

# LinearRegression

This class implements different linear regression models, using the least squares method to estimate the regression coefficients. Given input data  $x_{i,j}$  and response  $y_i$ , we want to find coefficients  $\beta_j$  that minimize the residuals of the form

$$\sum_i \left( y_i - \beta_0 - \sum_{j=1}^k \beta_j x_{i,j} \right)^2.$$

Using matrix notation, one may write this equation as

$$\min_{\beta} \|Y - X\beta\|_2,$$

where the first column of  $X$  is all 1, and the  $L_2$  norm is used.

Sometimes, one wants to use a basis of general functions  $\psi_j(t)$  with a minimization of the form

$$\sum_i \left( y_i - \sum_{j=1}^k \beta_j \psi_j(t_i) \right)^2.$$

For example, we could have  $\psi_j(t) = e^{-\lambda_j t}$  or some other functions. In that case, one has to choose the points  $t_i$  at which to compute the basis functions, and use a method below with  $x_{i,j} = \psi_j(t_i)$ .

```
package regres;
```

```
public class LinearRegression
```

## Methods

```
public static double[] calcCoefficients (double[] X, double[] Y)
```

Computes the regression coefficients using the least squares method. This is a simple linear regression with only 2 regression coefficients. The model is

$$y = \alpha + \beta x.$$

Given the  $n$  data points  $(X_i, Y_i)$ ,  $i = 0, 1, \dots, (n-1)$ , the method computes and returns the array  $[\alpha, \beta]$ .

```
public static double[] calcCoefficients0 (double[][] X, double[] Y)
```

Computes the regression coefficients using the least squares method. This is a model for multiple linear regression. There are  $k+1$  regression coefficients  $\beta_j$ ,  $j = 0, 1, \dots, k$  and  $k$  regressors variables  $x_j$ . The model is

$$y = \beta_0 + \sum_{j=1}^k \beta_j x_j.$$

There are  $n$  data points  $Y_i$ ,  $X_{ij}$ ,  $i = 0, 1, \dots, (n - 1)$ , and each  $X_i$  is a  $k$ -dimensional point. Given the response  $Y[i]$  and the regressor variables  $X[i][j]$ ,  $i = 0, 1, \dots, (n - 1)$ ,  $j = 0, 1, \dots, (k - 1)$ , the method computes and returns the array  $[\beta_0, \beta_1, \dots, \beta_k]$ . Restriction:  $n > k + 1$ .

```
public static double[] calcCoefficients (double[] [] X, double[] Y)
```

Computes the regression coefficients using the least squares method. This is a model for multiple linear regression. There are  $k$  regression coefficients  $\beta_j$ ,  $j = 0, 1, \dots, (k - 1)$  and  $k$  regressors variables  $x_j$ . The model is

$$y = \sum_{j=0}^{k-1} \beta_j x_j.$$

There are  $n$  data points  $Y_i$ ,  $X_{ij}$ ,  $i = 0, 1, \dots, (n - 1)$ , and each  $X_i$  is a  $k$ -dimensional point. Given the response  $Y[i]$  and the regressor variables  $X[i][j]$ ,  $i = 0, 1, \dots, (n - 1)$ ,  $j = 0, 1, \dots, (k - 1)$ , the method computes and returns the array  $[\beta_0, \beta_1, \dots, \beta_{k-1}]$ . Restriction:  $n > k$ .