华南理工大学

South China University of Technology

The Experiment Report of

Deep Learning

College	Software College
Subject	Software Engineering
Student ID	201722800100
E-mail	Az.yamen@gmail.com
Tutor	Mingkui Tan
Date submitted	2017.12 .22

Logistic Regression, Linear Classification and Stochastic Gradient Descent

Time: 2017-12-22

Reported By:

Ezzaddin Ahmed Othman Saeed

Abstract

Regression and Classification are two basic techniques of machine learning. In this report, We will introduce Logistic Regression and Linear Classification .And we will update two models parameters using Stochastic Gradient Descent(SGD). Among our experiments, we update two models parameters using four optimized methods(NAG, RMSProp, AdaDelta and Adam) and compare the effect of them.

Key words: Logistic Regression; Linear Classification; SGD;

I. Introduction:

In this report, we will introduce Logistic Regression and Linear Classification. We will use support vector machine(SVM) to solve Linear Classification in order to further understand the principles of SVM. And we will update two models parameters using Stochastic Gradient Descent. Among our experiments, We will use four optimized methods(NAG, RMSProp, AdaDelta and Adam) to update the model parameters and compare the effect of them.

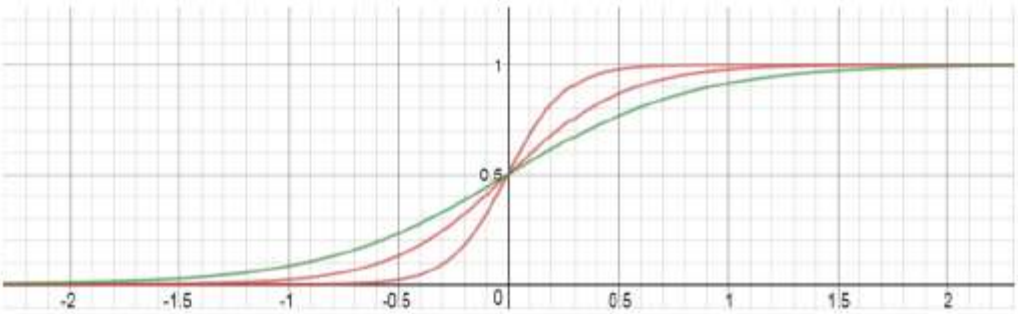
II. Purposes:

- Compare and understand the difference 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.

III. METHODS AND THEORY

In this section, I will illustrate chosen methods, the related theories, the related equations(loss function) and the derivation process(taking the gradient) ,for Logistic Regression :can be expressed as

$$g(z) = \frac{1}{1 + e^{-z}}$$



- This function is a continuous function.
  If z →-∞, then g (z) → 1; If z →∞, then g (z) → 0.

assume that the data is binary as follow:

D = {(x_1, y_1=±1),…., (x_n, y_n = ±1)}

then we can use the cross entropy error to measure the loss of the model as follow:

$$\mathbf{E}_{in}(\mathbf{w}) = \frac{1}{n}\sum_{i=1}^{n}\log(1 + e^{-y_n \cdot \mathbf{w}^\top \mathbf{x}})$$

We can also add regularization to avoid overfitting as follow:

$$J(\mathbf{w}) = \frac{1}{n}\sum_{i=1}^{n}\log(1+e^{-y_i \cdot \mathbf{w}^\top \mathbf{x}_i}) + \frac{\lambda}{2}||\mathbf{w}||_2^2$$

Assume that the data is binary as follow:

$D = \{(x_1,\ y_1 \in \{0,1\}), \ldots, (x_n,\ y_n \in \{0,1\})\}$

then the loss function of the model can be expressed as

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

for Linear Classification model : a linear classifier has the form:

$$f(x) = w^T x + b$$

Assume that $f(x) \in \{-1,1\}$, we will set a threshold that if the predictive value is higher than the threshold , then we set the classification result to 1,and if the predictive value is lower than the threshold , then we set the classification result to -1. We use the Support Vector Machine (SVM) model to solve Linear Classification. Then the loss function of the model takes the form

$$\min_{\mathbf{w},b} \frac{\|\mathbf{w}\|^2}{2} + C\sum_{i=1}^{n}\max(0, 1-y_i(\mathbf{w}^\top \mathbf{x}_i + b))$$

IV. EXPERIMENTS

A. Dataset:

These Experiments (Logistic Regression and Linear Classification) uses a9a of LIBSVM Data, including 32561/16281(testing) samples and each sample has 124/124 (testing) features.

B. Implementation :

for Experiment of Logistic Regression We set the learning-rate of NAG and RMSProp to 0.01, the **γ** of NAG and RMSProp to 0.9 , the **ε** of RMSProp to 10^{-8}, the **γ** of AdaDelta to 0.95, the **ε** of AdaDelta and Adam to 10^{-6} , the β_1 of Adam to 0.9, the β_2 of Adam to 0.999, the α of Adam to 0.01, the iterations to 200,and initialize both model parameters to zeros. the Figure1 shows that test-loss vary with the number of iterations using the four optimized method.

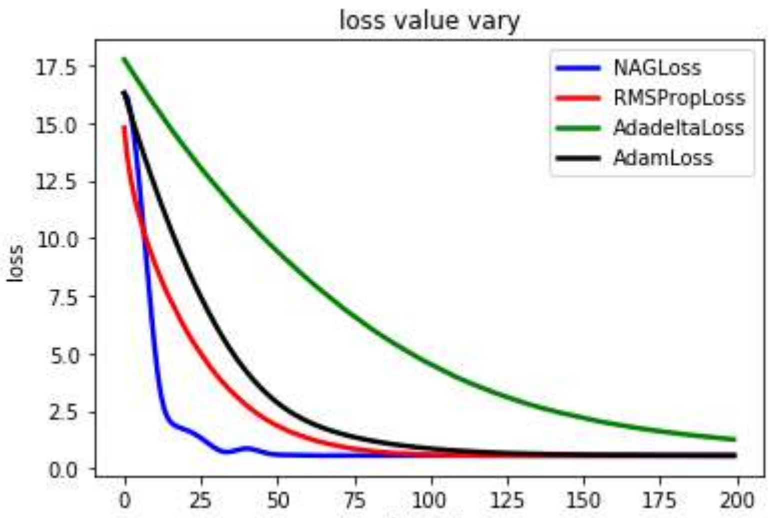


Figure1

For Experiment of Linear Classification We set the learning-rate of NAG and RMSProp to 0.01, the $\boldsymbol{\gamma}$ of NAG and RMSProp to 0.9, the $\boldsymbol{\varepsilon}$ of RMSProp to 10^{-8},the $\boldsymbol{\gamma}$ of AdaDelta to 0.95, the $\boldsymbol{\varepsilon}$ of AdaDelta and Adam to 10^{-6}, the β_1 of Adam to 0.9, the β_2 of Adam to 0.999, the α of Adam to 0.01, the iterations to 200,and initialize both model parameters to zeros. the Figure2 shows that test-loss vary with the number of iterations using the four optimized method.

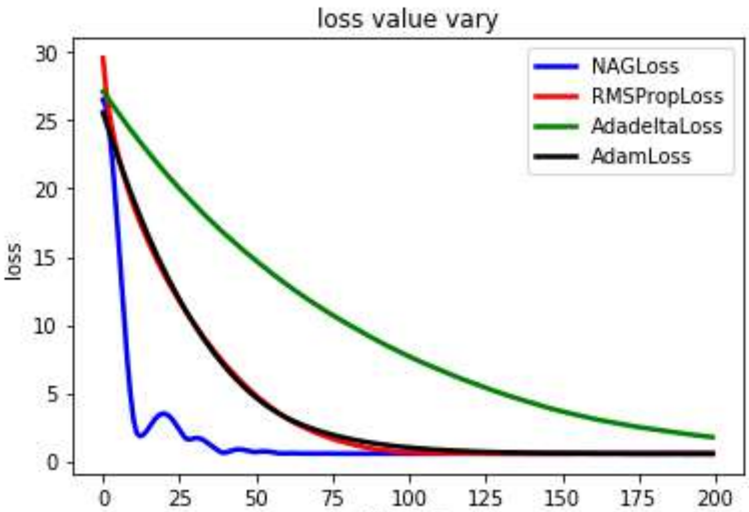


Figure2

from the Figure.2, we can find that both four optimized method can minimize the model. It is shown that NAG has the fastest convergence and AdaDelta has the slowest convergence.

V. CONCLUSION

In our experiments, we can find that four optimized methods are the good algorithms to minimize the loss function, when we use the cross entropy error as the loss function of Logistic Regression and use SVM to solve the Linear Classification. It is also shown that NAG has the fastest convergence and AdaDelta has the slowest convergence both in Logistic Regression and Linear Classification.