华南理工大学

South China University of Technology

The Experiment Report of Machine Learning

SCHOOL: SCHOOL OF SOFTWARE ENGINEERING

SUBJECT: SOFTWARE ENGINEERING

Author:
Zhaohui Huang

Supervisor:
Qingyao Wu

Student ID：201720145112

Grade: Graduate

December 15, 2017

Linear Regression, Linear Classification and Gradient Descent

Abstract—Gradient descent is a first-order iterative optimization algorithm for finding the minimum of a function. In this paper we use compare the difference between gradient descent and stochastic gradient descent in SVM.

I. Introduction

The main purposes of this report can be concluded as the following:

1. Compare and understand the difference between gradient descent and stochastic gradient descent.
2. Compare and understand the differences and relationships between Logistic regression and linear classification.
3. Further understand the principles of SVM and practice on larger data.

II. Methods and Theory

The loss function of logistic regression is:

$$L_{reg} = -\frac{1}{N}\sum_{i=1}^{N} y^{(i)} log(h(x^{(i)})) + (1-y^{(i)}) log(1-h(x^{(i)})),$$

$$\text{where } h(x^{(i)}) = \frac{1}{1+e^{-w^T x}}$$

The corresponding gradient with respect to weight in logistic regression

$$\frac{\partial L_{reg}}{\partial w} = \frac{1}{N}\sum_{i=1}^{N}(h(x^{(i)}) - y^{(i)}) * x_j^{(i)}$$

The loss function of linear classification, e.g.Support Vector Machine (SVM):

$$L_{cls} = \frac{\|w\|^2}{2} + C\frac{1}{N}\sum_{i=1}^{N} max(0, 1 - y_i(w^T x_i + b))$$

Stochastic Gradient Descent (SGD). Stochastic gradient descent (SGD) in contrast performs a parameter update for each training example x (i) and y (i) :

$$\theta = \theta - \eta \nabla_\theta L(\theta; x^{(i)}; y^{(i)})$$

III. Experiment

Dataset

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

Experiment Step

Logistic Regression and Stochastic Gradient Descent

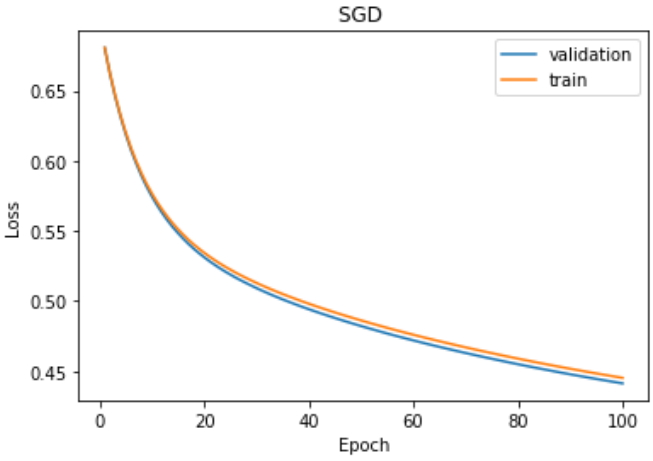
1. Load the training set and validation set.
2. Initalize logistic regression model parameters, you can consider initalizing zeros, random numbers or normal distribution.
3. Select the loss function and calculate its derivation, find more detail in PPT.
4. Calculate gradient toward loss function from **partial samples**.
5. **Update model parameters using different optimized methods(NAG，RMSProp，AdaDelta and Adam).**
6. Select the appropriate threshold, mark the sample whose predict scores **greater than the threshold as positive, on the contrary as negative**.

Linear Classification and Stochastic Gradient Descent

1. Load the training set and validation set.
2. Initalize SVM model parameters, you can consider initalizing zeros, random numbers or normal distribution.
3. Select the loss function and calculate its derivation, find more detail in PPT.
4. Calculate gradient toward loss function from **partial samples**.
5. **Update model parameters using different optimized methods(NAG，RMSProp，AdaDelta and Adam).**
6. Select the appropriate threshold, mark the sample whose predict scores **greater than the threshold as positive, on the contrary as negative**.

Result

Logistic Regression and Stochastic Gradient Descent



IV. CONCLUSION

Logistic regression is a statistical method for analyzing a dataset in which there are one or more independent variables that determine an outcome. The outcome is measured with a dichotomous variable (in which there are only two possible outcomes). Due to the reason of time, I only implement logistic regression by SGD. I will complete others after class.