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:

Changlin Yan

Supervisor:

Qingyao Wu

Student ID：

201530613368

Grade:

Undergraduat

December 12, 2017

Linear Regression, Linear Classification and Gradient Descent

Abstract—Logistic regression is a widely used classification machine learning algorithm that fits data to a logit function (or logistic function) to predict the probability of an event occurring.

Linear classification refers to trying to find the interface between two types (or classes) in the feature space.

I. Introduction

Logistic regression is a widely used classification machine learning algorithm that fits data to a logit function (or logistic function) to predict the probability of an event occurring.

Linear classification refers to trying to find the interface between two types (or classes) in the feature space.

The goal of our experiment is:

Compare and understand the differences and relationships between gradient descent and stochastic gradient descent.

Compare and understand the differences and relationships between logistic regression and linear classification.

Further understand the principles of SVM and practice on larger data

II. Methods and Theory

NAG

NAG not only adds the momentum term, but also subtracts the momentum term in the loss function to calculate $\nabla \theta J(\theta - \gamma v_{t-1})$ when calculating the gradient of the parameter, which predicts the value of the next parameter position. which is:

$$v_t = \gamma v_{t-1} + \eta \cdot \nabla \theta J(\theta - \gamma v_{t-1})$$

$$\theta = \theta - v_t$$

RMSprop.

RMSprop is a very effective, but currently unpublished adaptive learning rate method. Amusingly, everyone who uses this method in their work currently cites slide 29 of Lecture 6 of Geoff Hinton's Coursera class. The RMSProp update adjusts the Adagrad method in a very simple way in an attempt to reduce its aggressive, monotonically decreasing learning rate. In particular, it uses a moving average of squared gradients instead, giving:

```
cache = decay_rate * cache + (1 - decay_rate) * dx**2
x += - learning_rate * dx / (np.sqrt(cache) + eps)
```

AdaDelta

Adadelta is an extension of Adagrad, the initial program is still adaptive constraint on the learning rate, but a computational simplification. Adagrad accumulates all previous gradient squared, whereas Adadelta only accumulates fixed-size items and does not directly store these terms, only the approximate average of the approximate calculations.

Adam.

Adam is a recently proposed update that looks a bit like RMSProp with momentum. The (simplified) update looks as follows:

```
m = beta1*m + (1-beta1)*dx
v = beta2*v + (1-beta2)*(dx**2)
x += - learning_rate * m / (np.sqrt(v) + eps)
```

III. EXPERIMENT

Logistic Regression and Stochastic Gradient Decrease

(1) read the experimental training set and verification set.

(2) Logistic regression model parameter initialization, you can consider all zero initialization, random initialization or normal distribution initialization.

(3) Select Loss function and its derivative.

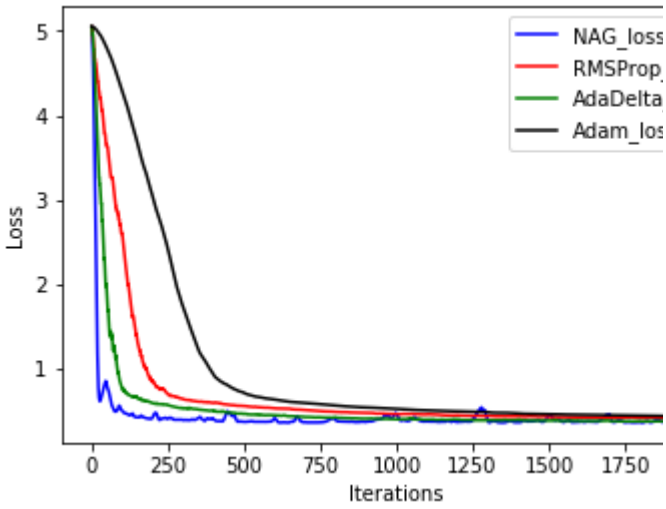
(4) Find the gradient G of some samples to Loss function.

(5) Update the model parameters using different optimization methods (NAG, RMSProp, AdaDelta and Adam).

(6) Select the appropriate threshold, the verification of centralized computing results greater than the threshold mark is positive, otherwise negative. Test on the validation set and get the Loss function values for different optimization methods L_{NAG}, $L_{RMSProp}$, $L_{AdaDelta}$ and L_{Adam}.

(7) Repeat steps 4-6 several times, draw L_{NAG}, $L_{RMSProp}$, $L_{AdaDelta}$ and L_{Adam} change with the number of iterations.

result:



Linear classification and stochastic gradient descent

(1) read the experimental training set and verification set.

(2) Initialization of support vector machine model parameters can consider all-zero initialization, random initialization or normal distribution initialization.

(3) Select the Loss function and its derivative, the process see courseware ppt.

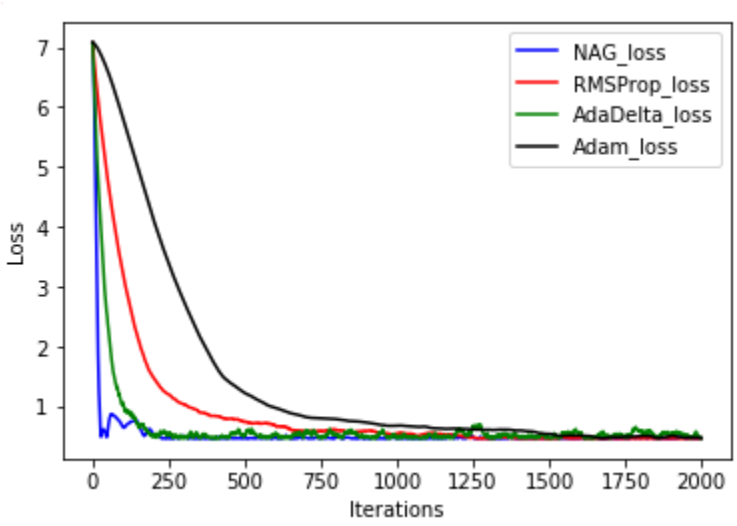
(4) Find the gradient of some samples to Loss function.

(5) Update the model parameters using different optimization methods (NAG, RMSProp, AdaDelta and Adam).

(6) Select the appropriate threshold, the verification of centralized computing results greater than the threshold mark is positive, otherwise negative. Test on the validation set and get the Loss function values for different optimization methods L_{NAG}, $L_{RMSProp}$, $L_{AdaDelta}$ and L_{Adam}.

(7) Repeat steps 4-6 several times, draw L_{NAG}, $L_{RMSProp}$, $L_{AdaDelta}$ and L_{Adam} change with the number of iterations.

result:



IV. CONCLUSION

Through this experiment, I understand more clearly the difference and connection between gradient descent and stochastic gradient descent.

Let me understand the difference and connection between logistic regression and linear classification.

Let me further understand the principles of SVM and practice on larger data

Let me know how to use different optimization methods to update the model parameters