REGRESIÓN LINEAL --> el target, y, es una variable numérica

feature

x	y
2	8
3	10
1	4

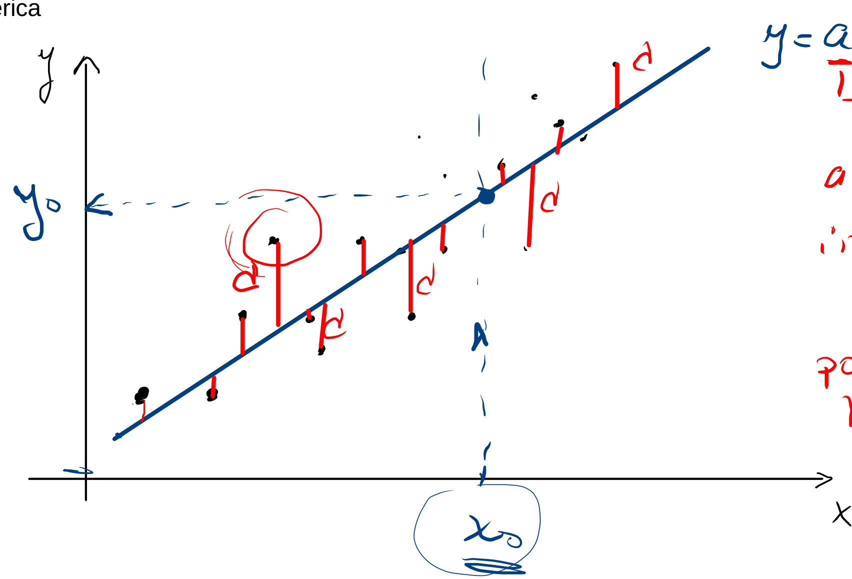
SCATTER PLOT

pendiente

$y = a\,x + b$

ordenada al origen (corte eje vertical)

$y = mx + n$



$y = a\,x + b$

a y b son las incógnitas

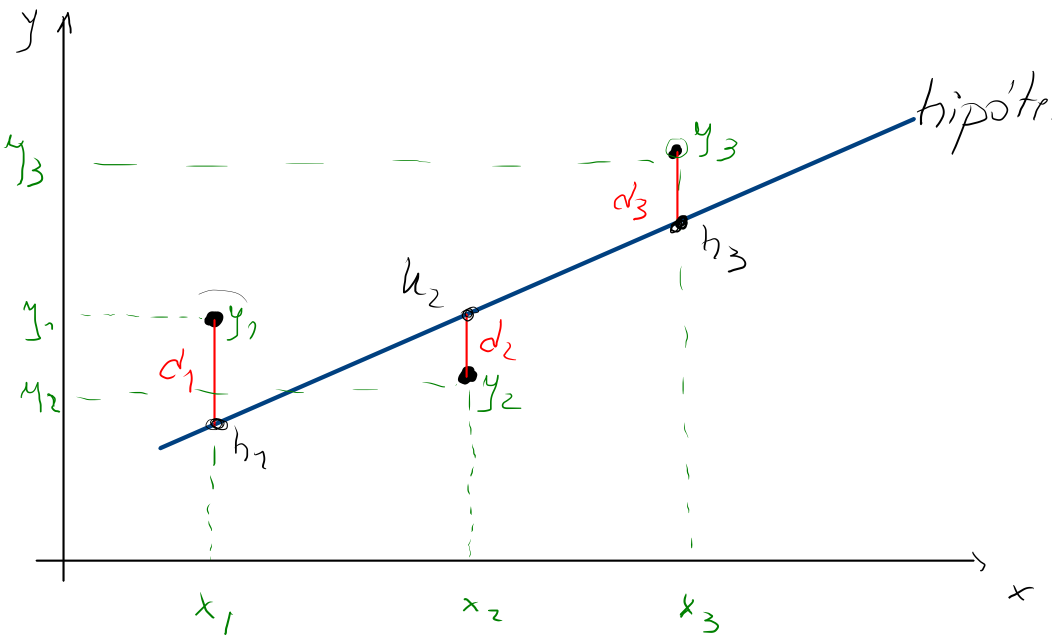
porque ellos definen la ecuación de la recta.

CON QUÉ CRITERIO BUSCAMOS LA RECTA??

QUEREMOS LA RECTA QUE MEJOR APROXIME A LOS PUNTOS.

???

x	y
x_1	y_1
x_2	y_2
x_3	y_3
$\vdots$	
x_m	y_m



hipótesis $\underline{h}(x) = \underline{a}\,x + \underline{b}$

$$\underline{d_1 + d_2 + d_3 + \cdots + d_m}$$

$\longrightarrow$ QUEREMOS QUE ESTA SUMA DE LAS "DISTANCIAS" SEA LO <u>MENOR</u> POSIBLE

$\downarrow$

$$\overbrace{(y_1 - h_1)}^{+} + \overbrace{(y_2 - h_2)}^{-} + \cdots + \overbrace{(y_m - h_m)}$$

$\longrightarrow$ Ojo con los signos!

$\longrightarrow |y_1 - h_1| + |y_2 - h_2| + \cdots |y_m - h_m|$

$\left.\begin{array}{l} \to |5| = 5 \\ \to |-5| = 5 \end{array}\right\}$ siempre es positivo

Esa forma de medir "distancia" como el valor absoluto de la diferencia se la denomina L_1 o ℓ_1.

↳ NO SE USA mucho porque el valor absoluto no es derivable

$$\boxed{(y_1-h_1)^2+(y_2-h_2)^2+\cdots+(y_m-h_m)^2}$$

si cada d se eleva al cuadrado son todos positivos

(se denomina L_2 o ℓ_2) y es DERIVABLE (se pueden usar todas las herramientas de análisis matemático)

QUEREMOS ENCONTRAR LOS VALORES DE a y b $(h=ax+b)$ QUE HACEN QUE LA SUMA

→ $(y_1-h_1)^2+(y_2-h_2)^2+\cdots+(y_m-h_m)^2$ SEA LO MÁS PEQUEÑA POSIBLE

→ $$\boxed{(y_1-(ax_1+b))^2+(y_2-(ax_2+b))^2+\cdots+(y_m-(ax_m+b))^2}$$

ESTE PROBLEMA DE MINIMIZACIÓN SE PUEDE RESOLVER DE MUCHAS FORMAS

→ "MÉTODO DE LA DERIVADA"

→ "MÉTODO ALGEBRAICO"

PRECISOS

→ MÉTODOS APROXIMADOS – MÉTODOS NUMÉRICOS COMO "GRADIENT DESCENT" (GD)

↳ MUCHAS FEATURES

↳ MUCHAS OBSERVACIONES

↳ POCO TIEMPO

↳ RED NEURONAL.

SÍMBOLO $\sum$ "SUMATORIA"

mayor valor

cómo es cada sumando

$$\sum_{i=1}^{4} 3^i = \underbrace{3^1 + 3^2 + 3^3 + 3^4}$$

menor valor de i

$$\sum_{i=1}^{i=7} i^i = \underbrace{1^1 + 2^2 + 3^3 + 4^4 + 5^5 + 6^6 + 7^7}$$

$$\sum_{i=1}^{i=4} \frac{1}{i} = \frac{1}{1} + \frac{1}{2} + \frac{1}{3} + \frac{1}{4}$$

En nuestro queremos minimizar:

$$(y_1 - (ax_1 + b))^2 + (y_2 - (ax_2 + b))^2 + \dots + (y_m - (ax_m + b))^2$$

con el símbolo de Σ

$$\boxed{\underset{a,b}{\text{minimizar}} \sum_{i=1}^{i=m} (y_i - (ax_i + b))^2} = \sum_{i=1}^{i=k} (y_i - h_i)^2$$

→ buscamos a y b !!

→ OLS (ORDINARY LEAST SQUARES)

"MÍNIMOS CUADRADOS ORDINARIOS"