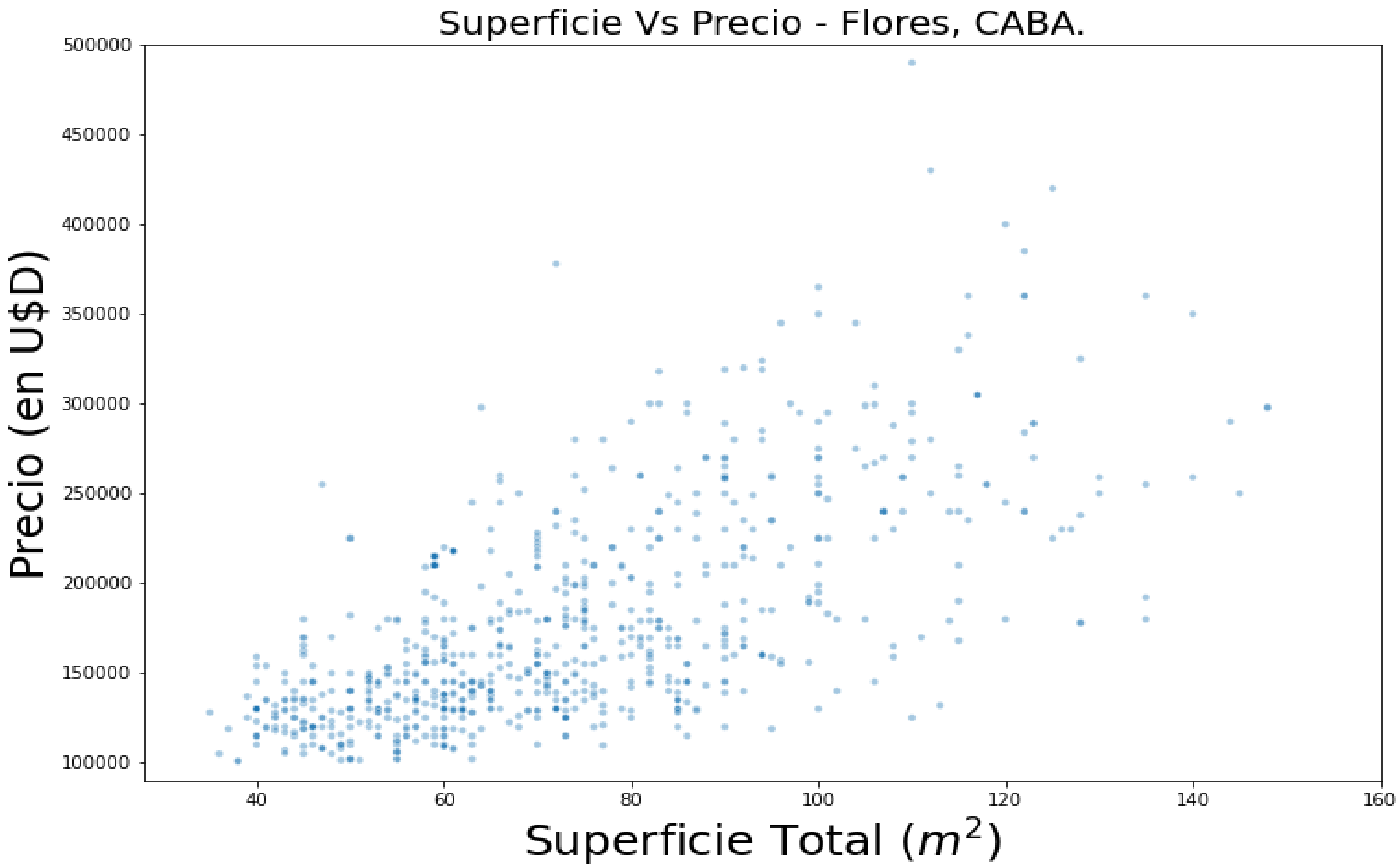
Regresión Lineal Simple

Palazzo Tomás Alejandro

Ciencias de Datos - FCEN - UBA

Motivación: Precios de propiedades



Regresión Lineal Simple

Este modelo trata de explicar la relación lineal entre una variable **dependiente** y con una variable **independiente** x

Nuestros Datos

A partir de ahora

- x = **Superficie Total(m^2)**
- y = **Precio**

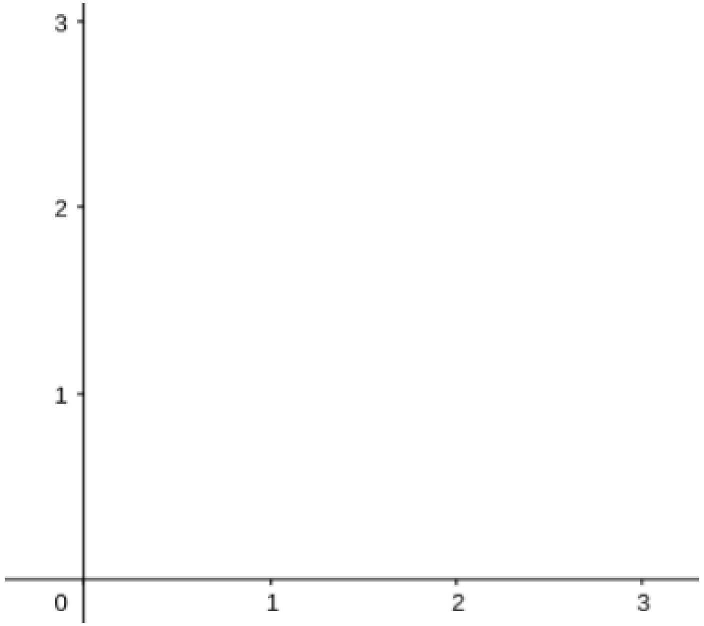
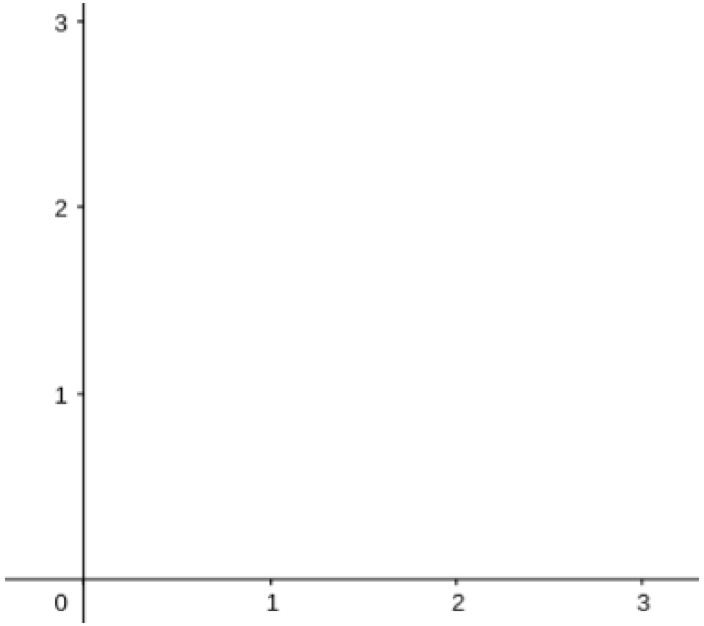
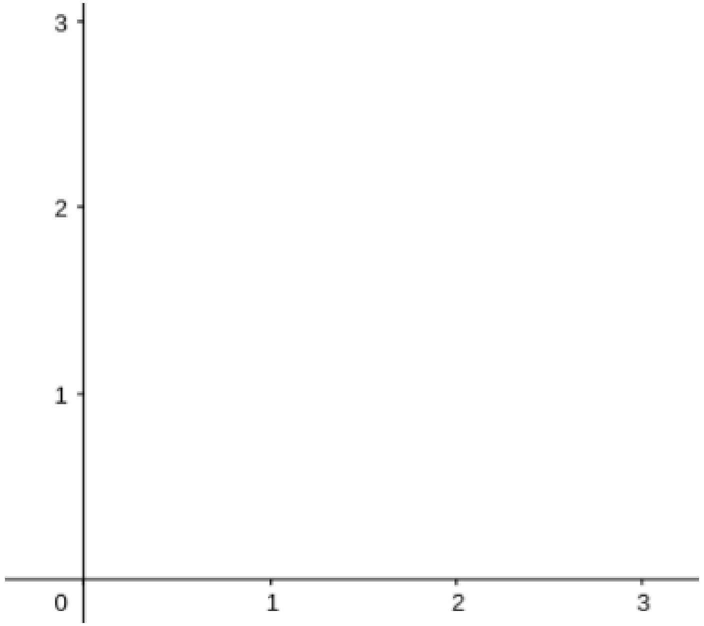
Observemos que $y = (y_1, y_2, ..., y_n)$ y $x = (x_1, x_2, ..., x_n)$. Donde x va a ser nuestra variable **dependiente** e y nuestra variable **dependiente**.

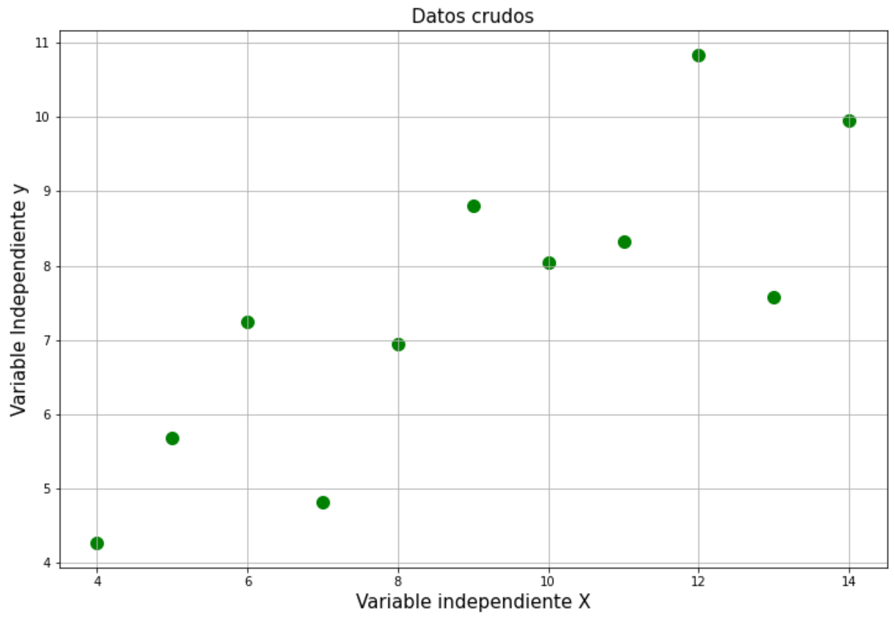
y vamos a llamar a

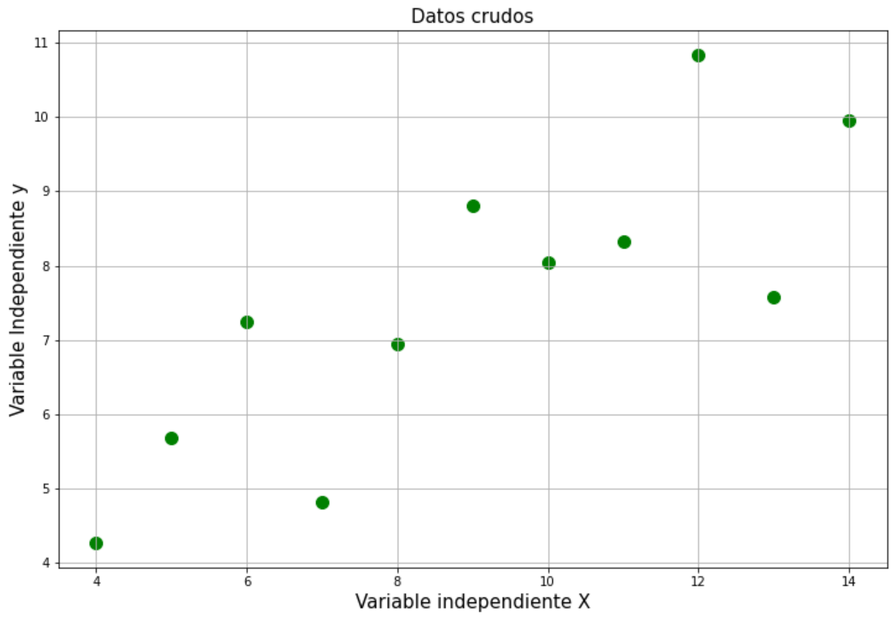
$$\hat{y} = \beta + \alpha x = \hat{precio}$$

- **Idea**: Encontrar los β y α tales que $\hat{y}$ sea lo más parecida posible a y

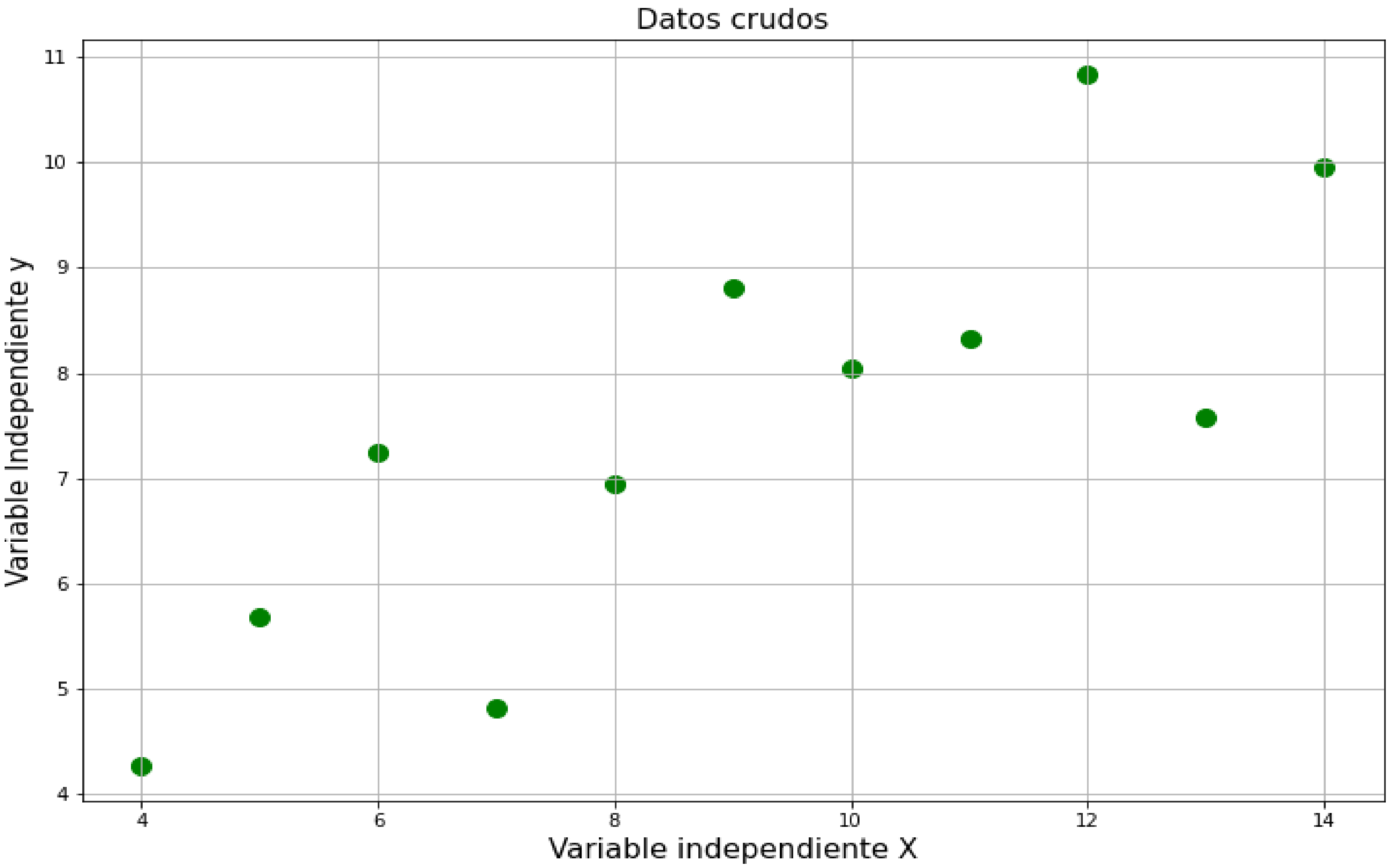
$$\hat{y} = \beta + \alpha x$$

$\alpha = 0, \beta = 2$	$\alpha = 1, \beta = 0$	$\alpha = -1, \beta = 3$
		


Datos crudos
Variable Independiente y
Variable independiente X


Datos crudos
Variable Independiente y
Variable independiente X

Función de Costo



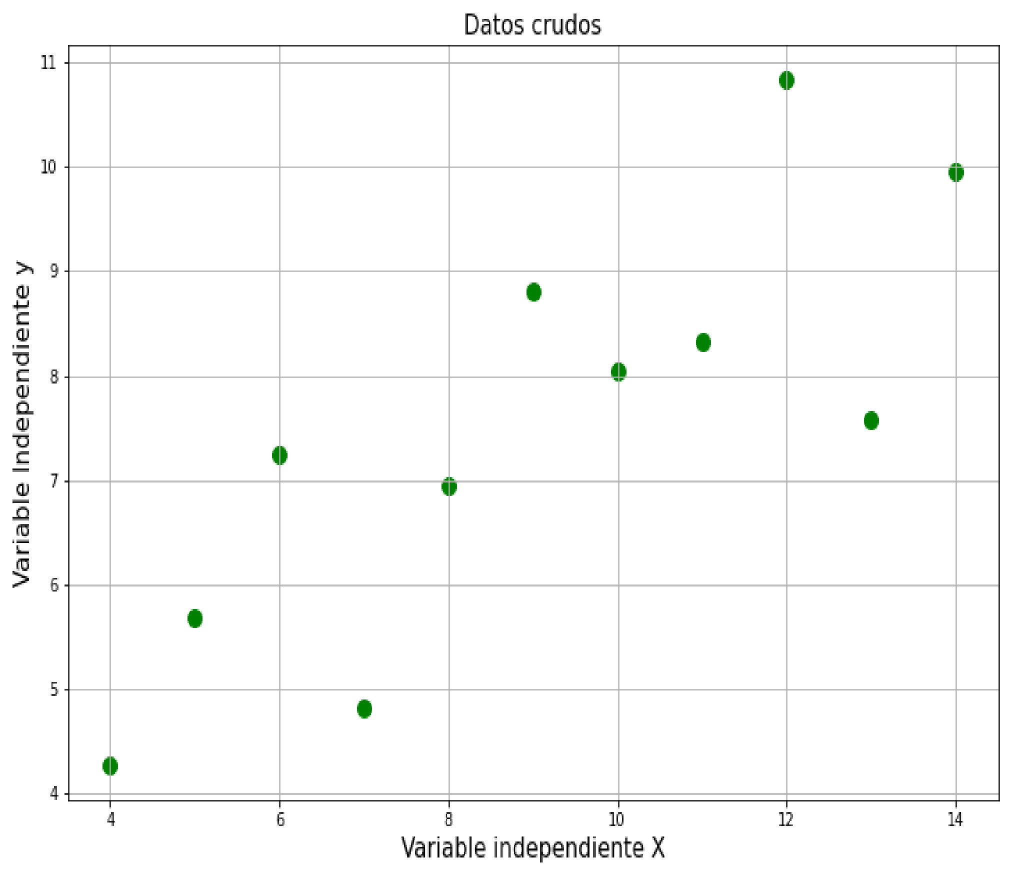
Función de Costo

Lo que tratamos de buscar es que nuestro modelo tenga *el menor error posible*.

Pero, ¿cómo calculamos ese error? Una forma es la siguiente:

$$L(y, \hat{y}) = \frac{1}{2n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$

$$\hat{y} = \beta + \alpha x$$



$$min_{\alpha,\beta} \frac{1}{2n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$

¿Cómo minimizamos el error?

En otras palabras. ¿Cómo calculamos $min_{\alpha,\beta} \frac{1}{2n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$?

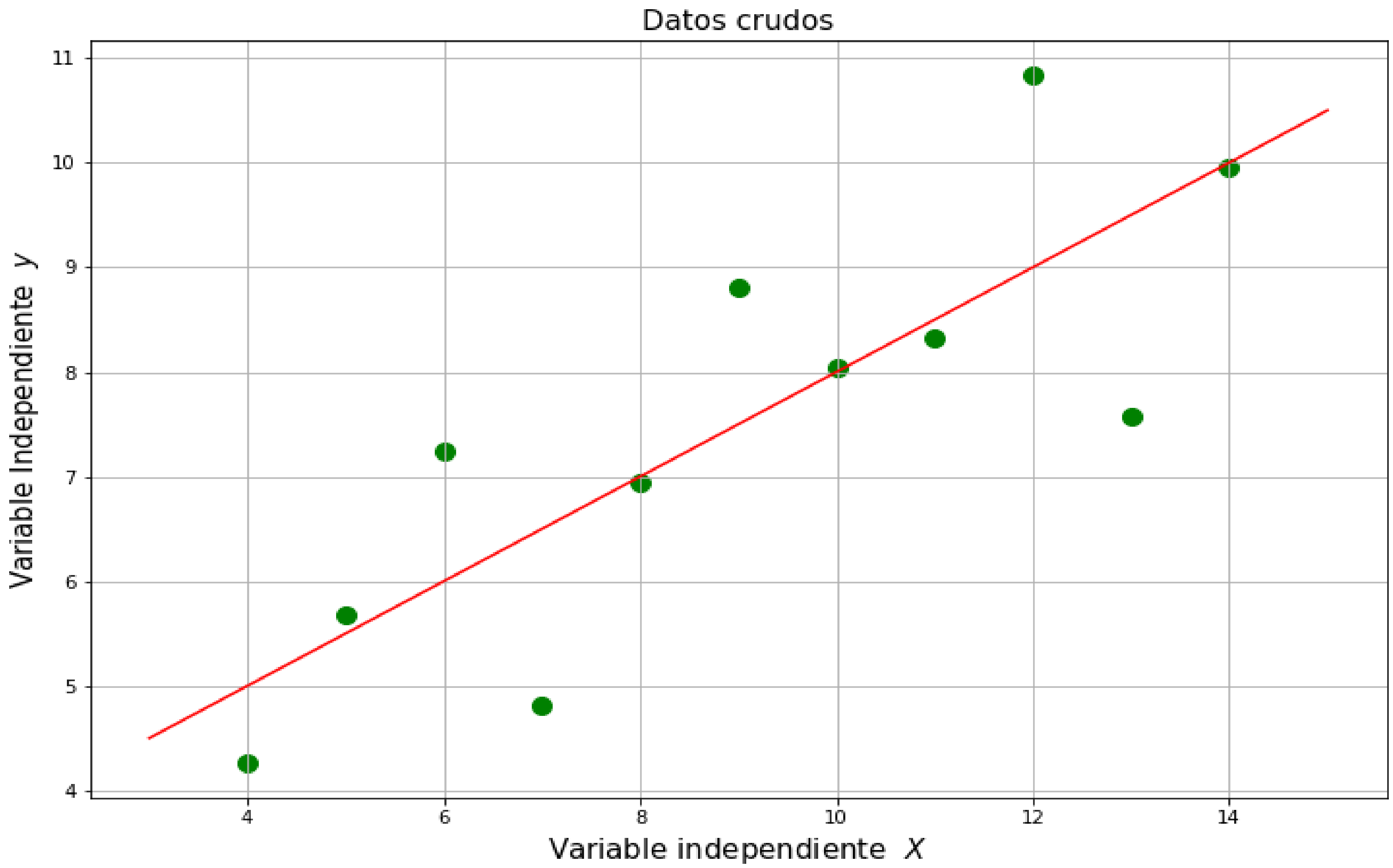
Un método puede ser Gradiente Descendente (no incluido en el video).

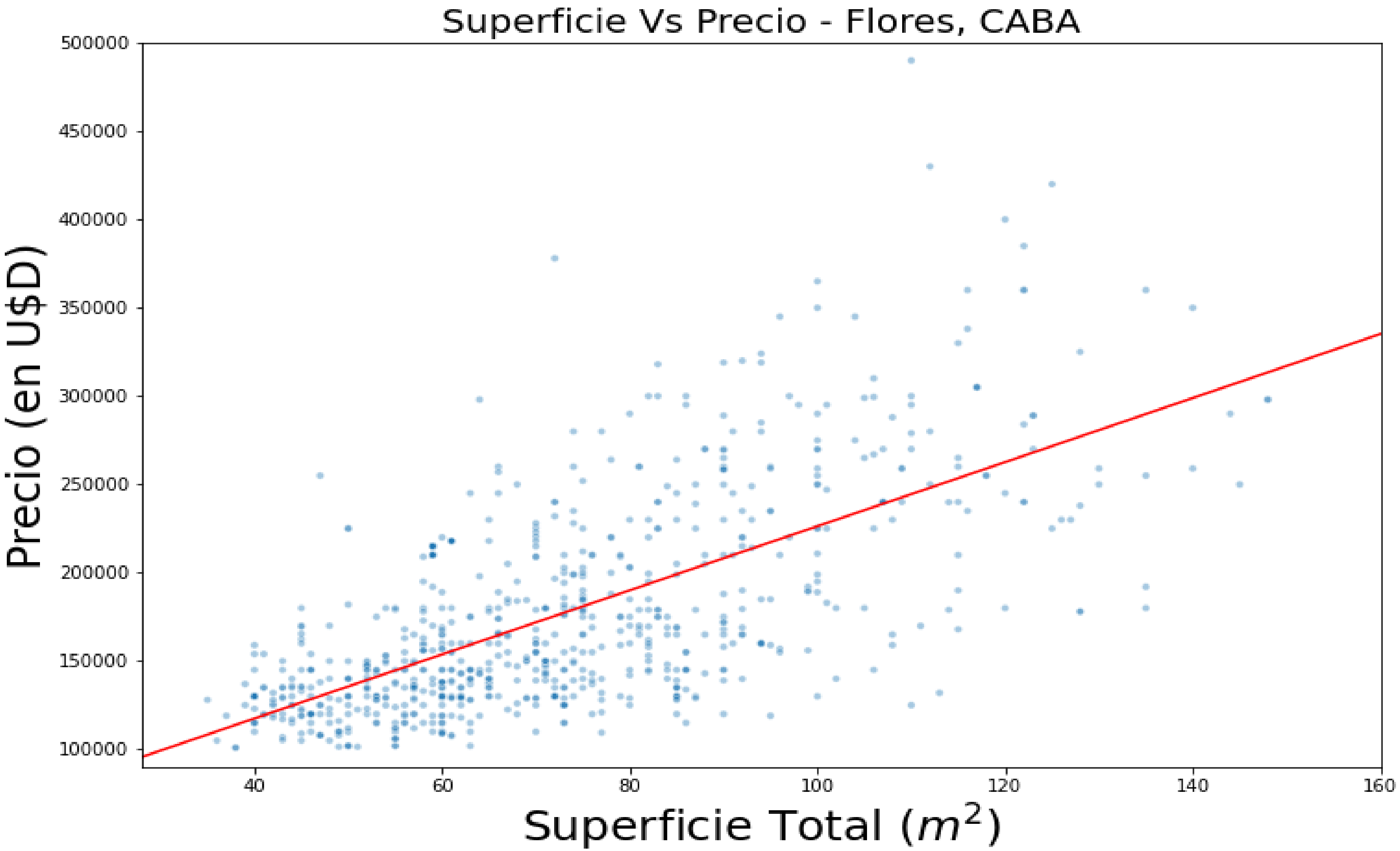
- $\hat{y} = \beta + \alpha x$

- Parámetros: α, β

- Función de Costo: $L(y, \hat{y}) = \frac{1}{2n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$

- Objetivo: $\min_{\alpha,\beta} \frac{1}{2n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$


Datos crudos
Variable Independiente y
Variable independiente X
11
10
9
8
7
6
5
4
4
6
8
10
12
14



Superficie Vs Precio - Flores, CABA
500000
450000
400000
350000
300000
250000
200000
150000
100000
Precio (en U$D)
40
60
80
100
120
140
160
Superficie Total (m^2)