Introduction to Linear Regression(LR)

Notes of class

Iván Andrés Trujillo Abella

FACULTAD DE INGENIERÍA

---

1 Linear regression

Before that we introduce to linear regression and the derivarion of its properties it is important recall some properties of summatory:

Remeber that;

$$\sum_{i=1}^{n} x_i = x_1 + x_2 + x_3 + ... + x_n \tag{1}$$

$$\sum a = na \tag{2}$$

$$\sum (ax) = a \sum x \tag{3}$$

$$\sum (x_i + y_i) = \sum x_i + \sum y_i \tag{4}$$

$$\sum (ax + by) = a \sum x_1 + b \sum y_1 \tag{5}$$

$$\begin{aligned}
&\sum x_i y_i - n\bar{x}\bar{y} \\
&= \sum x_i y_i - \sum x_i \bar{y} \\
&= x_1 y_1 + x_2 y_2 + ... + x_n y_n - (x_1 + x_2 + ... + x_n) y_n \\
&= x_1 y_1 - x_1 \bar{y} + x_2 y_2 - x_2 \bar{y} + ... + x_n y_n - x_n \bar{y} \\
&= x_1 (y_1 - \bar{y}) + x_2 (y_2 - \bar{y}) + ... + x_n (y_n - \bar{y}) = \sum x_i (y_i - \bar{y}) = \sum y_i (x_i - \bar{x})
\end{aligned} \tag{6}$$

With this properties and examples we could verify that :

$$\sum (x_i - \bar{x})^2 = \sum x_i^2 - n\bar{x}^2 \tag{7}$$

and

$$\sum (y_i - \bar{y})(x_i - \bar{x}) = \sum x_i y_i - n\bar{x}\bar{y} \tag{8}$$

2 OLS

$u_i = y_i - \hat{y}_i$ and $\hat{y}_i = \beta_0 + \beta_1 x_i$ and our goal is :

$$\min \sum (y_i - \hat{y}_i)^2 = \min \sum u_i^2 \tag{9}$$

to solve the problem note that $u_i(\beta_0, \beta_1)$.

$$\frac{\sum u_i^2}{\partial \beta_k} = \frac{\sum u_i^2}{\partial u_i} \frac{\partial u_i}{\partial \beta_k} = 0 \quad k = 0, 1. \tag{10}$$

$$\begin{aligned} \frac{\partial u_i}{\partial \beta_0} &= -2 \sum u_i = 0 \\ \frac{\partial u_i}{\partial \beta_1} &= \sum u_i x_i = 0 \end{aligned} \tag{11}$$

rewriting the equations:

$$\begin{aligned} \sum (y_i - \beta_0 - \beta_1 x_i) &= 0 \\ \sum (y_i - \beta_0 - \beta_1 x_i) x_i &= 0 \end{aligned} \tag{12}$$

applying:

$$\begin{aligned} n\bar{y} - n\beta_0 - n\beta_1 \bar{x} &= 0 \\ \sum x_i y_i - n\beta_0 \bar{x} - \beta_1 \sum x_i^2 &= 0 \end{aligned} \tag{13}$$

now of $n\bar{y} - n\beta_0 - n\beta_1 \bar{x}$ we can get:

$$\beta_0 = \bar{y} - \beta_1 \bar{x} \tag{14}$$

now replace β_0 in $\sum x_i y_i - n\beta_0 \bar{x} - \beta_1 \sum x_i^2$

$$\begin{aligned} \sum x_i y_i - n\bar{x}\bar{y} + n\beta_1 \bar{x}^2 - \beta_1 \sum x_i^2 &= 0 \\ \sum x_i y_i - \bar{x}\bar{y} &= \beta_1 (\sum x_i^2 - n\bar{x}^2) \\ \beta_1 &= \frac{\sum x_i y_i - \bar{x}\bar{y}}{\sum x_i^2 - n\bar{x}^2} \end{aligned} \tag{15}$$

According to the properties defined previously:

$$\beta_1 = \frac{\sum x_i y_i - \bar{x}\bar{y}}{\sum x_i^2 - n\bar{x}^2} = \frac{\sum (y_i - \bar{y})(x_i - \bar{x})}{\sum (x_i - \bar{x})^2} = \frac{cov(x, y)}{var(x)}. \tag{16}$$

We could make own implementations to solve this problem;

3 Multiple linear regression

4 LR gradient descendent