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:
Zilong Li

Supervisor:
Qingyao Wu

Student ID：
201721045398

Grade:
Graduate

December 14, 2017

Linear Regression, Linear Classification and Gradient Descent

Abstract

We implemented a Linear regression and a Linear classification algorithm ,which get very low loss in the LIBSVM Housing and Australian data.

I. INTRODUCTION

In statistics, linear regression is a linear approach for modeling the relationship between a scalar dependent variable y and one or more explanatory variables (or independent variables) denoted X. The case of one explanatory variable is called simple linear regression. For more than one explanatory variable, the process is called multiple linear regression.(This term is distinct from multivariate linear regression, where multiple correlated dependent variables are predicted, rather than a single scalar variable.)

In linear regression, the relationships are modeled using linear predictor functions whose unknown model parameters are estimated from the data. Such models are called linear models.Most commonly, the conditional mean of y given the value of X is assumed to be an affine function of X; less commonly, the median or some other quantile of the conditional distribution of y given X is expressed as a linear function of X. Like all forms of regression analysis, linear regression focuses on the conditional probability distribution of y given X, rather than on the joint probability distribution of y and X, which is the domain of multivariate analysis.

Linear regression was the first type of regression analysis to be studied rigorously, and to be used extensively in practical applications.This is because models which depend linearly on their unknown parameters are easier to fit than models which are non-linearly related to their parameters and because the statistical properties of the resulting estimators are easier to determine.

Linear regression has many practical uses. Most applications fall into one of the following two broad categories:

1.If the goal is prediction, or forecasting, or error reduction, linear regression can be used to fit a predictive model to an observed data set of y and X values. After developing such a model, if an additional value of X is then given without its accompanying value of y, the fitted model can be used to make a prediction of the value of y.

2.Given a variable y and a number of variables X1, ..., Xp that may be related to y, linear regression analysis can be applied to quantify the strength of the relationship between y and the Xj, to assess which Xj may have no relationship with y at all, and to identify which subsets of the Xj contain redundant information about y.

II. METHODS AND THEORY

Given a data set $\{y_i, \boldsymbol{x_i}\}_{i=1}^{n}$ of n n statistical units, a linear regression model assumes that the relationship between the dependent variable y_i and the p-vector of regressors $x_i$is linear.This relationship is modeled through a disturbance term or error variable εi — an unobserved random variable that adds noise to the linear relationship between the dependent variable and regressors. Thus the model takes the form

$$y_i = \theta_0 1 + \theta_1 x_{i1} + \cdots + \theta_p x_{ip} + \varepsilon_i$$

Which can also written in the function form:

$$y_{\mathrm{i}} = \theta_0 + \sum_{i=1}^{p} \theta_i x_i + \varepsilon_i = \theta_0 + \boldsymbol{\theta_1 x} + \varepsilon_i$$

So, we can get the model as follow:

$$h_\theta(x) = \theta_0 + \boldsymbol{\theta_1 x}$$

Usually,we call $\boldsymbol{x}$ as "feature" ,*h(x)* as "hypothesis"

III. EXPERIMENT

So, to get the "correct" $\theta_0 + \boldsymbol{\theta}_1$,We used the batch Gradient Descent Algorithm. The Loss function is as follow:

$$L(\theta_0, \theta_1) = \frac{1}{2m}\sum_{i=1}^{m}(\mathrm{h}_\theta(x^i) - y^i)^2$$

And the gradient function is as bellow:

$$Grad(\theta_0) = \frac{1}{m}\sum_{i=1}^{m} \mathrm{h}_\theta(x^i) - y^i$$

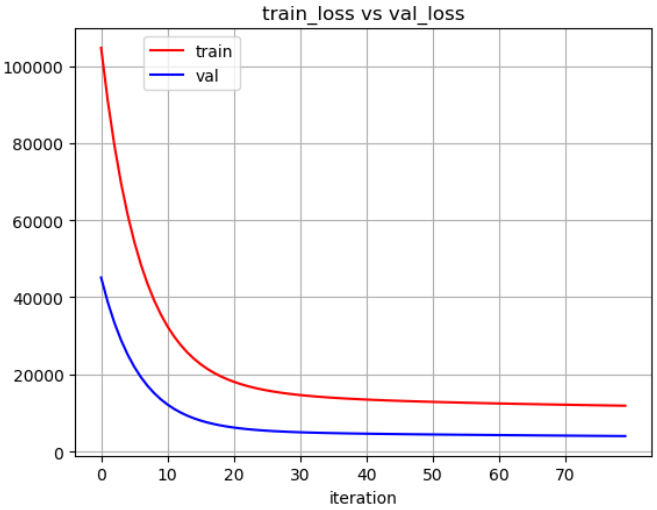
$$Grad(\theta_1) = \frac{1}{m}\sum_{i=1}^{m} (\mathrm{h}_\theta(x^i) - y^i) * x^{(i)}$$

So,We update our linear regression as follow:

$$\theta_0 := \theta_0 - \alpha * Grad(\theta_0)$$
$$\theta_1 := \theta_1 - \alpha * Grad(\theta_1)$$

In our experiment,we set the initial value of θ as all zeros, and the learning rate is 0.005.We conduct the regression experiment in the housing data set,and separate the traning data set and the vilification as 2:1.

The is the training loss and validation vs iterations:



What's more ,we also conducted a Linear Classification algorithm. The loss function is as follows:

$$L(\theta_0,\theta_1)=\begin{cases}\frac{1}{2m}\sum_{i=1}^{m}(\mathrm{h}_\theta(x^i)-y^i)^2, if\ \mathrm{h}_\theta(x^i)*y^i<0\\ 0, if\ \mathrm{h}_\theta(x^i)*y^i\geq 0\end{cases}$$

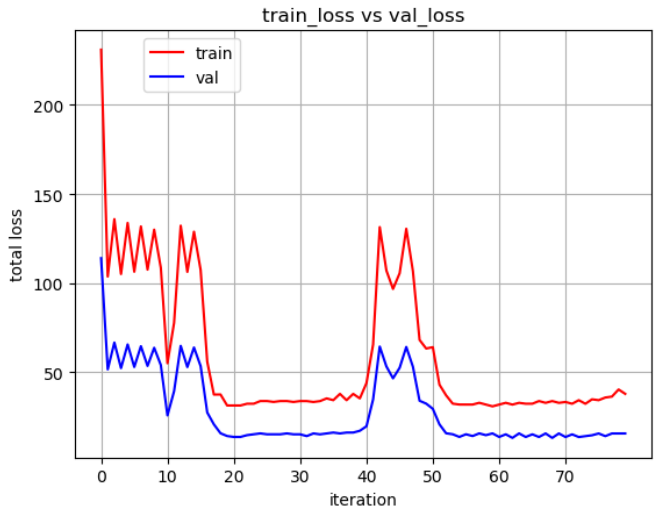
$$Grad(\theta_0)=\begin{cases}\frac{1}{m}\sum_{i=1}^{m}\mathrm{h}_\theta(x^i)-y^i, if\ \mathrm{h}_\theta(x^i)*y^i<0\\ 0, if\ \mathrm{h}_\theta(x^i)*y^i\geq 0\end{cases}$$

$$Grad(\theta_1)=\begin{cases}\frac{1}{m}\sum_{i=1}^{m}(\mathrm{h}_\theta(x^i)-y^i)*x^{(i)}, if\ \mathrm{h}_\theta(x^i)*y^i<0\\ 0, if\ \mathrm{h}_\theta(x^i)*y^i\geq 0\end{cases}$$

We conduct the experiment in the

In our experiment,we set the initial value of θ as all zeros, and the learning rate is 0.005.We conduct the classification experiment in the Australian data set,and separate the traning data set and the vilification as 2:1.

The is the training loss and validation vs iterations:



IV. CONCLUSION

We conducted two linear algorithm ,and got very good results. The linear model is very easy and convenient to conduct and it's widely used in many aspects of both CS and finance field. But to the truth ,its still a weak classifier which cant distinguish hard example