The lab report of the Multiple regression model

Zhang Junwei
eMail: 474650206@qq.com

Nov 19th, 2017

Abstract: In regression analysis,if there are two or more than two independent variables,it is called multiple regression. In fact,a phenomenon often is associated with many factors,the optimal combination of multiple variable to predict or estimate the dependent variable,compared with a variable forecast or estimates is more effective,more practical,so multiple linear regression than a linear regression of greater practical significance. In practical problems,a variable is often affected by multiple variables. This paper introduces the idea of multiple linear regression model and polynomial regression model.
Keywords: multiple linear regression, Polynomial regression, Machine learning, statistical learning method

1 Introduction

In the regression analysis, if there are two or more than two independent variables, it is called multiple regression.regression of a dependent variable with two or more than two independent variables multiple linear regression. Multivariate regression including multiple linear regression and polynomial regression, linear regression is the polynomial regression when the highest index is 1.

this paper mainly introduces the polynomial regression, by adjusting the parameters in the polynomial regression, obtain optimal fitting curve of NACA0012 airfoil data.

The rest of the chapter is organized as follows. Section 2 discusses the the establishment of multiple regression equations and polynomial regression, and provide a case study of one such combined approach by using the NACA0012 airfoil data. Finally, we determine the most suitable parameters of data fitting, through the comparison of some evaluation standards.

2 The algorithm method

2.1 Multiple regression model

The simplest definition of regression is that a point set D is given, and a function is used to fit the set of points, and the error between the point set and the fitting function is minimized. If the function curve is a straight line, then the linear regression is called linear regression.

Assuming that the function relationship between the predicted value and the sample feature is linear, the task of regression analysis is to estimate the function H according to the observed values of the sample X and Y, and seek the approximate functional relationship between variables. Definition:

$$h_\theta(x) = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \cdots + \theta_n x_n$$

Among them, n is the characteristic item, and x is the value of the j characteristic of each training sample.It can be considered that the j value of the vector is, and the x_0 is 1, then the multivariate linear regression can be regarded as the value of the item:

$$h_\theta(x) = \theta^T x$$

The solution of the parameters can be introduced into the loss function by using the gradient descent algorithm.

$$J(\theta_0, ..., \theta_n) = \frac{1}{2m} \sum (h_\theta(x^{(i)})) - y^{(i)})^2$$

For the prediction value of the corresponding sample $x^{(i)})$, it subtracts the $y^{(i)})$ in the sample, and the upper form can be regarded as the distance between the predicted value and the true value. So we minimize this distance, which makes our model closer to our given training data. Therefore, our objective function is:

$$minJ(\theta_0, ..., \theta_n)$$

Gradient descent algorithm:
repeat until convergence

$$\theta_j := \theta_j - \alpha \frac{\partial}{\partial \theta_j} J(\theta_0, ..., \theta_n)(for\ j = 0\ and\ j = 1)$$

2.2 Common problems

2.2.1 Feature scaling

Normalized features similar to the zoom feature, which is unique to the multivariate sample, because the sample has a number of features, and each feature dimension is not exactly the same as the maximum value and the minimum value of the range is different, so we need to put them into a unified range. Specifically, for one of the features, we extract all the values of the sample in the feature dimension and compute the mean and standard values. Then use the following formula to normalize:

$$X_n = \frac{X_n - \mu_n}{S_n}$$

μ_nWhich is the mean, $s_n$is the standard deviation

2.2.2 Learning rate

The learning rate in the gradient descent method determines the extent to which it drops after the fastest descent direction is found. It can be said that the gradient descent method, this kind of omnipotent solution to the optimization of the players received the impact of two factors, in addition to the initial point settings, there is one of it. To solve the effect of gradient descent learning rate influenced by size, if the alpha is too small, so the convergence speed is slow, the number of iterations required for many; if alpha is very large, it may make the update of the time over the local minimum, leading to convergence.

2.2.3 Regularization

When the value of the polynomial number is large, if the blind pursuit of improving the ability to predict the training data, the complexity of the selected model is often higher than the real model, there will be over fitting phenomenon, cause this phenomenon not limited training data fully reflects a good model. So we use regularization to preserve all the characteristic variables, which can reduce the order of the characteristic variables.

2.3 Polynomial regression model

In statistics, polynomial regression is a form of regression analysis in which the relationship between the independent variable x and the dependent variable y is modelled as an nth degree polynomial in x. Polynomial regression fits a nonlinear relationship between the value of x and the corresponding conditional mean of y,and has been used to describe nonlinear phenomena.Polynomial regression fits a nonlinear model to the data, as a statistical estimation problem it is linear. The polynomial regression is considered to be a special case of multiple linear regression.

$$h_\theta(x) = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \cdots + \theta_n x_n$$

order:$x_1 = (size), ..., x_n = x_n = (size)^n$

$$h_\theta(x) = \theta_0 + \theta_1(size) + \theta_2(size)^2 + \cdots + \theta_n(size)^n$$

The gradient descent and learning rates of polynomial regression models are similar to those of multivariate linear regression models.

3 Experiment

In this part, we use the NACA0012 airfoil data set to select the batch gradient model and the loss function as the objective function. The polynomial model uses regularization term, and the minimum loss function is selected by adjusting the parameters.

3.1 mathematical model

Multiple liner regression model

Input:Training data set, test data set, training label set, learning rate
Output:Parameter, test set label, minimum loss function
Initialization:The data of the training datasets are processed by feature scaling, and the training data set is scaled from -1 to 1.
repeat:The loss of the tag vector and the training tag is calculated, and the parameters are calculated by the gradient descent algorithm.
Calculation:minimum loss function

Polynomial regression model

Input:Training data set, test data set, training label set, learning rate,L1 paradigm, L2 paradigm
Output:Parameter, test set label, minimum loss function
Initialization:The data of the training datasets are processed by feature scaling, and the training data set is scaled from -1 to 1.
repeat:The loss of the tag vector and the training tag is calculated, and the parameters are calculated by the gradient descent algorithm.
Calculation:minimum loss function

3.2 experimental data

The NASA data set comprises different size NACA 0012 airfoils at various wind tunnel speeds and angles of attack. The span of the airfoil and the observer position were the same in all of the experiments.The data includes frequency, in hertzs, angle of attack, in degrees, chord length, in meters, free-stream velocity, in meters per second, suction side displacement thickness, in meters. The training dataset has 1052 data and the test dataset has 451 data

This is the table 1 that the minimum loss function varies with the change of learning rate for multiple liner regression model.

Rate	min loss function
0.01	11.1444
0.05	11.1442
0.1	11.1441

This is the table 2 that the minimum loss function varies with the change of learning rate for polynomial regression model.That the L1 paradigm is 0.0001, L2 paradigm is 0.0001, degree is 2.

Rate	min loss function
0.001	18.1549
0.005	18.1423
0.01	18.1410

This is the table 3 that the minimum loss function varies with the change of learning rate for polynomial regression model.That the rate is 0.01, degree is 2.

Regularization	min loss function
0.001	18.2607
0.005	18.14077
0.01	18.14076

This is the table 4 that the minimum loss function varies with the change of learning rate for polynomial regression model.That the rate is 0.01, the L1 paradigm is 0.001, L2 paradigm is 0.001.

degree	min loss function
1	18.37853657
2	18.2607
3	18.4686

4 Summary

The regularization term adjusts the parameters of the linear regression model and the polynomial regression model, and the optimal parameters are obtained by looking at the minimum loss function to predict the labels of the test set.

Through experiments, we can get that in the linear regression model, the learning rate has an impact on the size of the loss function. In the polynomial regression model, the learning rate, the size of the normal form and the exponent of the polynomial can also affect the loss function.