華南理工大學

South China University of Technology

The Experiment Report of *Machine Learning*

SCHOOL: SCHOOL OF SOFTWARE ENGINEERING

SUBJECT: SOFTWARE ENGINEERING

Author:
Biquan Wang

Supervisor:
Qingyao Wu

Student ID:
201821038853

Grade:
Graduate

January 1, 2019

Logistic Regression and Support Vector Machine

***Abstract*—Logistic regression and Support Vector Machine are simple and efficient classification models,both of them are widely used in production. In this report, we try to compare the result with different learn rate and find out the best.**

I. Introduction

LOGISTIC regression is a classification model in machine learning. Due to its simplicity and efficiency, it is widely used in production. Svm (Support Vector Machine) is a linear classifier, proposed by Cortes and Vapnik in 1995, and has been applied in the fields of handwriting recognition and text categorization, etc. This paper focuses on the mathematical model and parameter solving method of logistic regression algorithm and support vector machine algorithm. Futher more, we shows the implemention of tthese two algorithm and the result of them.

II. Methods and Theory

In this section, we will give a complete introduction to these experiments including the loss function and the method of updating parameter $\boldsymbol{w}$ of logistic regression, the loss function and the method of updating parameter $\boldsymbol{w}$ of linear classification.

A. Theory of logistic regression

Assume that the labels are binary: $y_i \in \{0, 1\}$

$$h_{\boldsymbol{w}}(\boldsymbol{x}) = g(\boldsymbol{w}^T\boldsymbol{x}) = \frac{1}{1+e^{-\boldsymbol{w}^T\boldsymbol{x}}}$$

Probability:

$$p = \begin{cases} h_w(x_i) & y_i = 1 \\ 1 - h_w(x_i) & y_i = 0 \end{cases}$$

$$\begin{aligned} max \prod_{i=1}^{n} P(y_i|\boldsymbol{x_i}) &\Leftrightarrow max\ log(\prod_{i=1}^{n} P(y_i|\boldsymbol{x_i})) \\ &\equiv max \sum_{i=1}^{n} logP(y_i|\boldsymbol{x_i}) \\ &\Leftrightarrow min - \frac{1}{n}\sum_{i=1}^{n} logP(y_i|\boldsymbol{x_i}) \end{aligned} \tag{1}$$

$$P(y_i|\boldsymbol{x_i}) = h_{\boldsymbol{w}}(\boldsymbol{x_i})^{y_i} \cdot (1 - h_{\boldsymbol{w}}(\boldsymbol{x_i}))^{(1-y_i)}$$

We can get the loss function:

$$J(\boldsymbol{w}) = -\frac{1}{n}\left[\sum_{i=1}^{n} y_i log h_{\boldsymbol{x}_i} + (1-y_i)log(1 - h_{\boldsymbol{w}}(\boldsymbol{x}_i))\right] \tag{2}$$

The gradient of the loss function:

$$\frac{\partial J(\boldsymbol{w})}{\partial \boldsymbol{w}} = \frac{1}{n}\sum_{i=1}^{n}(h_{\boldsymbol{w}}(\boldsymbol{x}_i) - y)\boldsymbol{x}_i \tag{3}$$

So, the update function of w is:

$$\boldsymbol{w} := \boldsymbol{w} - \frac{1}{n}\sum_{i=1}^{n}(h_{\boldsymbol{w}}(\boldsymbol{x}_i) - y)\boldsymbol{x}_i \tag{4}$$

B. Theory of linear regression

Select two parallel hyperplanes that separate the two classes of data and let the distance between them as large as possible, The region bounded by these two hyperplanes is called the **”margin”**.

$$L_i = \sum_{j \neq y_i} max(f_j - f_{y_i} + \Delta)$$

$$f = w * x$$

So:

$$L_i = \sum_{j \neq y_i} max(w_j^T x_i - w_{y_i}^T x_i + \Delta)$$

Average on the train set:

$$L = \frac{1}{N}\sum_{i=1}^{N}\sum_{j \neq y_i} max(0, w_j^T x_i - w_{y_i}^T x_i + \Delta)$$

Contain regularization:

$$\begin{aligned} L \quad = \quad & \frac{1}{N}\sum_{i=1}^{N}\sum_{j \neq y_i} max(0, w_j^T x_i - w_{y_i}^T x_i + \Delta) \\ & + \lambda \sum_k \sum_l W_{k,l}^2 \end{aligned} \tag{5}$$

Gradient/derivative:

$$\frac{\partial L_i}{\partial w_{y_i}} = -(\sum_{j \neq y_i} 1(w_j^T x_i - w_{y_i}^T x_i + \Delta > 0))x_i + 2\lambda w_{y_i} \tag{6}$$

$$\frac{\partial L_i}{\partial w_j} = 1(w_j^T x_i - w_{y_i}^T x_i + \delta > 0)x_i + 2\lambda w_j \tag{7}$$

III. Experiments

A. Dataset

Experiment uses a9a of LIBSVM Data, including 32561/16281(testing) samples and each sample has 123/123 (testing) features.

B. Step of Logistic regression

The steps of our Logistic regression are as the following:

1. Use load_svmlight_file function in sklearn library to load the a9a as train set a9a.t as valid set.
2. change the value of y from (1,-1) to (1,0).
3. Initialize logistic regression model parameters. Set all parameter into zero, initialize it randomly or with normal distribution.
4. Get a mini-batch data from train set randomly.
5. Update the parameters $\boldsymbol{w}$ with function (4)
6. Calculate the loss of train set and validation set with function (2).
7. Calculate the threshold loss of validation set, mark the sample whose predict scores greater than the threshold as positive, on the contrary as negative.
8. Repeate setp 4-7 for n times, and return the losses of train set, losses of validation set and the threshold losses.

C. Result of Logistic regression

In logistic regression, we discuss the impact with different value of learning rate, other prameter is same: interaction:1000,threshold:0.5, and batch size:64. By visualized as a figure, we try to get a proper value of learning rate and analyze the causes of different results.

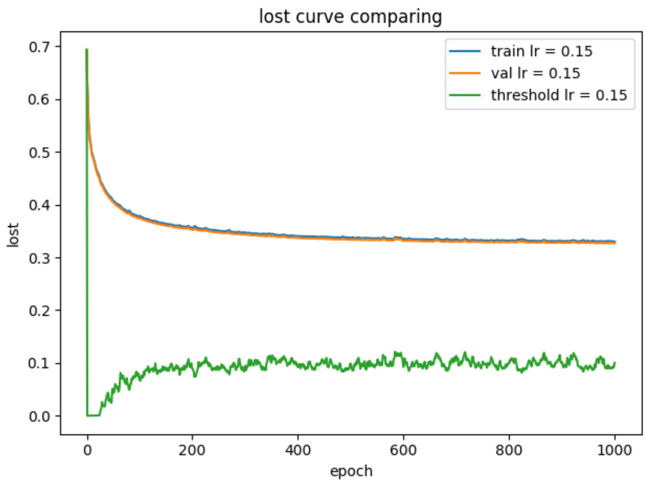

Fig. 1. Logistic regression learn rate:0.15

In figure 1 2 3 4, we can find out that during the learning rate increasing, the validation loss decrease progressively. However, when the learning rate is out of a boundary, the loss will be vibrating and can not achieve the best performance.

D. Step of SVM

The steps of our SVM are as the following:

1. Use load_svmlight_file function in sklearn library to load the a9a as train set a9a.t as valid set.
2. change the value of y from (1,-1) to (1,0).
3. Initialize SVM model parameters. Set all parameter into zero, initialize it randomly or with normal distribution.

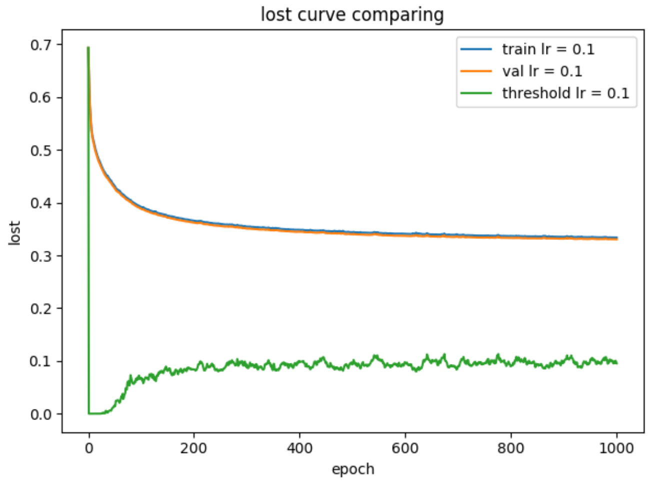

Fig. 2. Logistic regression learn rate:0.1

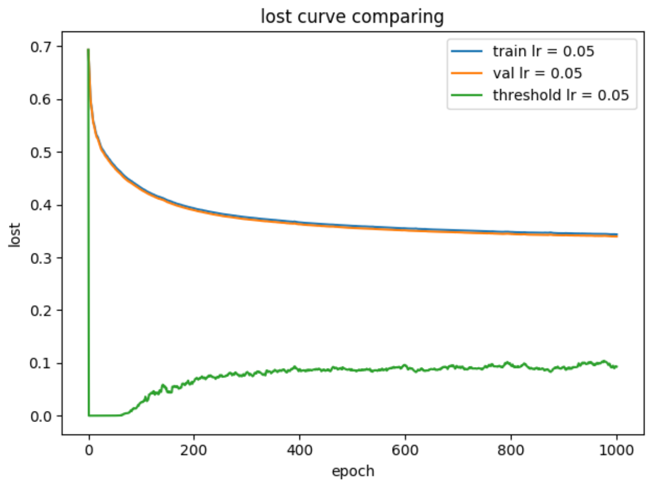

Fig. 3. Logistic regression learn rate:0.05

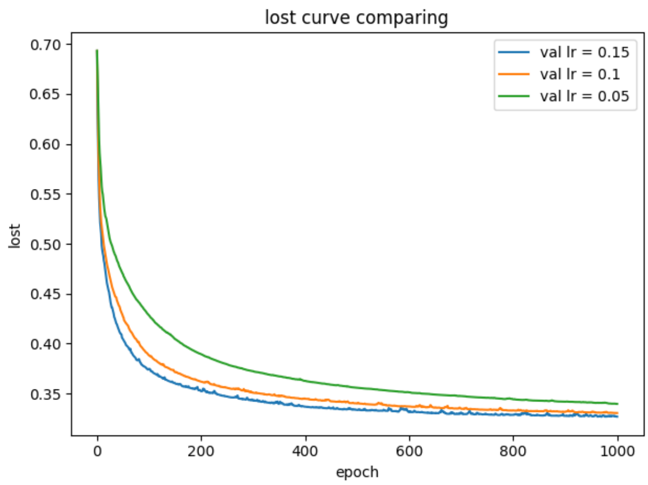

Fig. 4. Valid loss compare of different learn rate

4. Get a mini-batch data from train set randomly.
5. Update the parameters $\boldsymbol{w}$ with function (6) and (7)
6. Calculate the loss of train set and validation set with function (5).
7. Calculate the threshold loss of validation set, mark the sample whose predict scores greater than the threshold as positive, on the contrary as negative.
8. Repeate setp 4-7 for n times, and return the losses of train set, losses of validation set and the threshold losses.

E. Result of SVM

In SVM, we try to compare the different loss with different learning rate, other prameter is same: interaction:500,regular:0.5, and batch size:64.

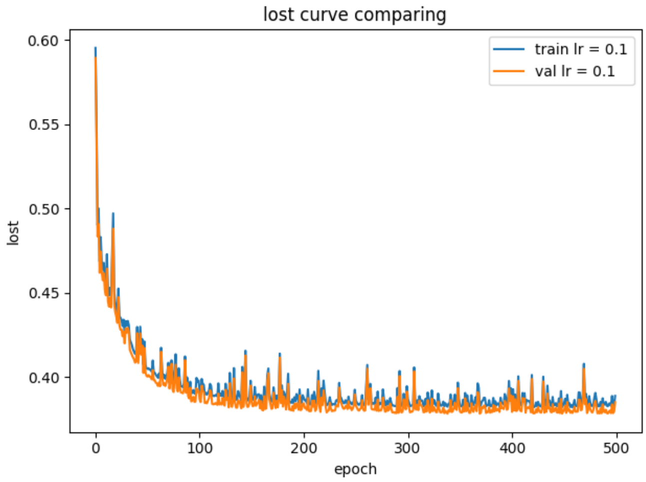

Fig. 5. SVM learn rate:0.1

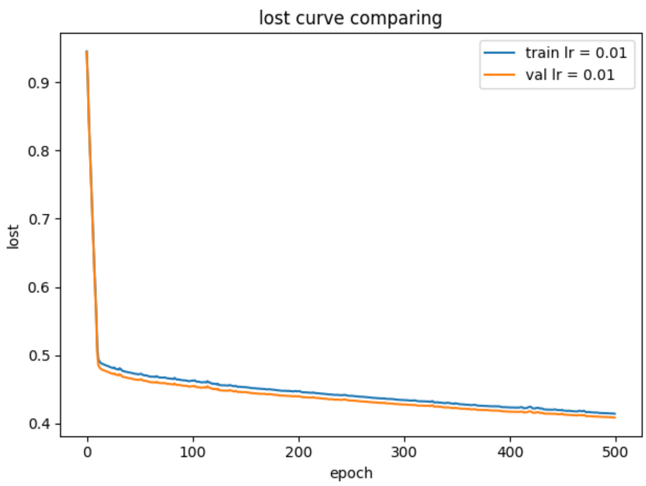

Fig. 6. SVM learn rate:0.01

IV. Conclusion

In this reprot, we mentioned Logistic regression and Support Vector Machine, both of them use SGD to approach the solution. For esaier to understand, we try to conduct some experiments and visualize the results of logistic regression and SVM and try to compare and analys the result.

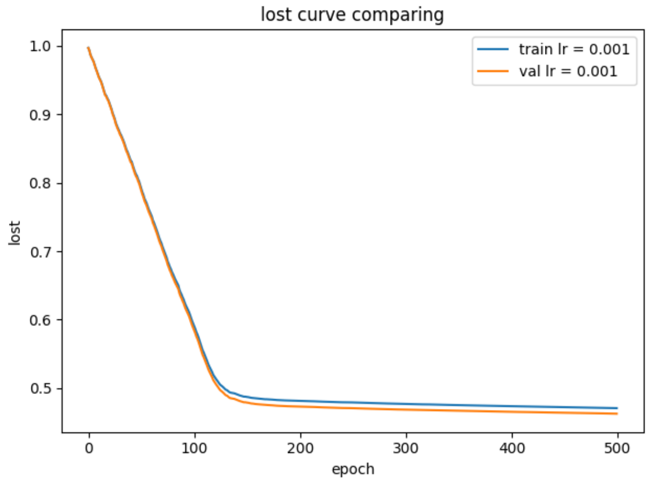

Fig. 7. SVM learn rate:0.001

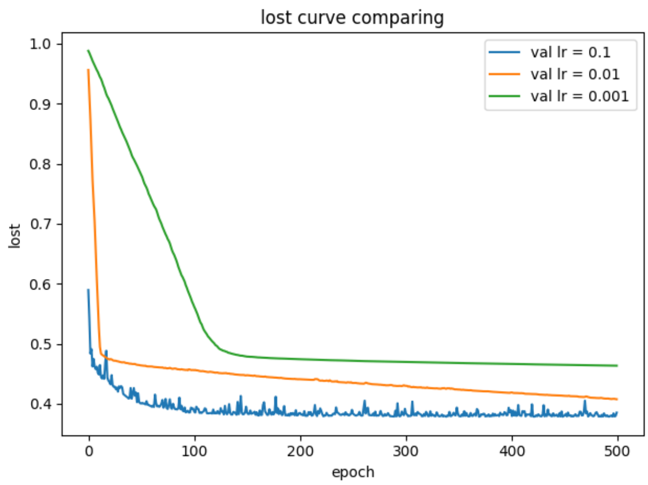

Fig. 8. SVM Valid loss compare of different learn rate