The Experiment Report of *Machine Learning*

SCHOOL: SCHOOL OF SOFTWARE ENGINEERING

SUBJECT: SOFTWARE ENGINEERING

Author:
Haipeng Deng

Supervisor:
Mingkui Tan

Student ID:
201730686193

Grade:
Undergraduate

November 7, 2019

Logistic Regression and Support Vector Machine

***Abstract*—This report introduces my work in Lab2 : Logistic Regression and Support Vector Machine. In this lab, a logistic regression model and a SVM are constructed and Mini-Batch Stochastic Gradient Descent is the method of minimizing the loss of the two models.**

I. Introduction

REGRESSION and classification are the two main categories in the field of supervised learning. A regression model is good at predicting continuous values while classification is always used to predicting discrete values i.e. classifying. However, the logistic regression model has nothing to do with the term 'logic' nor regression. It computes how probable that the test input is of a certain type. Compared to logistic regression, the support vector machine provides a mean to find a hyperplane which divides the training data in the hyperplane where those data scattered. An SVM has better robustness, thus it is used in many tasks and shows many advantages. This lab focuses on understanding mini-batch stochastic gradient descent, discovering the differences and relationships between Logistic regression and linear classification and practising on a larger dataset.

II. Methods and Theory

Logistic Regression

The logistic regression model is a kind of linear model but used for classification tasks. The function that was used in this lab is:

$$h_w(x) = g(\sum_{i=1}^{m} w_i x_i) = g(w^T x) \quad where \quad g(z) = \frac{1}{(1+e^-z)} \tag{1}$$

The function can be considered as a good function if:

$$\begin{cases} h_w(x) \approx 1 & y = 1 \\ h_w(x) \approx 0 & y = -1 \end{cases} \tag{2}$$

A simple function of measuring the loss is:

$$E_{in}(h) = \frac{1}{n}\sum_{i=1}^{n}(h_w(x_i) - \frac{1}{2}(1+y_i))^2 \tag{3}$$

But it is not convenient to use cause it is hard to minimize by finding the gradient of it. Instead, we can use a function with a regulation parameter λ which trades off between fitting the training set well and keeping the model relatively simple like:

$$J(w) = \frac{1}{n}\sum_{i=1}^{n} log(1 + e^{-y_i w^T x_i}) + \frac{\lambda}{2}||w||_2^2 \tag{4}$$

To apply Mini-batch Stochastic Gradient Descent in logistic regression, we computes the gradient and update the w in the below way:

$$\frac{\partial J(w)}{\partial w} = \frac{1}{n}\sum_{i=1}^{n}(h_w(x_i) - y)x_i \tag{5}$$

$$w := w - \alpha \frac{\partial J(w)}{\partial w} \tag{6}$$

Support Vector Machine

The implementation and goal of constructing a support vector machine is easy to understand. What an SVM do is basically selects two parallel hyperplanes that separate the two classes of data and let the distance between them as large as possible, i.e.

$$\begin{cases} w^T x_+ + b = +1 \\ w^T x_- + b = -1 \end{cases} \tag{7}$$

and the margin is $\frac{2}{||w||}$. However, the training data can not be separated linearly sometimes

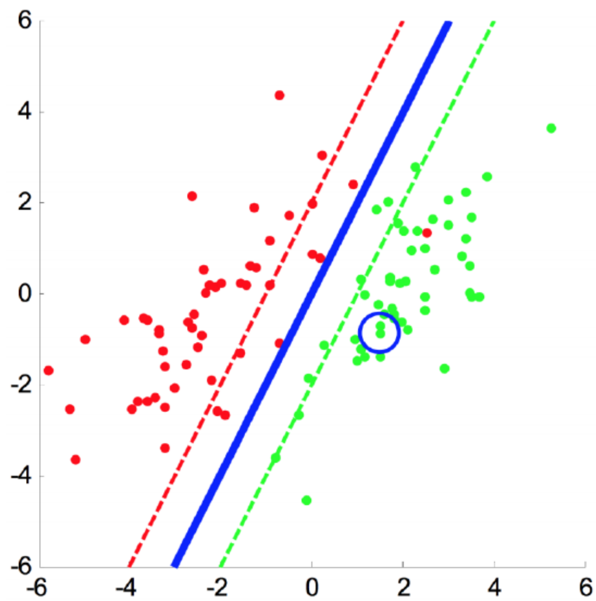

Fig. 1. Example of not linearly separable data

To address this problem, we introduce variable $\varepsilon_i \geq 0$, for each i, which represents how much example i is on wrong side of margin boundary. Thus in this lab, we use hinge loss function to evaluate the goodness of our model:

$$\varepsilon_i = max(0, 1 - y_i(w^T x_i + b)) \tag{8}$$

In this lab, the optimization problem then becomes:

$$min(\frac{||w||^2}{2} + \frac{C}{n}\sum_{i=1}^{n} max(0, 1 - y_i(w^T x_i + b)) \tag{9}$$

The MSGD method is used to reduce the loss of the model.

III. Experiments

A. Dataset

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

B. Implementation

Logistic Regression and Batch Stochastic Gradient Descent

Firstly,load the training set and validation set. Then,To initialize logistic regression model parameter, I initialize my w with a matrix whose all entries are zero. Thirdly, I determine the batch size(1024) and randomly take some samples,calculate gradient G toward loss function from partial samples and use the MSGD optimization method to update the parametric model Finally, to evaluate the model I select the appropriate threshold, mark the sample whose predict scores greater than the threshold as positive, on the contrary as negative, predict under validation set and get the loss on validation set.

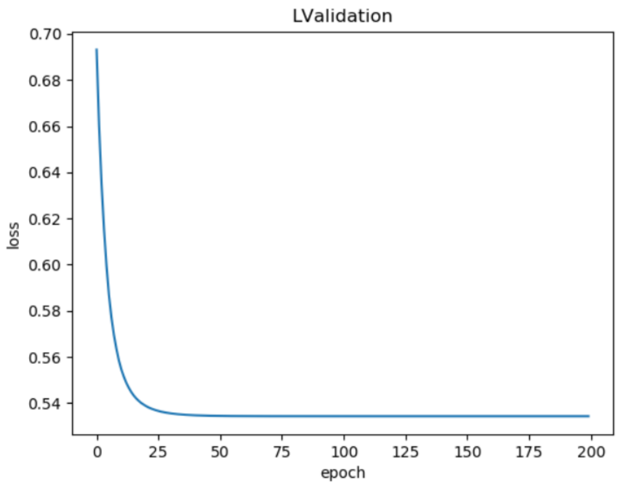

Fig. 2. Lvalidation chage of logistic regression

Linear Classification and Batch Stochastic Gradient Descent

After loading the training set and validation set I initialize SVM model parameters w into a matrix of zeros. The loss function that I select has been introduced above. Next I choose 1024 as my batch size and randomly take some samples,calculate gradient G toward loss function from partial samples. MSGD optimization is the following step and finally I get the result graph as below:

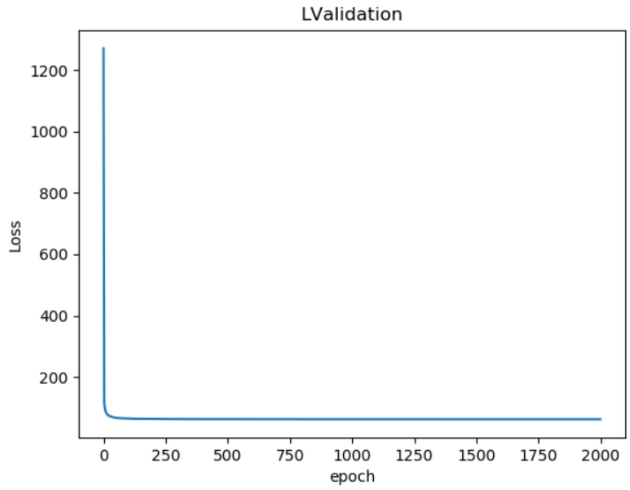

Fig. 3. Lvalidation chage of SVM.

IV. Conclusion

This lab is quite challenging compared to lab1, especially in the SVM part. After constructing my naive model, I found that my model has a problem that a pal told me. The problem is that after training, my model can simply mark all the testing set data x_i into -1. Days after, my classmate tells me that I may ignore the C constant in the model. Then I make some researches online and adjust my C in the model, the problem is addressed! This lab is fabulously inspiring