South China University of Technology

The Experiment Report of Machine Learning

SCHOOL: SCHOOL OF SOFTWARE ENGINEERING

SUBJECT: SOFTWARE ENGINEERING

Author:
Haijun You

Supervisor:
Mingkui Tan or Qingyao Wu

Student ID：
201721045312

Grade:
Graduate

December 9, 2017

Linear Regression, Linear Classification and Gradient Descent

Abstract—In this paper, logistical regression and linear classification with stochastic gradient descent are introduced and performed with experiment, respectively.

I. Introduction

This report is of the second experiment. This experiment contains two portions, which are as follows: a) logistical regression and stochastic gradient descent are need performed with 9a9 data set in LIBCVM data; b) linear classification and stochastic gradient descent are applied in australian data set.

II. METHODS AND THEORY

1) Initialization

a) logistical regression and stochastic gradient descent: parameters of weight are set to zero , and learning rate is set to 0.1. The min-batch is set to 100. The other parameters were set variously according to specific situations.

b) linear classification and gradient descent: parameters of weight and basis are set zero, and learning rate is set to 0.001. The min-batch is set to 100. The other parameters were set variously according to specific situations.

2) Loss function and gradient

a) logistical regression and stochastic gradient descent:

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

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

$$\frac{\partial J(\mathbf{w})}{\partial \mathbf{w}} = \frac{1}{n}\sum_{i=1}^{n}\left(h_{\mathbf{w}}(\mathbf{x}_i) - y\right)\mathbf{x}_i$$

$$\mathbf{w} := \mathbf{w} - \frac{1}{n}\sum_{i=1}^{n}\alpha\left(h_{\mathbf{w}}(\mathbf{x}_i) - y_i\right)\mathbf{x}_i$$

b) linear classification and stochastic gradient descent:

$$\min_{\mathbf{w},b} \quad f : \frac{\|\mathbf{w}\|^2}{2} + C\sum_{i=1}^{N}\max(0, 1 - y_i(\mathbf{w}^\top x_i + b))$$

$$\nabla_{\mathbf{w}} L(\mathbf{w}, b) = \mathbf{w} + \frac{C}{|\mathcal{S}_k|}\sum_{i\in\mathcal{S}_k} g_{\mathbf{w}}(\mathbf{x}_i)$$

$$\nabla_b L(\mathbf{w}, b) = \frac{C}{|\mathcal{S}_k|}\sum_{i\in\mathcal{S}_k} g_b(\mathbf{x}_i)$$

III. Experiment

a) logistical regression and stochastic gradient descent

The loss results of train and validation are drawn and shown in Figure 1,whose parameters are set with iteration is 500. And figure 2 shows the loss result with RMSprop parameters optimization. Figure 3 shows the loss result with Adm parameters optimization. Figure 4 shows the loss result with NAG parameters optimization.

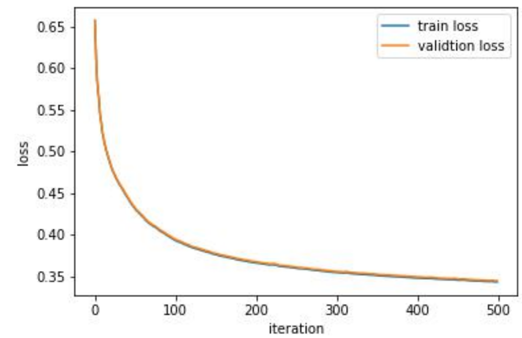

Figure 1 Loss results of logistical regression

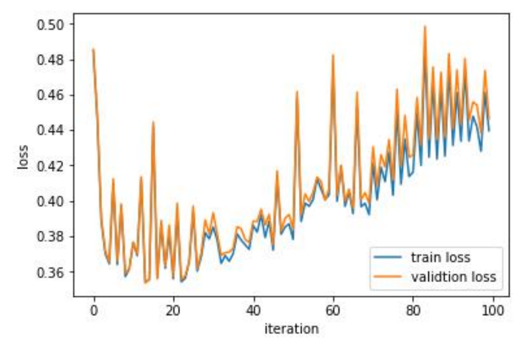

Figure 2 Loss result with RMSprop

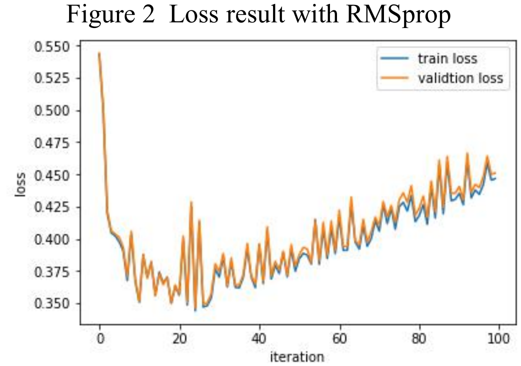

Figure 3 Loss result with Adm

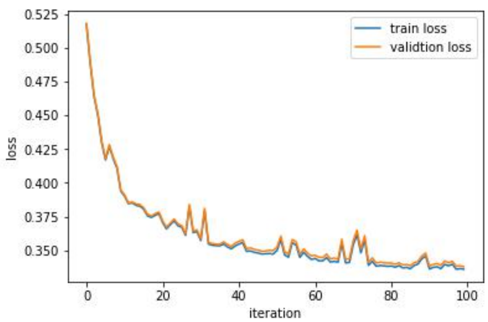

Figure 4 Loss result with NAG

b)linear classification and logistical gradient descent

The loss results of train and validation are drawn and shown in Figure 5,whose parameters are set with iteration is 100. And figure 6 shows the loss result with RMSprop parameters optimization. Figure 7 shows the loss result with Adm parameters optimization. Figure 8 shows the loss result with NAG parameters optimization.

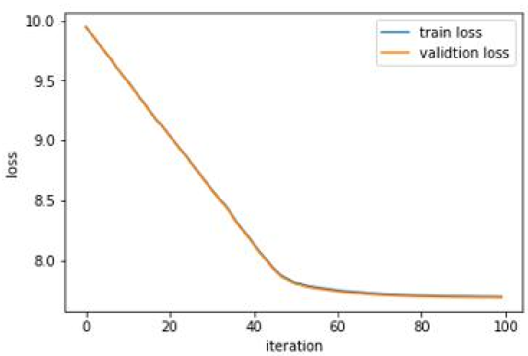

Figure 5 Loss results of linear classification

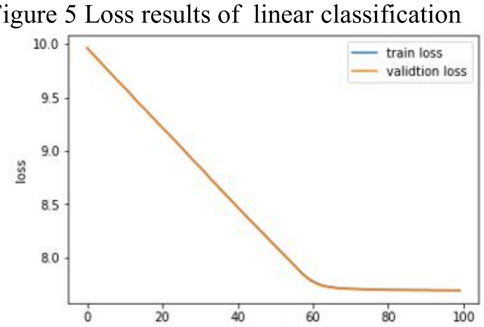

Figure 6 Loss result with RMSprop

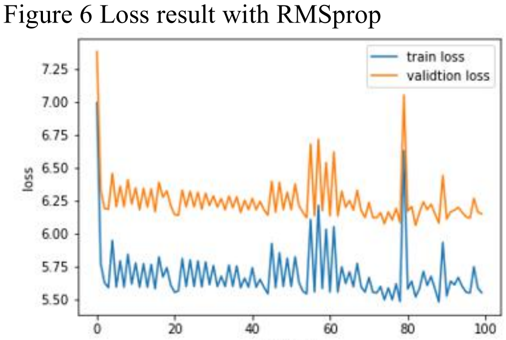

Figure 7 Loss result with Adm

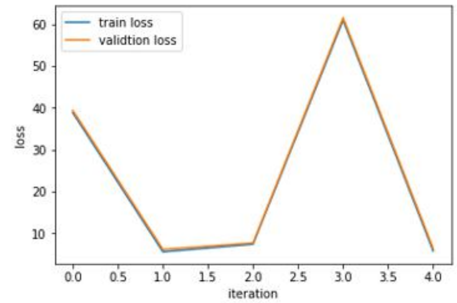

Figure 8 Loss result with NAG

IV. CONCLUSION

These two models apply logistical gradient descent to match and classify.The differences is the loss function, derivation function and gradient. The steady loss values were obtained of these two different approached in the end.

In general, the training time of SGD is longer than other optimization method. The optimization method with self-adaptation is useful if the faster convergence is necessary. Adadelta, RMSprop, Adm are similar algorithms with similar performance in the same condition.