Linear Regression Lab Report

WangXinyi

College of Software Engineering,ChongQing University

Abstract

Linear Regression is probably one of the most frequently used model in machine learning nowadays. 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.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.

Keywords:Linear Regression, statistics

1 Introduction

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

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.

Given a variable y and a number of variables $X_1, ..., X_p$ that may be related to y, linear regression analysis can be applied to quantify the strength of the relationship between y and the X_j, to assess which X_j may have no relationship with y at all, and to identify which subsets of the X_j contain redundant information about y.

Linear regression models are often fitted using the least squares approach, but they may also be fitted in other ways, such as by minimizing the ”lack of fit” in some other norm (as with least absolute deviations regression), or by minimizing a penalized version of the least squares loss function as in ridge regression (L2-norm penalty) and lasso (L1-norm penalty). Conversely, the least squares approach can be used to fit models that are not linear models. Thus, although the terms ”least squares” and ”linear model” are closely linked, they are not synonymous.

2 Algorithm

2.1 Linear Regression Introduction

2.1.1 Model

linear:hθ(x)=θ_0+θ_1*x_1+θ_2*x_2+θ_3*x_3+θ_4*x_4+...+θ_n*x_n

now all the θ are scalar value

polynomial:hθ(x)=θ_0+θ_1*X+θ_2*X^2+θ_3*X^3+θ_4*X^4+...+θ_n*X^n

now θ_0 is scalar value,the rest θ are vectors

2.1.2 CostFunction

J(θ_0,θ_1,θ_2,...,θ_n)=$\frac{1}{m}\sum_{i=1}^{m}$(h$\theta$($x^{(i)}$)-$y^{(i)}$)^2

2.2 Gradient Descent Introduction

gradient descent formula:

θ_0=θ_0-$\alpha\frac{1}{m}\sum_{i=1}^{m} h\theta$($x^{(i)}$)-$y^{(i)}$)

θ_1=θ_1-$\alpha\frac{1}{m}\sum_{i=1}^{m} h\theta$($x^{(i)}$)-$y^{(i)}$)$x^{(1)}$

θ_2=θ_2-$\alpha\frac{1}{m}\sum_{i=1}^{m} h\theta$($x^{(i)}$)-$y^{(i)}$)$x^{(2)}$

and so on

θ_n=θ_n-$\alpha\frac{1}{m}\sum_{i=1}^{m} h\theta$($x^{(i)}$)-$y^{(i)}$)$x^{(n)}$

matters need attention:gradient update must be done **simultaneously**:

right:

temp0=θ_0-$\alpha\frac{1}{m}\sum_{i=1}^{m} h\theta$($x^{(i)}$)-$y^{(i)}$)

temp1=θ_1-$\alpha\frac{1}{m}\sum_{i=1}^{m} h\theta$($x^{(i)}$)-$y^{(i)}$)$x^{(1)}$

θ_0=temp0

θ_1=temp1

wrong:

temp0=θ_0-$\alpha\frac{1}{m}\sum_{i=1}^{m} h\theta$(x$^{(i)}$)-y$^{(i)}$)

θ_0=temp0

temp1=θ_1-$\alpha\frac{1}{m}\sum_{i=1}^{m} h\theta$(x$^{(i)}$)-y$^{(i)}$)x$^{(1)}$

θ_1=temp1

2.3 Advantage and Disadvantage of Linear Regression

Linear regression is perhaps the easiest model we have ever learned,but in fact it plays an important role in machine learning.It can be easily calculated and it provides an acceptable result,it appears in many places such as a part of logistic regression.However,if the model is too simple,it may not fit the training set well.Think of a straight line trying to fit a complex training set,this won't give us a good result,thus effort should be put into finding an appropriate complexity of parameters.

3 Experiment

In this part we will demonstrate how the algorithms are implemented and the result of the experiments.

3.1 Algorithm Description

input:train Set,test Set,model Dimension

output: train Set error,test Set error,Iteration number,parameter

initialization:input train Set,test set and test label, save them into different lists step 1:observe the data set,run gradient descent and decide whether it needs feature scaling,in our data set,the answer is yes,during this step an appropriate learning rate should be decided

step 2:run gradient descent on train set,decide whether we need polynomial model or not depending on the error,in this experiment we simply define polynomial model by just adding x^2,x^3,...,x^n terms instead of adding mixed terms such as x_1x_2,x_1x_2 and so on

step 3:run gradient descent on train set and and use the parameters on test set, in this step regularization rate should be decided

step 4:run different model and compare train error and test error of different situation

3.2 Experiment

Be aware of that we consider the gradient has converged when the change of cost function during one iteration fall bellow 10^{-4},the maximum iteration we set in our program is 5000.Learning rate won't make a big difference on the final result if chosen in an appropriate range.Since our gradient had converged in all situations,we didn't record all the learning rate α

As we can see from the table,model complexity and regularization rate have a great influence on the final result of our experiment.With the regularization rate set to 0,model tends to perform better on train set with the rise of complexity.However,the performance on test set doesn't go the same,complex model may lead to over-fitting,which is to say while the more complex model performs better on train set,it may have a poor performance on the test set(for example,our 3-dimension model have a better performance on train set than the 2-dimension

Table 1: łŁ

Model Dim	Regularization Rate	Iterations	Train Error	Test Error
1	0	383	11.149391	12.42138
2	0	1744	10.315140	11.65928
2	0.3	1388	13.349955	11.71665
2	0.1	1604	11.349354	11.66521
2	0.01	1729	10.419665	11.65898
3	0	1182	10.298759	11.65987
3	0.3	1021	13.236707	11.64659
3	0.1	1120	11.290291	11.64429
3	0.01	1175	10.398558	11.65771

model.But it has a contrary result when it comes to the test set).To deal with over-fitting,we introduce a regularization parameter consisting of the quadratic sum of the parameters θ.The regularization parameter is also called a penalized term,it penalize parameter θ,preventing model from becoming too complex,which in the end may lead to over-fitting.Regularization parameter may cause the model become easier,which means the model may not fit the train set well,but it may have a better generalization performance.However,too large regularization rate will penalize the model too much,causing it become too simple,this won't lead to a good result,either in train set or test set.Thus,choosing a suitable regularization rate is of great importance.

4 Conclusion

In this report we mainly introduced the 2 factors that may have great impact on the result of linear regression.We gave the main steps of our experiment on linear regression and analyzed the result.Based on the analyze on different parameters we drew the conclusion that in order to obtain an acceptable result we must keep a good tradeoff between bias and variance.