South China University of Technology

The Experiment Report of Machine Learning

SCHOOL: SCHOOL OF SOFTWARE ENGINEERING

SUBJECT: SOFTWARE ENGINEERING

Author:
Haiying Zhou

Supervisor:
Mingkui Tan

Student ID：
201630666547

Grade:
Undergraduate or Graduate

October 21, 2018

Logistic Regression and Support Vector Machine

Abstract—The experiment includes two parts:closed-form solution of Linear Regression and Linear Regression and Stochastic Gradient Descent

I. INTRODUCTION

This experiment intends to use Linear Regression and Stochastic Gradient Descent respectively for conducting some experiments under small scale data set and realizing the process of optimization and adjusting parameters.

II. METHODS AND THEORY

A. *closed-form solution of Linear Regression*

linear regression model:

$$f(\mathbf{x}) = \mathbf{w}^\top \mathbf{x} + b$$

Loss function:

Absolute value loss:

$$l(\hat{y}_i, y_i) = |\hat{y}_i - y_i|$$

Least squares loss:

$$l(\hat{y}_i, y_i) = \frac{1}{2}(\hat{y}_i - y_i)^2$$

Total loss(loss function):

$$\mathcal{L}_D(\mathbf{w}) = \sum_{i=1}^{n} l(\hat{y}_i, y_i)$$

Closed-form of linear regression:

$$\begin{aligned}
\mathcal{L}_D(\mathbf{w}) &= \frac{1}{2}(\mathbf{y} - \mathbf{X}\mathbf{w})^\top(\mathbf{y} - \mathbf{X}\mathbf{w}) \\
&= \frac{1}{2}(\mathbf{y}^\top\mathbf{y} - 2\mathbf{w}^\top\mathbf{X}^\top\mathbf{y} + \mathbf{w}^\top\mathbf{X}^\top\mathbf{X}\mathbf{w}) \\
\frac{\partial \mathcal{L}_D(\mathbf{w})}{\partial \mathbf{w}} &= \frac{1}{2}\left(\frac{\partial \mathbf{y}^\top\mathbf{y}}{\partial \mathbf{w}} - \frac{\partial 2\mathbf{w}^\top\mathbf{X}^\top\mathbf{y}}{\partial \mathbf{w}} + \frac{\partial \mathbf{w}^\top\mathbf{X}^\top\mathbf{X}\mathbf{w}}{\partial \mathbf{w}}\right) \\
&= \frac{1}{2}(-2\mathbf{X}^\top\mathbf{y} + (\mathbf{X}^\top\mathbf{X} + (\mathbf{X}^\top\mathbf{X})^\top)\mathbf{w}) \\
&= -\mathbf{X}^\top\mathbf{y} + \mathbf{X}^\top\mathbf{X}\mathbf{w}
\end{aligned}$$

B. *Linear Regression and Stochastic Gradient Descent*

Gradient (vector of partial derivatives):

$$\frac{\partial \mathcal{L}_D(\mathbf{w})}{\partial \mathbf{w}} = \begin{bmatrix} \frac{\partial \mathcal{L}_D(w_1)}{\partial w_1} \\ \frac{\partial \mathcal{L}_D(w_2)}{\partial w_2} \\ \vdots \\ \frac{\partial \mathcal{L}_D(w_n)}{\partial w_n} \end{bmatrix}$$

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

III. EXPERIENCE

A. *Data set*

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

B. *Implementation*

1) closed-form solution of Linear Regression

i) Load the experiment data and divide data set.
ii) Initialize linear model parameters.
iii) Select a Loss function and calculate the value of the Loss function of the training set,denoted as loss.
iv) Get the formula of the closed-form solution and get the value of parameter W by the closed-form solution, and update the parameter W.
v) Get the loss,loss_train under the training set and loss_val by validating under validation set and output the value of loss, loss_train and loss_val.

2) Linear Regression and Stochastic Gradient Descent

vi) Load the experiment data and divide data set.
vii) Initialize linear model parameters.
viii) Choose loss function and derivation.
ix) Calculate G toward loss function from each sample and denote the opposite direction of gradient G as D.
x) Update model and get the loss loss_train under the training set and loss_val by validating under validation set.

C. *Result*

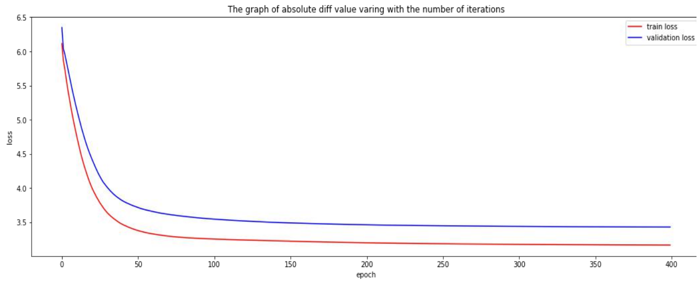
1) closed-form solution of Linear Regression

```
In [7]: w=numpy.dot(numpy.dot(numpy.linalg.inv(numpy.dot(x_train.transpose(),x_train)),x_train.transpose()),y_train)
        y_predict=numpy.dot(x_train,w)
        loss_train=numpy.average(numpy.abs(y_predict-y_train))
        y_predict=numpy.dot(x_val,w)
        loss_val=numpy.average(numpy.abs(y_predict-y_val))

        print(loss)
        print(loss_train)
        print(loss_val)

22.731398416886545
3.3655656257716524
3.0925756691108695
```

2) Linear Regression and Stochastic Gradient Descent



D. *Analysis*

From experience's result, we can find that the descent process of the loss of train and the loss of evaluation are similar, which means that the hyper-parameters chosen are suitable. In addition, the gradient descent is very smooth, which means that the descent process meets the requirements of this lab.

IV. CONCLUSION

It is my first time to write code to implement the basic methods of machine learning. Through this experiment, I Further understand of linear regression, closed-form solution and Stochastic gradient descent.What's more, the practice give me a chance to realize the process of optimization and adjusting parameters.