then use the numpy lib to compute the gradient descent of the training dataset and use the formula 1.4 to find the best gradient descent by repeat the process and try to find a good learning rate to make the curve of the loss be more smooth in a epoch which is not so large that it need so much time to run out the answer

TABLE I
The experimental result of using closed-form solution

	value
Loss	300.024
Loss_train	10.589
Loss_val	12.971

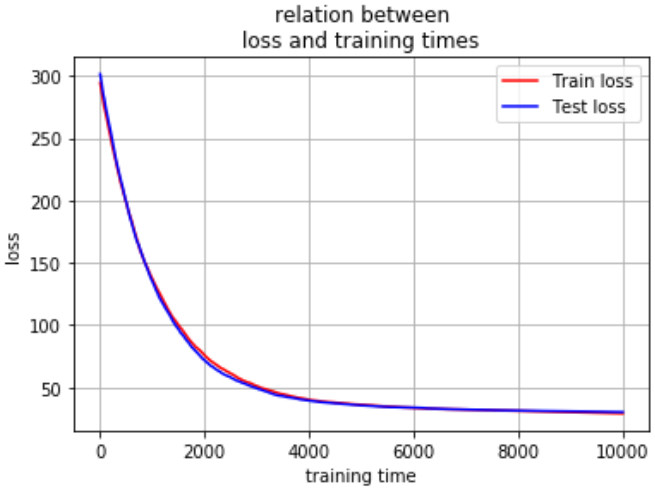

Fig. 1. results of lab1.2 using SGD to find the w

TABLE II
The result of using Stochastic Gradient Descent

	value
Loss_train	29.110
Loss_val	30.230

IV. CONCLUSION

From the result of these two experiments we can know that the final loss of the closed-form solution is better than the loss of the SGD solution. However, it is obvious that the closed-form solution takes a much more time. Because the inverse takes O(m3) complexity to compute. If we set a large learning rate at first it just take so little time to get a not bad W.