The Experiment Report of Machine Learning

SCHOOL: SCHOOL OF SOFTWARE ENGINEERING

SUBJECT: SOFTWARE ENGINEERING

Author:
潘楷乐

Supervisor:
Qingyao Wu

Student ID: 201530612583

Grade:
Undergraduate

December 9, 2017

Linear Regression, Linear Classification and Stochastic Gradient Descent

1.Introduction

Motivation of Experiment

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.

2.Method And Theory

Logistic Regression

Loss function:

$$\sum_{n} -\left[\underbrace{\hat{y}^n ln f_{w,b}(x^n) + (1-\hat{y}^n) ln\left(1 - f_{w,b}(x^n)\right)}_{\text{Cross entropy between two Bernoulli distribution}}\right]$$

gradient descent:

$$w_i \leftarrow w_i - \eta \sum_n -\left(\underline{\hat{y}^n - f_{w,b}(x^n)}\right) x_i^n$$

Linear Classification

Loss function:

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

gradient descent:

$$\nabla f(\beta) = \begin{cases} \mathbf{w}^\top - C\mathbf{y}^\top \mathbf{X} & 1 - y_i(\mathbf{w}^\top x_i + b) >= 0 \\ \mathbf{w}^\top & 1 - y_i(\mathbf{w}^\top x_i + b) < 0 \end{cases}$$

optimized methods

NAG

$$\begin{aligned} &\mathbf{g}_t \leftarrow \nabla J(\boldsymbol{\theta}_{t-1} - \gamma \mathbf{v}_{t-1}) \\ &\mathbf{v}_t \leftarrow \gamma \mathbf{v}_{t-1} + \eta \mathbf{g}_t \\ &\boldsymbol{\theta}_t \leftarrow \boldsymbol{\theta}_{t-1} - \mathbf{v}_t \end{aligned}$$

RMSProp

$$
\begin{aligned}
\mathbf{g}_t &\leftarrow \nabla J(\boldsymbol{\theta}_{t-1}) \\
G_t &\leftarrow \gamma G_t + (1-\gamma)\mathbf{g}_t \odot \mathbf{g}_t \\
\boldsymbol{\theta}_t &\leftarrow \boldsymbol{\theta}_{t-1} - \frac{\eta}{\sqrt{G_t + \epsilon}} \odot \mathbf{g}_t
\end{aligned}
$$

AdaDelta

$$
\begin{aligned}
\mathbf{g}_t &\leftarrow \nabla J(\boldsymbol{\theta}_{t-1}) \\
\mathbf{m}_t &\leftarrow \beta_1 \mathbf{m}_{t-1} + (1-\beta_1)\mathbf{g}_t \\
G_t &\leftarrow \gamma G_t + (1-\gamma)\mathbf{g}_t \odot \mathbf{g}_t \\
\alpha &\leftarrow \eta \frac{\sqrt{1-\gamma^t}}{1-\beta^t} \\
\boldsymbol{\theta}_t &\leftarrow \boldsymbol{\theta}_{t-1} - \alpha \frac{\mathbf{m}_t}{\sqrt{G_t + \epsilon}}
\end{aligned}
$$

Adam

$$
\begin{aligned}
\mathbf{g}_t &\leftarrow \nabla J(\boldsymbol{\theta}_{t-1}) \\
\mathbf{m}_t &\leftarrow \beta_1 \mathbf{m}_{t-1} + (1-\beta_1)\mathbf{g}_t \\
G_t &\leftarrow \gamma G_t + (1-\gamma)\mathbf{g}_t \odot \mathbf{g}_t \\
\alpha &\leftarrow \eta \frac{\sqrt{1-\gamma^t}}{1-\beta^t} \\
\boldsymbol{\theta}_t &\leftarrow \boldsymbol{\theta}_{t-1} - \alpha \frac{\mathbf{m}_t}{\sqrt{G_t + \epsilon}}
\end{aligned}
$$

3.Method And Theory

A.Dataset

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

B.Implementation

Initalize logistic regression model parameters: random numbers

Hyper-parameter selection:

Sgd: lr=0.01

Nag: lr=0.01, gamma=0.9, momentum is random

Rmsprop:lr=0.1, Gt=1, gamma=0.9, epsilon=10e-7

Adam: eplison=10e-8, beta=0.9, gamma=0.999, lr=0.1, m=0, Gt=0

Adadelta: Gt=0, delta_t=0, gamma=0.95, epsilon=10e-7, lr=10

Experiment Step

The experimental code and drawing are completed on jupyter.

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**. Predict under validation set and get the different optimized method loss L_{NAG}, $L_{RMSProp}$， $L_{AdaDelta}$ and L_{Adam} .

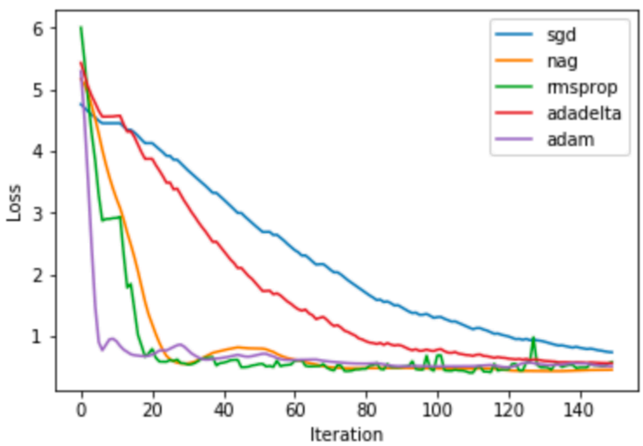
7. Repeate step 4 to 6 for several times, and **drawing graph of** L_{NAG}**,** $L_{RMSProp}$, $L_{AdaDelta}$ and L_{Adam} **with the number of iterations**.

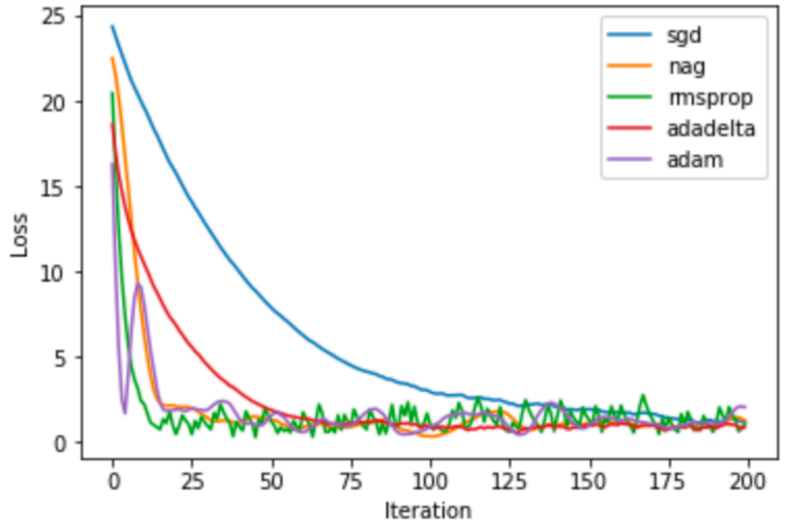
Linear Classification and Stochastic Gradient Descent

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**. Predict under validation set and get the different optimized method loss L_{NAG}, $L_{RMSProp}$, $L_{AdaDelta}$ and L_{Adam} .
7. 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**.

Experimental results and curve

Logistic Regression





4.Conclusion

Through this experiment, I have a deeper understanding of logical regression and linear classification, and also improve my practical ability and thinking ability. But I also have insufficient aspects, I finish the homework speed is very low, and more practice in the future.