The Experiment Report of Machine Le

SCHOOL: SCHOOL OF SOFTWARE ENGINEERING

SUBJECT: SOFTWARE ENGINEERING

Author:
黄品超

Supervisor:
Mingkui Tan

Student ID：201530611760

Grade:
Undergraduate

December 9, 2017

Linear Regression, Linear Classification and Gradient Descent

Abstract—

I. Introduction

In this experiment, the first target is to compare and understand the difference between gradient descent and stochastic gradient descent.

Secondly, after the whole process, a realization of differences and connection between Logistic regression and linear classification should have cultivated.

Finally, implement practicing over larger data, for further understanding principles of SVM.

II. Methods and Theory

This experiment subjectively focus on the classification problem, in which I have to try different ways of linear classification and linear regression. For improvement, Four different optimize method were chosen, NAG, RMSProp,AdaDelta and Adam.

The loss function I have chosen is Cross Entropy Error, as follow shows:

$$E_{\text{in}}(\mathbf{w}) = \frac{1}{N}\sum_{n=1}^{N} \ln(1 + e^{-y_n \cdot \mathbf{w}^{\top}\mathbf{x}_n})$$

And the derivatives:

$$\frac{1}{n}\sum_{i=1}^{n} \frac{y_i \mathbf{x}_i}{1 + e^{y_i \cdot \mathbf{w}^{\top}\mathbf{x}_i}}$$

III. Experiment

The following content is completely came from the experiment guide.

The experimental code and drawing are completed on jupyter.

Logistic Regression and Stochastic Gradient Descent:

Load the training set and validation set.
Initalize logistic regression model parameters, you can consider initalizing zeros, random numbers or normal distribution.
Select the loss function and calculate its derivation, find more detail in PPT.
Calculate gradient toward loss function from partial samples.
Update model parameters using different optimized methods(NAG，RMSProp，AdaDelta and Adam).
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 different optimized method loss L_NAG ， L_RMSProp，L_AdaDelta and L_Adam .

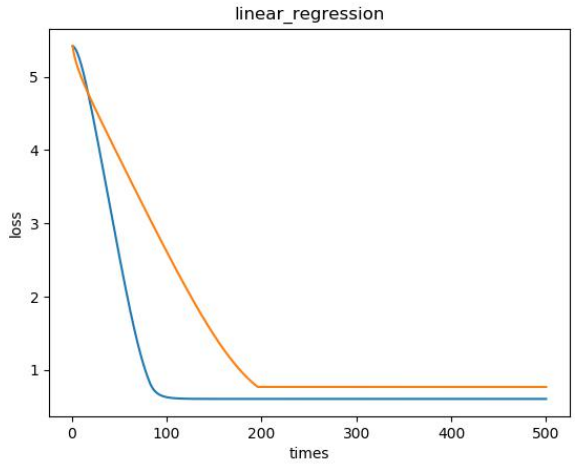
Repeate step 4 to 6 for several times, and drawing graph of L_NAG，L_RMSProp，L_AdaDelta and L_Adam with the number of iterations.

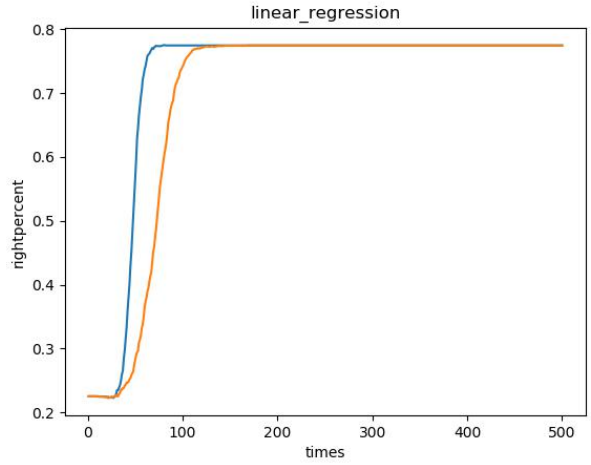
Linear Classification and Stochastic Gradient Descent:

Load the training set and validation set.
Initalize SVM model parameters, you can consider initalizing zeros, random numbers or normal distribution.
Select the loss function and calculate its derivation, find more detail in PPT.
Calculate gradient toward loss function from partial samples.
Update model parameters using different optimized method loss L_NAG ， L_RMSProp，L_AdaDelta and L_Adam .

Repeate step 4 to 6 for several times, and drawing graph of L_NAG，L_RMSProp，L_AdaDelta and L_Adam with the number of iterations.

Since I only finished part of the experiment, I only give the loss result of linear regression here, where I have used both NAG and RMSProp on it.





IV. CONCLUSION

This experiment focus on the two method of linear regression and linear classification. As a conclusion of the experiment, both of them come out with good result on prediction jobs. In my perspective, the only major difference between them is that the linear regression is really munch more easier to understand, though it seems to be harder for me to put that in to practice when coding. Ps. The linear regression is weaker than classification in terms of performance, for it implements with exactly more functions and complicate models, that was in no surprise.