→ Batch gradient descent

mini batch 2^n

Descenso de gradiente:

learning rate

$\theta := \theta - \alpha \nabla_\theta J(\theta)$

paso de actualización todos

$\theta_j := \theta_j - \alpha \boxed{\dfrac{\partial J(\theta)}{\partial \theta_j}}$?

función de error

$h - y = \text{error}$

$$J(\theta) = \frac{1}{2} \sum_{i=1}^{m} \left(h_\theta(x^{(i)}) - y^{(i)}\right)^2$$

m es muy grande

m son millones

miles de millones

- dificultades
- alternativas

Para un ejemplo:

$$J(\theta) = \frac{1}{2} \left(h_\theta(x) - y\right)^2$$

$$\frac{\partial J(\theta)}{\partial \theta_j} = \frac{\partial}{\partial \theta_j}\left(\frac{1}{2}(h_\theta(x)-y)^2\right) = \frac{1}{\cancel{2}} \cancel{2} \cdot (h_\theta(x)-y) \cdot \frac{\partial (h_\theta(x)-\cancel{y})}{\partial \theta_j}$$

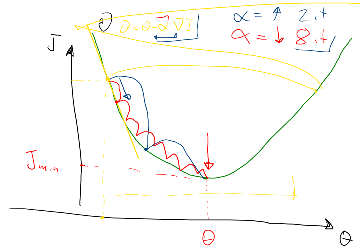
$$h_\theta(x) = \sum_{j=0}^{n-1} \theta_j x_j = \theta_0 x_0 + \theta_1 x_1 \cdots \theta_j x_j \cdots \theta_{n-1} x_{n-1}$$

$$\frac{d\,ax}{dx} = a$$

$$\frac{\partial J(\theta)}{\partial \theta_j} = (h_\theta(x) - y) \cdot x_j$$

$$\frac{d\, f(x) + g(x)}{dx} = \frac{df}{dx} + \frac{dg}{dx}$$

$$\frac{\partial J(\theta)}{\partial \theta_j} = \sum_{i=1}^{m} \left(h_\theta(x^{(i)}) - y^{(i)}\right) \cdot x_j^{(i)}$$



α define el tamaño del paso en el D.G.

α pequeño ⇒ entrenamiento + lento

α muy grande ⇒ riesgo de divergencia

Regresión lineal

Entrada: x_1 x_2 x_3 x_4 → Salida: y

5
3
6
1

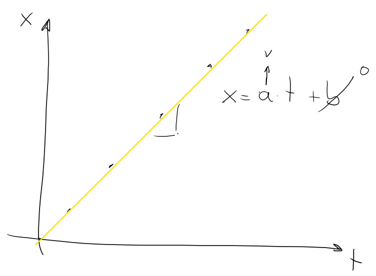
$\theta : h_\theta(x) \cong y$

$$y \cong h = \theta^T X = \theta_0 \cdot x_0 + \theta_1 x_1 + \cdots + \theta_n \cdot x_n$$

x_1	x_2	x_3	x_4	x_5	x_6	
#baños	sup. terreno	sup. const.	esquina	dormitorios	garage	precio

Importante

t [s]	x [m]
0	0
1.5	0.5
2	0.7
3	1.3
4	1.9
5	2.6
6	?



$$V = \frac{x}{t}$$

Regresión logística → Clasificación

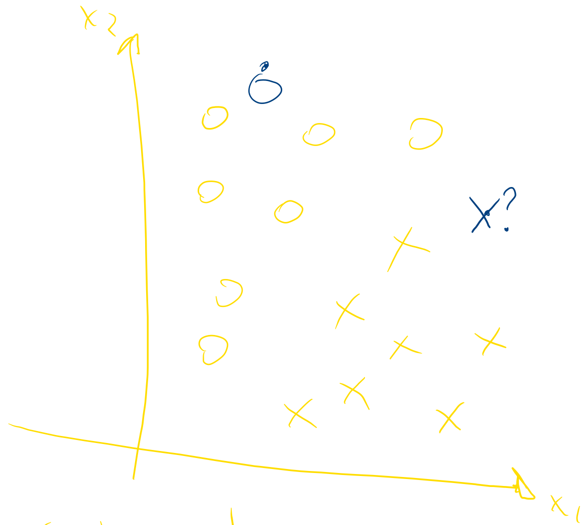
$$h_\theta(x) = g(\theta^T x)$$

$$g(z) = \frac{1}{1+e^{-z}} = \sigma(z)$$

función Sigmoide

o

x

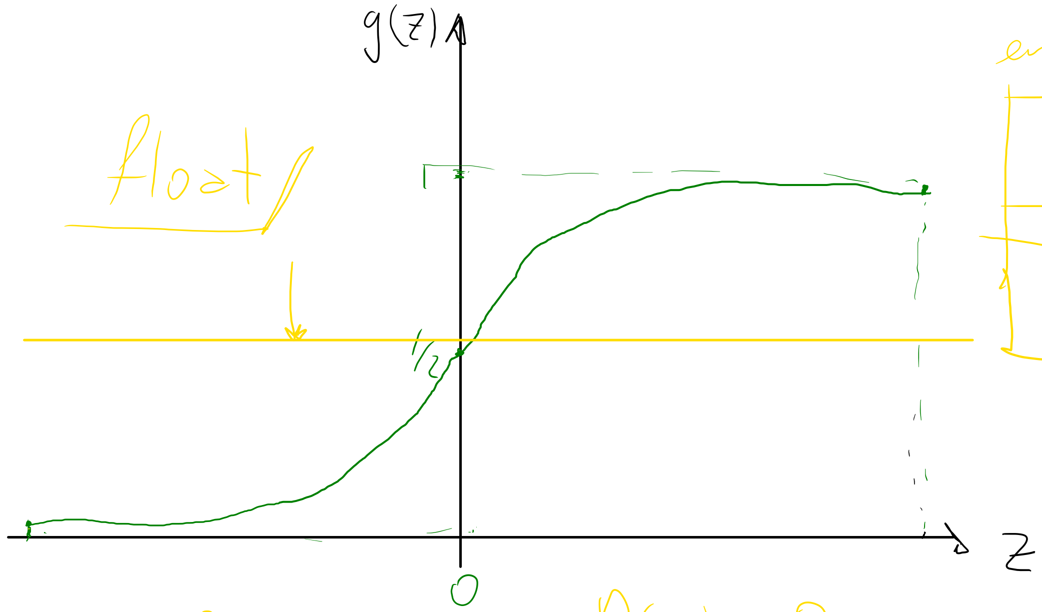


$$h_\theta(x) = \frac{1}{1+e^{-\theta^T x}}$$

$$g(z) = \frac{1}{1+e^{-z}}$$

32
64

float



entrada salida

2 clases

0

1

$0 < P() < 1$

$P(x) = 0$ →

$P(x) = 1$ → siempre

$$\text{clase}(x) = \begin{cases} 0 & g(x) \le 0.5 \\ 1 & g(x) > 0.5 \end{cases}$$

$x = \begin{bmatrix} x_1 \\ x_1 \\ x_3 \\ x_4 \end{bmatrix}$ $g \to h(x) = 0 \ldots 1$