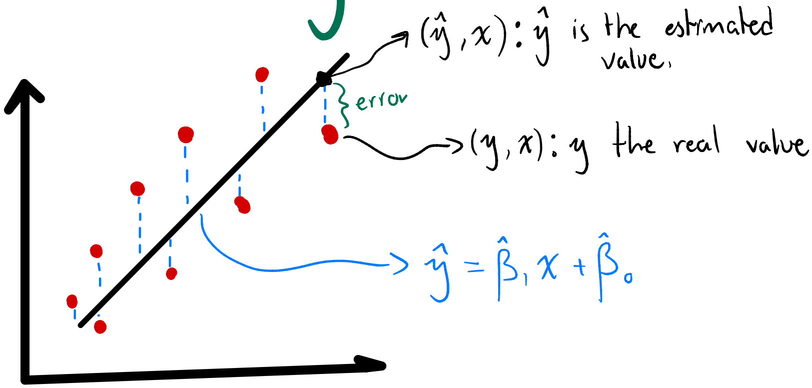
Data Science UCA

Spring 2025

Rafael Dénk

Linear Regression

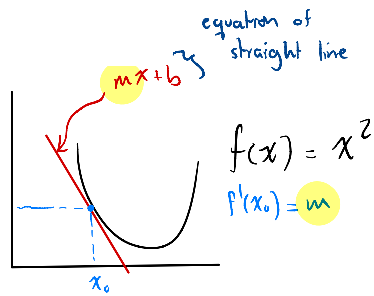


$$\text{error} = y - \hat{y} = e$$

$$e^2 = (y - \hat{y})^2$$

$$RSS = \sum_{i=1}^{n} e_i^2 = \sum (y - \hat{\beta}_1 x - \hat{\beta}_0)^2$$

Residual Sum of Squares



To find the minimum we must derivate and equal the derivative to zero. In other words, the derivative is the slope

We should compute:

$$\frac{\partial RSS}{\partial \beta_0} = 0 \quad (1) \qquad \text{and} \qquad \frac{\partial RSS}{\partial \beta_1} = 0 \quad (2)$$

To know more about Calculus go here

https://youtu.be/WUvTyaaNkzM

From (1)

$$\frac{\partial \sum (y - \hat{\beta}_1 x - \hat{\beta}_0)^2}{\partial \beta_0} = 0$$

$$\rightarrow \sum \frac{\partial (y - \hat{\beta}_1 x - \hat{\beta}_0)^2}{\partial \beta_0} = 0 \rightarrow \sum \cancel{2}(\cancel{-1})(y - \hat{\beta}_1 x - \hat{\beta}_0) = 0$$

Rafael Dávila

$$\sum(y-\hat{\beta}_1 x-\hat{\beta}_0)=0$$

$$\sum y-\hat{\beta}_1\sum x-\sum\hat{\beta}_0=0$$

$$n\bar{y}-n\hat{\beta}_1\bar{x}-n\hat{\beta}_0=0$$

$$\bar{x}=\frac{\sum_{i=1}^{n}x_i}{n}$$

$$\rightarrow n\bar{x}=\sum_{i=1}^{n}x_i$$

$$n(\bar{y}-\beta_1\bar{x}-\hat{\beta}_0)=0$$

$$\bar{y}-\beta_1\bar{x}=\hat{\beta}_0 \quad (3)$$

From (2)

$$\frac{\partial}{\hat{\beta}_1}\sum(y-\hat{\beta}_1 x-\hat{\beta}_0)^2=0$$

$$\sum\frac{\partial}{\partial\beta_1}(y-\hat{\beta}_1 x-\hat{\beta}_0)^2=0$$

$$\sum 2(-x)(y-\hat{\beta}_1 x-\hat{\beta}_0)=0$$

$$\cancel{-2}\sum\left[x(y-\hat{\beta}_1 x-\hat{\beta}_0)\right]=0$$

$$\sum\left[xy-\hat{\beta}_1 x^2-\hat{\beta}_0 x\right]=0$$

$$\sum xy-\hat{\beta}_1\sum x^2-\hat{\beta}_0\sum x=0 \quad (4)$$

From (3) and (4)

$$\sum xy-\hat{\beta}_1\sum x^2-(\bar{y}-\hat{\beta}_1\bar{x})\sum x=0$$

$$\sum xy-\hat{\beta}_1\sum x^2-\bar{y}\sum x+\hat{\beta}_1\bar{x}\sum x=0$$

$$\sum xy-\hat{\beta}_1\sum x^2-n\bar{x}\bar{y}+\hat{\beta}_1 n\bar{x}^2=0$$

$$\sum xy-n\bar{x}\bar{y}=\hat{\beta}_1\sum x^2-\hat{\beta}_1 n\bar{x}^2$$

$$\frac{\sum xy-n\bar{x}\bar{y}}{\sum x^2-n\bar{x}^2}=\hat{\beta}_1$$

$$\hat{\beta}_1=\frac{\sum xy-n\bar{x}\bar{y}}{\sum x^2-n\bar{x}^2}$$

$$\hat{\beta}_1=\frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sum_{i=1}^{n}(x_i-\bar{x})^2}$$

Easy to check from bottom to top

Rafael Dávila