The Experiment Report of *Machine Learning*

SCHOOL: SCHOOL OF SOFTWARE ENGINEERING

SUBJECT: SOFTWARE ENGINEERING

Author:
Anran Lin

Supervisor:
Qingyao Wu

Student ID:
201630665045

Grade:
Undergraduate

October 27, 2018

Linear Regression and Stochastic Gradient Descent

***Abstract*—This experience is about linear regression and stochastic gradient descent.The main purpose is to have further understand of linear regression ,closed-form solution and Stochastic gradient descent by conducting some experiments under small scale dataset.**

I. Introduction

In statistics, linear regression is a linear approach to modelling the relationship between a scalar response (or dependent variable) and one or more explanatory variables (or independent variables).Stochastic gradient descent is an iterative method for optimizing a differentiable objective function, a stochastic approximation of gradient descent optimization.In this experiment,I would like to use stochastic gradient decent method to solve a linear regression problem for further understand of linear regression and gradient descent.My aim is to use the housing-scale dataset to train and test and try to make the loss as low as possible.

II. Methods and Theory

A. Linear Regression

Hypothesis: h($\mathbf{x}$;ω) with

$$parameters : \omega \in \{R\}, \omega_0 \in \mathbb{R}$$
$$input : x\ where\ x_j \in \mathbb{R}\ for\ j \in 1, ...(m-1)\ features$$

Model Function:

$$\begin{aligned} h(\mathbf{x}; \omega_0, \omega) &= \omega_0 + \omega_1 x_1 + ... + \omega_{m-1} x_{m-1} \\ &= \sum_{j=1}^{m-1} \omega_j x_j + \omega_0 \\ &= \omega^T x + \omega_0 \end{aligned} \tag{1}$$

B. Loss Function

Squared loss

$$\begin{aligned} L_D(\omega) &= \frac{1}{2} \sum_{i=1}^{n} (y_i - h(x_i; \omega))^2 \\ &= \frac{1}{2} \sum_{i=1}^{n} (y_i - \widehat{y_i})^2. \end{aligned} \tag{2}$$

Training:find minimizer of this loss(least squares)

$$\omega^* = arg \min_{\omega} L_D(\omega)$$

C. Linear Regression and Stochastic Gradient Descent

To minimize the loss function L(ω) use the iterative update

$$\omega_{t+1} \leftarrow \omega_t - \eta_t \frac{\partial L(\omega_t)}{\partial \omega_t} \tag{3}$$

where η is the learning rate and $\frac{\partial L(\omega)}{\partial \omega}$ is:

$$\begin{aligned} \frac{\partial L(\omega)}{\partial \omega} &= -\sum_{i-1}^{n} (y_i - \omega^T x_i) xi \\ &= -\sum_{i-1}^{n} y_i x_i + (\sum_{i-1}^{n} x_i xi^T) \omega \end{aligned} \tag{4}$$

D. closed-form solution of Linear Regression

Set equation(1) to 0,and solve for optimal parameters $\omega^(*)$

$$\omega^* = (\mathbf{X}^{\mathrm{T}} \mathbf{X})^{-1} \mathbf{X}^{\mathrm{T}} y = arg \min_{\omega} L_D(\omega)$$

III. Experiments

A. Dataset

The data set used in this experiment is housing-scale,provided by LIBSVM Data,including 506 samples and each sample has 13 features.I divided it into training set and validation set equally.

B. Implementation

1) closed-form solution of Linear Regression:

a) initialization and parameters: First,load the experiment data using load_svmlight_file function in sklearn library.Then divide dataset using train_test_split function and I choose to initialize linear model parameters randomly.

b) process: Calculate the loss using loss function equation(2).Then get the value of parameter$\omega$by the closed-form solution, and update the parameterω.Finally,calculate the loss again under the training set and validation set.

c) result: It is show that the loss has been minimized using closed form solution successfully.See Table I.

TABLE I
LOSS USING CLOSED FORM SOLUTION

Loss	training set	validation set
initial Loss	219.67849370889007	231.59647054171768
closed-form	5.19398968642731e-19	0.02691910042021025

2) Regression and Stochastic Gradient Descent:

a) initialization and parameters: First,load the experiment data using load_svmlight_file function in sklearn library.Then divide dataset using train_test_split function and I choose to initialize linear model parameters randomly.

b) process: Calculate gradient G toward loss function from each sample,then denote the opposite direction of gradient G as D.Update model $\omega_t = \omega_{t-1} + \eta D$,according to equation(3).Also,I had a try to adjust $\eta$to get different result.Finally,calculate the loss again under the training set and validation set.

c) result: see Table II. and Fig.1. It is show that the loss has been minimized using gradient descent successfully.

TABLE II
LOSS USING GRADIENT DESCENT

Loss	training set	validation set
initial Loss	219.67849370889007	231.59647054171768
gradient descent	0.10066842602334752	0.21471788009638246

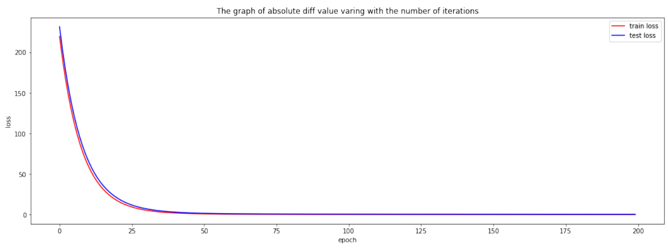

Fig. 1. Loss using gradient descent

IV. CONCLUSION

Through this experiment,I have experience solving a linear regression problem of a small scale dataset as well as the process of optimization and adjusting parameters.Also,I get the loss under closed-form solution and stochastic gradient descent and compare them(see TABLE III).It is apparent that closed-form solution for regression problem is precise but stochastic gradient descent is also a good method to handle the task.

TABLE III
COMPARISON

Loss	training set	validation set
initial Loss	219.67849370889007	231.59647054171768
closed-form	5.19398968642731e-19	0.02691910042021025
gradient descent	0.10066842602334752	0.21471788009638246