

**The Experiment Report of**

***Machine Learning***

**College Software College**

**Subject Software Engineering**

**Members**  **王松盛**

**Student ID 201530612880**

**E-mail 975801238@qq.com**

**Tutor**   **谭明奎**

**Date submitted** **2017.12.8**

**1. Topic:** **Linear Regression, Linear Classification and Gradient Descent**

**2. Time: 2017.12.8**

**3. Reporter:王松盛**

**4. Purposes:**

**4.1** **Further understand of linear regression and gradient descent.**

**4.2 Conduct some experiments under small scale dataset.**

**4.3 Realize the process of optimization and adjusting parameters.**

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

**We conduct the Linear Regression experiment on Housing dataset in SIBSVM data, including 506 samples and each sample has 13 features. For Linear Classification problem, We use Australian in LIBSVM Data, including 690 samples and each sample has 14 features.**

**6. Experimental steps:**

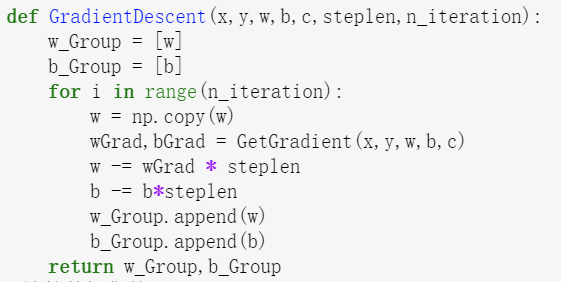
**In Linear Regression and Gradient Descent experiment, we did the experiment with the following steps:**

1. **Load the experiment data. We use load\_svmlight\_file function in sklearn library to load the data.**
2. **Devide dataset. We divide dataset into training set and validation set using train\_test\_split function.**
3. **Initialize linear model parameters. We set all parameter into zero in the experiment.**
4. **Choose loss function and derivation:We use error of mean square for linear regression problem, use hinge loss for linear classification problem.**
5. **Calculate gradient G toward loss function from all samples.**
6. **Denote the opposite direction of gradient G as D.**
7. **Update model:Wt = Wt-1 + nD, n is learning rate, a hyper-parameter that we can adjust.**
8. **Get the loss Ltrain under the training set and Lvalidation by validating under validation set.**

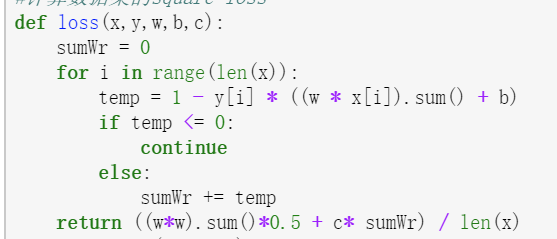
**9. Repeate step 5 to 8 for several times, and drawing graph of Ltrain as well as Lvalidation with the number of iterations.**

**7. Code:**

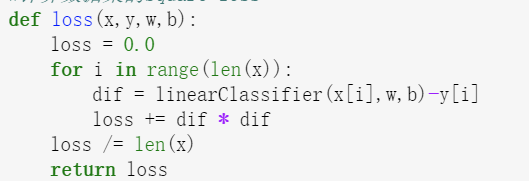
**Gradient descent:**



**Hinge loss:**



**Mean-square loss:**



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

**We use hold-out method for both experiments.**

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

**We initialize all parameter to zero in linear regression experiment; we use random value to initialize the parameter in linear classification experiment.**

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

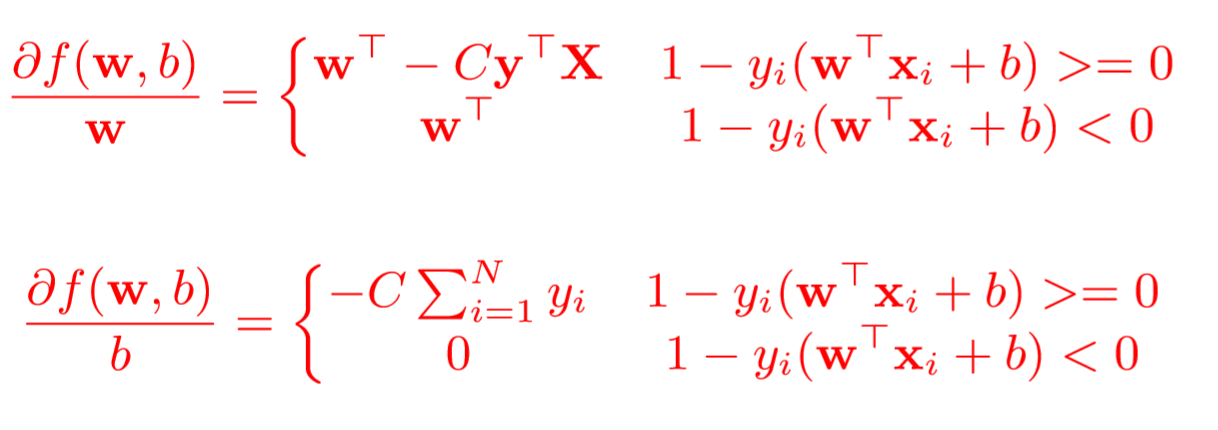
**For linear regression experiment, we use mean-square loss function:**

**The derivation of the function is:**

**For linear classification experiment, we use hinge loss function:**

**Hinge loss = ξi = max(0,1−yi(w>xi + b))**

**The derivation of the function is:**



1. **Experimental results and curve:**

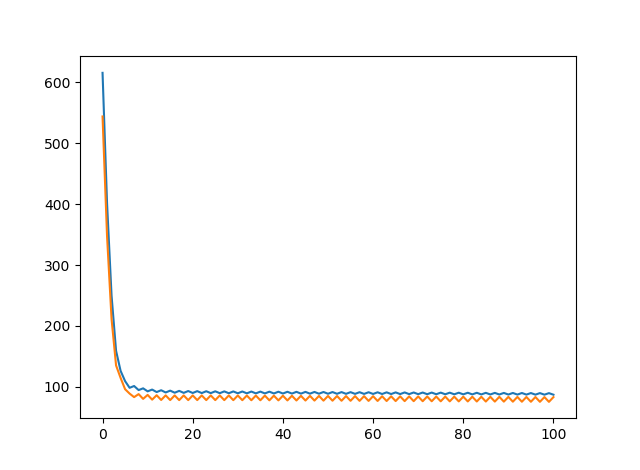
**Linear Regression experiment:**

**In this experiment, we set learning rate to 0.002, making 100 iterations.**

## Assessment Results (based on selected validation): loss value = 82.98

## Predicted Results (Best Results): loss value = 86.92

## Loss curve:



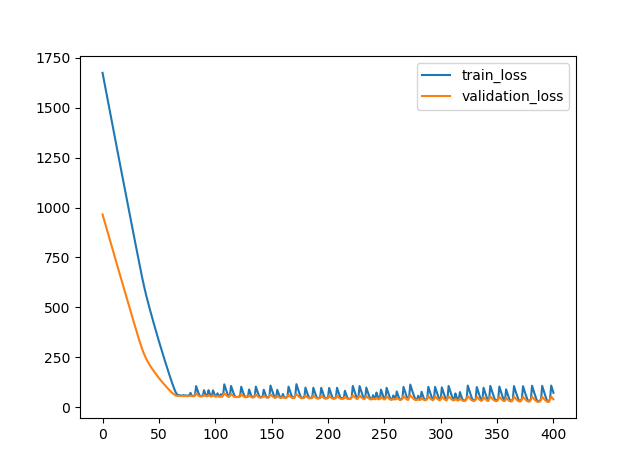
Linear classification experiment:

In this experiment, we choose learning rate 0.001,making 400 iterations.

## Assessment Results (based on selected validation): loss value = 39.15

## Predicted Results (Best Results): loss value = 72.98

Loss curve:



1. **Results analysis:**

In both experiment, the loss function decrease significantly at the first several iterations, then get down to a steady value.

We also found out that the loss curve can be effected by the initial value , learning rate and the time of iterations.

1. **Similarities and differences between linear regression and linear classification:**

**Similarity:**

1. **We use a linear model to solve both problem, trying to split the positive and negative sample with a hyper surface, or trying to predict the sample’s value with a hyper surface.**
2. **We use the same optimization algorithm: gradient descent algorithm to optimize the parameter of both models.**

**Difference:**

1. **The objective function of linear regression and classification is different. For linear regression, we need the predicted value to get closer to our answer; for linear classification, we hope to get a larger margin with a given loose parameter.**
2. **The parameter of Gradient descent function is different. As we use two different dataset, we should retry the parameter of gradient descent to get a better answer.**
3. **Summary:**

**In this experiment, we conduct the linear classification and regression experiment to two different dataset. We use the gradient descent algorithm to optimize the parameter of the learning model. We realize the importance of setting proper initial value and learning rate to get the better answer.**