Module-04, Python for Machine Learning

Regression Algorithms(Linear Regression)

Dostdar Ali
Instructor

Data science and Artificial Intelligence
3-Months Course
at
Karakaroum international Univrsity

February 19, 2024

Table of Contents

Linear Regression Function

- Function

$$f(x_i) = a_1 X_i + a_0 \tag{1}$$

- The objective is to find a function that fits the data points over all.
- Error Equations

$$r_i = y_i - f(x_i) \tag{2}$$

- The overall error can be calculated in different ways.

Calculating Error Methods

One simple way is to add the residuals of all the points

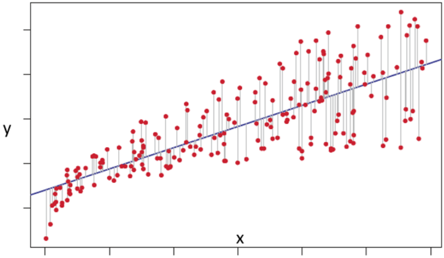
$$E = \sum_{i=1}^{n} r_i = \sum_{i=1}^{n} [y_i - (a_1 X_i + a_0)]$$

Residual

Residual –

The difference between residual the *i*th observed response value and the *i*th response value that is predicted by our linear model is known as residual

$$e_i = y_i - \hat{y}_i$$



Linear least square method

Note

This is not a good choice to sum all : both +ve and -ve residuals both can be large zero or very close to zero.

- The next possibility: Absolute $||.||$

$$E = \sum_{i=1}^{n} r_i = \sum_{i=1}^{n} |y_i - (a_1 X_i + a_0)|$$

Now positive: (total error) but it cannot be used for determining the constants of the function that give the best fit. This is because the measure is not unique.

- Now we apply square on the all over error E is:

$$E = \sum_{i=1}^{n} r_i = \sum_{i=1}^{n} [y_i - (a_1 X_i + a_0)]^2$$

Linear least square method

Note

This is not a good choice to sum all : both +ve and -ve residuals both can be large zero or very close to zero.

- The next possibility: Absolute $||.||$

$$E = \sum_{i=1}^{n} r_i = \sum_{i=1}^{n} |y_i - (a_1 X_i + a_0)|$$

Now positive: (total error) but it cannot be used for determining the constants of the function that give the best fit. This is because the measure is not unique.

- Now we apply square on the all over error E is:

$$E = \sum_{i=1}^{n} r_i = \sum_{i=1}^{n} [y_i - (a_1 X_i + a_0)]^2$$

Linear least square method

Note

This is not a good choice to sum all : both +ve and -ve residuals both can be large zero or very close to zero.

- The next possibility: Absolute $||.||$

$$E = \sum_{i=1}^{n} r_i = \sum_{i=1}^{n} |y_i - (a_1 X_i + a_0)|$$

Now positive: (total error) but it cannot be used for determining the constants of the function that give the best fit. This is because the measure is not unique.

- Now we apply square on the all over error E is:

$$E = \sum_{i=1}^{n} r_i = \sum_{i=1}^{n} [y_i - (a_1 X_i + a_0)]^2$$

Calculate a_1 and a_0

- $$a_1 = \frac{nS_{xy} - S_x S_y}{nS_{xx} - (S_x)^2}$$

  $$a_0 = \frac{S_{xx} S_y - S_{xy} S_x}{nS_{xx} - (S_x)^2}$$

- Where:

  a_1 : Coefficients for the respective independent variables

  a_0 : Intercept

  S_x : Sum of x

  S_y : Sum of y

  S_{xy} : Sum of x and y product

  S_{xx} : Sum of x product itself

Calculate a_1 and a_0

- $$a_1 = \frac{nS_{xy} - S_x S_y}{nS_{xx} - (S_x)^2}$$
$$a_0 = \frac{S_{xx} S_y - S_{xy} S_x}{nS_{xx} - (S_x)^2}$$

- Where:

a_1 : Coefficients for the respective independent variables

a_0 : Intercept

S_x : Sum of x

S_y : Sum of y

S_{xy} : Sum of x and y product

S_{xx} : Sum of x product itself

Example For Linear Regression

Example

We want to predict the atmospheric pressure on given and in between temperature the data for this task as given below,

X_i	T(c^0)	0^0	30^0	70^0	100^0
y_i	P(atm)	0.94	1.05	1.17	1.28

$$S_x = \sum_{i=1}^{4} x_i = 0 + 30 + 70 + 100 = 200$$

$$S_y = \sum_{i=1}^{4} y_i = 0.94 + 1.05 + 1.17 + 1.18 = 4.44$$

Example For Linear Regression

- Similarly we have

$$S_{xx} = \sum_{i=1}^{4} x_i^2 = 15800$$

$$S_{xy} = \sum_{i=1}^{4} (x_i y_i) = 241.4$$

- Putting these values in parameters equations we get,

$$a_1 = \frac{4(241.4) - 200 \times 4.44}{4(15800) - (200)^2} = 0.003345$$

$$a_0 = 0.9428$$

- The Linear Equation is

$$\hat{y}_i = 0.003345X_i + 0.9428.$$

Example For Linear Regression

- Similarly we have

$$S_{xx} = \sum_{i=1}^{4} x_i^2 = 15800$$

$$S_{xy} = \sum_{i=1}^{4} (x_i y_i) = 241.4$$

- Putting these values in parameters equations we get,

$$a_1 = \frac{4(241.4) - 200 \times 4.44}{4(15800) - (200)^2} = 0.003345$$

$$a_0 = 0.9428$$

- The Linear Equation is

$$\hat{y}_i = 0.003345X_i + 0.9428.$$

Example For Linear Regression

- Similarly we have

$$S_{xx} = \sum_{i=1}^{4} x_i^2 = 15800$$

$$S_{xy} = \sum_{i=1}^{4} (x_i y_i) = 241.4$$

- Putting these values in parameters equations we get,

$$a_1 = \frac{4(241.4) - 200 \times 4.44}{4(15800) - (200)^2} = 0.003345$$

$$a_0 = 0.9428$$

- The Linear Equation is

$$\hat{y}_i = 0.003345X_i + 0.9428.$$

Great Job

Thank yo