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:
Shuai Zou

Supervisor:
Qingyao Wu

Student ID：
201720145167

Grade:
graduate

December 14, 2017

Linear Regression, Linear Classification and Gradient Descent

Abstract—The experiments contains Linear Regression and Linear Classification using Gradient Descent to adjust parameters. Through the experiments, we should understand linear regression and gradient descent deeper and then realize the process of optimization and adjusting parameters.

I. Introduction

The experiments are based on python3 and use the following packages which include sklearn,numpy,jupyter and matplotlib. We use the function train_test_split to split the datasets into training set and validation set randomly. We use the train set to adjust the parameters through GD and then use the model we get to validate on our validate dataset. Finally, we get the loss on the train and validate datasets according to the iterations and show it on the figure1 and 2.

II. Methods and Theory

In the experiment of Linear Regression, we first initial a linear model by setting all parameters into zero. The linear model is shown as $\mathrm{f(x)} = \theta * x + b$.
Then we choose the loss function

$$\mathrm{Loss} = (y - f(x))^2$$

to calculate gradient G from all samples. Choose the opposite direction of gradient G as D to update model

$$\theta_t = \theta_{t-1} - \alpha * \mathrm{D}$$

where α stands for the learning rate. The D can be calculated as

$$\frac{\partial L}{\partial \theta} = (f(x) - y) * x$$

$$\frac{\partial L}{\partial b} = (f(x) - y)$$

We get the loss L_{train} under the training set and $L_{validate}$ by validating under validation set.
Finally we repeat the updating progress of θ and draw the corresponding loss L_{train} , $L_{validate}$ according to the iterations.

In the experiment of Linear Classification which is also called SVM, the loss function is defined as

$$Loss = \frac{\|\theta\|^2}{2} + C * max(0, 1 - y(\theta * x + b)).$$

The D can be calculated as

$$\frac{\partial L}{\partial \theta} = \begin{cases} 0 & 1 - y * f(x) \leq 0 \\ -c * x * y & 1 - y * f(x) > 0 \end{cases}$$

$$\frac{\partial L}{\partial b} = \begin{cases} 0 & 1 - y * f(x) \leq 0 \\ -c * y & 1 - y * f(x) > 0 \end{cases}$$

III. Experiment

A. Dataset

The Housing in LIBSVM Data which contains 506 samples with 13 features is used in the experiment of Linear Regression.

The Australian in LIBSVM Data which contains 690 samples with 14 features is used in the experiment of Linear classification.

B. Implementation

- Linear Regression

  We have a 100 iteration and the learning rate is defined as 0.01 and the bias is set to 0.1. All θ are set to 0 in the beginning.
  The Loss on train and validate is shown below:

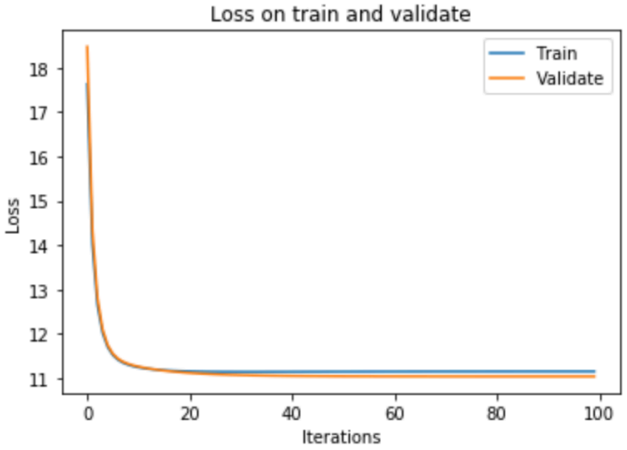

Figure 1

- Linear Classification

  We have a 1000 iteration and the learning rate is defined as 0.01 and the bias is set to 0.01.c is set to 0.01. All θ are set to 0 in the beginning.
  The Loss on train and validate is shown below:

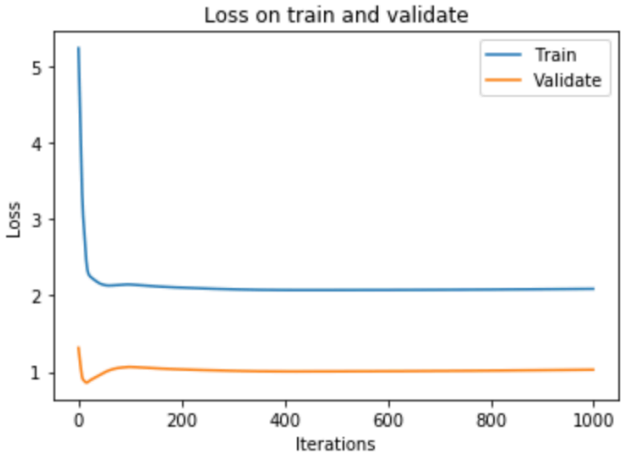

Figure 2

IV. CONCLUSION

From the experiments we can know the detail of the progress of Gradient Decent to adjust the parameters and we can get an intuitive feeling through the loss. The loss on both training set and validate set decrease and tend to converge with the adjusted parameters. We can set different hyper-parameters to get different output loss images. Proper hyper-parameters are very import in the progress of GD.