Least squares method is a mathematical optimization technique. It seeks the best functional match of data by minimizing the sum of squares of errors. The unknown data can be easily obtained by using the least square method, and the sum of squares of the errors between the obtained data and the actual data is minimized. The least squares method can also be used for curve fitting. Other optimization problems can also be expressed by least square method by minimizing energy or maximizing entropy.

The theoretical background of the least square method is solving a set of overdetermined linear equations. The solution we want to find is the model coefficients. The model is

We need to find out the best set of to fit the curve. The process is basically solving the overdetermined function:

The coefficient for the fitting curve is generalized by solving the function with the help of backlash. A matrix contains the coefficient for different basis functions and the regularization parameter . B matrix has the values for these functions. In this case, B matrix should be the set of y value for the data points.

We set up one array to store the x value for the data which is called . Based on this array, we create a new matrix :

Then we will splice a column matrix which stand for the coefficient for that is all 1 with b adduction from the left.

Then we create a new matrix whose the main diagonal is :

Matrix A is the combination of Matrix c and Matrix b from vertical direction:

文本, 信件

描述已自动生成

Picture 1. Composition of A matrix

Meanwhile, we need to deal with the y values of the data points as well. We set up one array to store the y value for the data which is called . To carry out the matrix calculation, should be an matrix. So we need to provide enough zeros for it:

So by calculation:

We can find out the solution for this least square problem.

Cross-validation method is used to find out the best for the regularization. We first divide the LargeData into five parts evenly and choose one part as the Test Data. The rest four part are used as Training Data. The error is calculated and recorded after the training and testing. Then, another part is chosen to be the test data, and the rest are for training. In the end, every part t has been used as the test data. Then we take the average value for each Lambda which should have five different Test Error for each.