

**The Experiment Report of**

***Machine Learning***

**College Software College**

**Subject Software Engineering**

**Members**   **洪海滨**

**Student ID 201530611586**

**E-mail Ocean233@live.com**

**Tutor**   **谭明奎**

**Date submitted** **2017.12.6**

1. **Topic:**

Logistic Regression, Linear Classification and Stochastic Gradient Descent

1. **Time:**

2017.12.2

1. **Reporter:**

洪海滨

**4. Purposes:**

Compare and understand the difference between gradient descent and stochastic gradient descent.

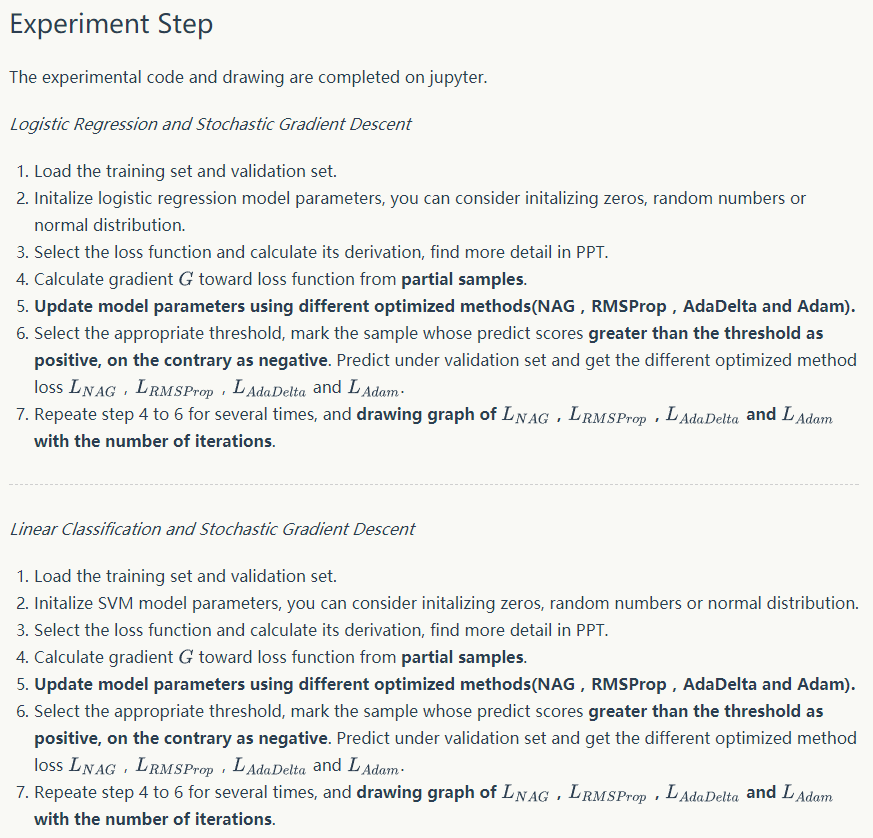
Compare and understand the differences and relationships between Logistic regression and linear classification.

Further understand the principles of SVM and practice on larger data.

1. **Data sets and data analysis:**

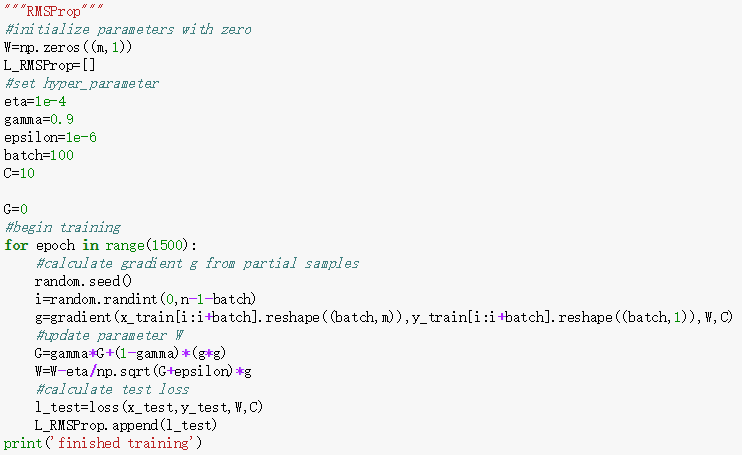
Experiment uses a9a of LIBSVM Data, including 32561/16281(testing) samples and each sample has 123/123 (testing) features.

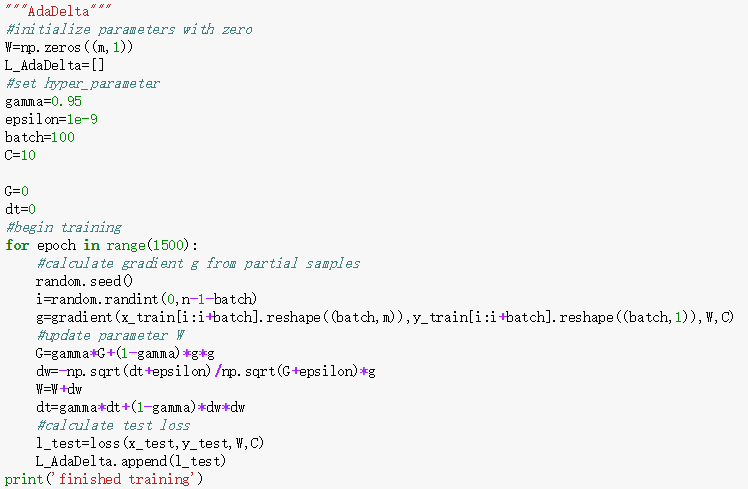
1. **Experimental steps:**

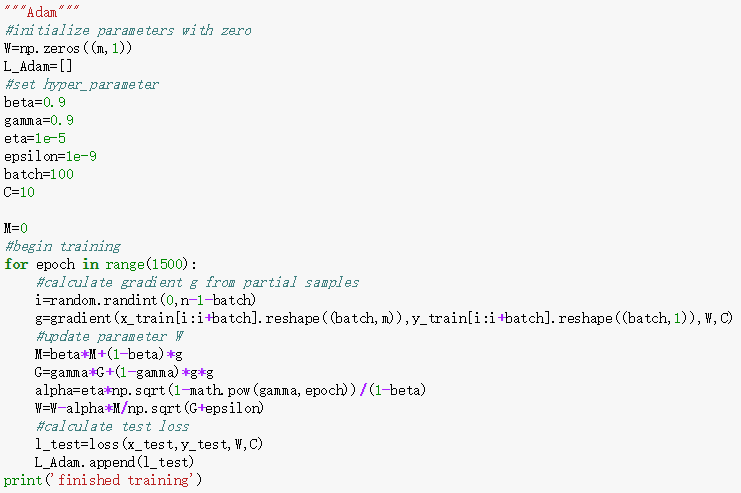
****

1. **Code:**









1. **Selection of validation (hold-out, cross-validation, k-folds cross-validation, etc.):**

hold-out

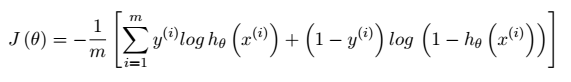
1. **The initialization method of model parameters:**

Initialize parameters with zero

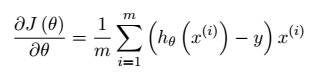
1. **The selected loss function and its derivatives:**

Logistic Regression:

loss function:

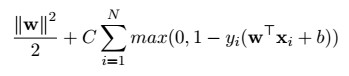
****

derivatives:

****

Linear Classification:

loss function:

****

derivatives:



**11. Experimental results and curve:**

## Hyper-parameter selection (η, epoch, etc.):

batch=100,epoch=1500

Logistic Regression:

NAG:eta=0.1,gamma=0.1,

RMSProp:eta=0.01,gamma=0.9,epsilon=1e-3

AdaDelta:gamma=0.9,epsilon=1e-6

Adam:eta=0.001,beta=0.9,gamma=0.999,epsilon=1e-6

Linear Classification:

NAG:eta=1e-6,gamma=1e-5

RMSProp:eta=1e-4,gamma=0.9,epsilon=1e-6

AdaDelta:gamma=0.95,epsilon=1e-9

Adam:eta=1e-5,beta=0.9,gamma=0.999,epsilon=1e-9

## Assessment Results (based on selected validation):

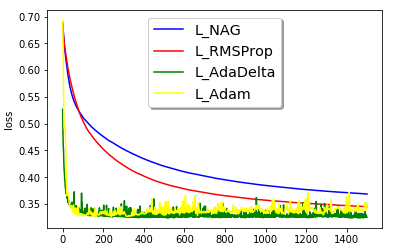
Logistic Regression：batch=100

NAG:eta=0.01,gamma=0.1

RMSProp:eta=0.001,gamma=0.9,epsilon=1e-3

AdaDelta:gamma=0.9,epsilon=1e-3

Adam:eta=0.1,beta=0.9,gamma=0.999,epsilon=1e-6

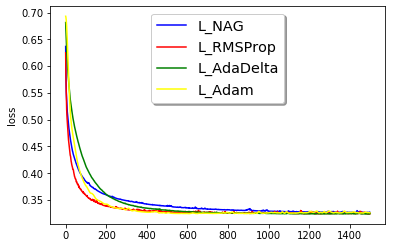


NAG:eta=0.1,gamma=0.1

RMSProp:eta=0.01,gamma=0.9,epsilon=1e-3

AdaDelta:gamma=0.9,epsilon=1e-6

Adam:eta=0.001,beta=0.9,gamma=0.999,epsilon=1e-6



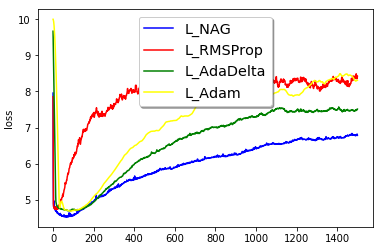
Linear Classification：batch=100,C=10

NAG:eta=1e-4,gamma=1e-5

RMSProp:eta=0.01,gamma=0.9,epsilon=1e-3

AdaDelta:gamma=0.95,epsilon=1e-6

Adam:eta=1e-3,beta=0.9,gamma=0.999,epsilon=1e-6

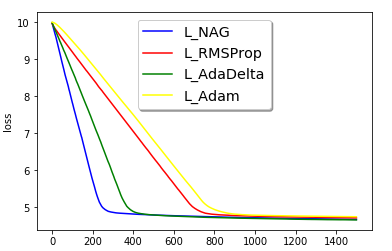


NAG:eta=1e-6,gamma=1e-5

RMSProp:eta=1e-4,gamma=0.9,epsilon=1e-6

AdaDelta:gamma=0.95,epsilon=1e-9

Adam:eta=1e-5,beta=0.9,gamma=0.999,epsilon=1e-9



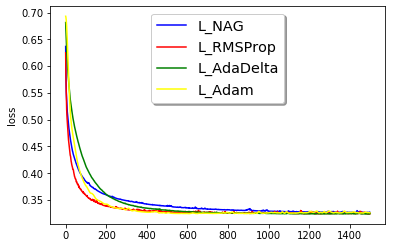
## Predicted Results (Best Results):

Logistic Regression：min\_loss≈0.32

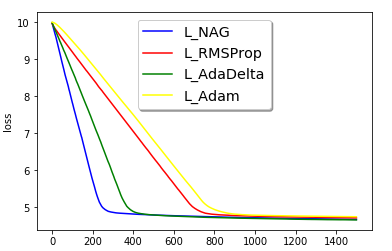
Linear Classification：min\_loss≈4.8

## Loss curve:

Logistic Regression：



Linear Classification：



**12. Results analysis:**

Logistic Regression:By using different optimized methods, all losses converge to the same value at similar speeds.

Linear classification:By using different optimized methods, all losses converge to the same value at different speeds.

**13. Similarities and differences between logistic regression and linear classification:**

Similarities:They are both used for classification.

Differences:The predicted values of logistic regression are probabilities and are therefore restricted to (0,1). The predicted values of linear classification are 0 and 1.

**14. Summary:**

Logistic regression performs better than linear classification at the data set. Under the same conditions, logistic regression spend less time computing loss and gradient. Also it’s loss converge more fast.

Hyper parameters cannot be learned directly from the data in the standard model training process and need to be predefined. They can be decided by setting different values, training different models, and choosing the values that test better.