$$a = \begin{bmatrix} 0 \\ 2 \\ 1 \\ 1 \\ 2 \end{bmatrix} \quad (5 \times 1)$$

$$b = \begin{bmatrix} 0 & 1 & 2 \end{bmatrix} \quad (1 \times 3)$$

$$a == b$$

$$\begin{bmatrix} T & F & F \\ F & F & T \\ F & T & F \\ F & T & F \\ F & F & T \end{bmatrix} \quad (5 \times 3)$$

REGRESIÓN LINEAL:

$$\{(\vec{x}_1, t_1), (\vec{x}_2, t_2), \ldots, (\vec{x}_N, t_N)\}$$

$$\vec{x}_i \in \mathbb{R}^d \qquad \vec{x}_i = (x_{i1}, x_{i2}, \ldots, x_{id})^T$$

$$t_i \in \mathbb{R}$$

$$y_i = f(\vec{x}_i) = w_1 x_{i1} + w_2 \cdot x_{i2} + \cdots + w_d \cdot x_{id} + b = \sum_{j=1}^{d} w_j x_{ij} + b = \underline{\vec{w}}^T \cdot \vec{x}_i + \underline{b}$$

$$\vec{w} = (w_1, w_2, \ldots, w_d)^T$$

PARÁMETROS

COSTE - ERROR CUADRÁTICO:

$$L = \frac{1}{2} \sum_{i=1}^{N} (y_i - t_i)^2$$

ENCONTRAR $\vec{w}, b$ QUE MINIMIZAN L.

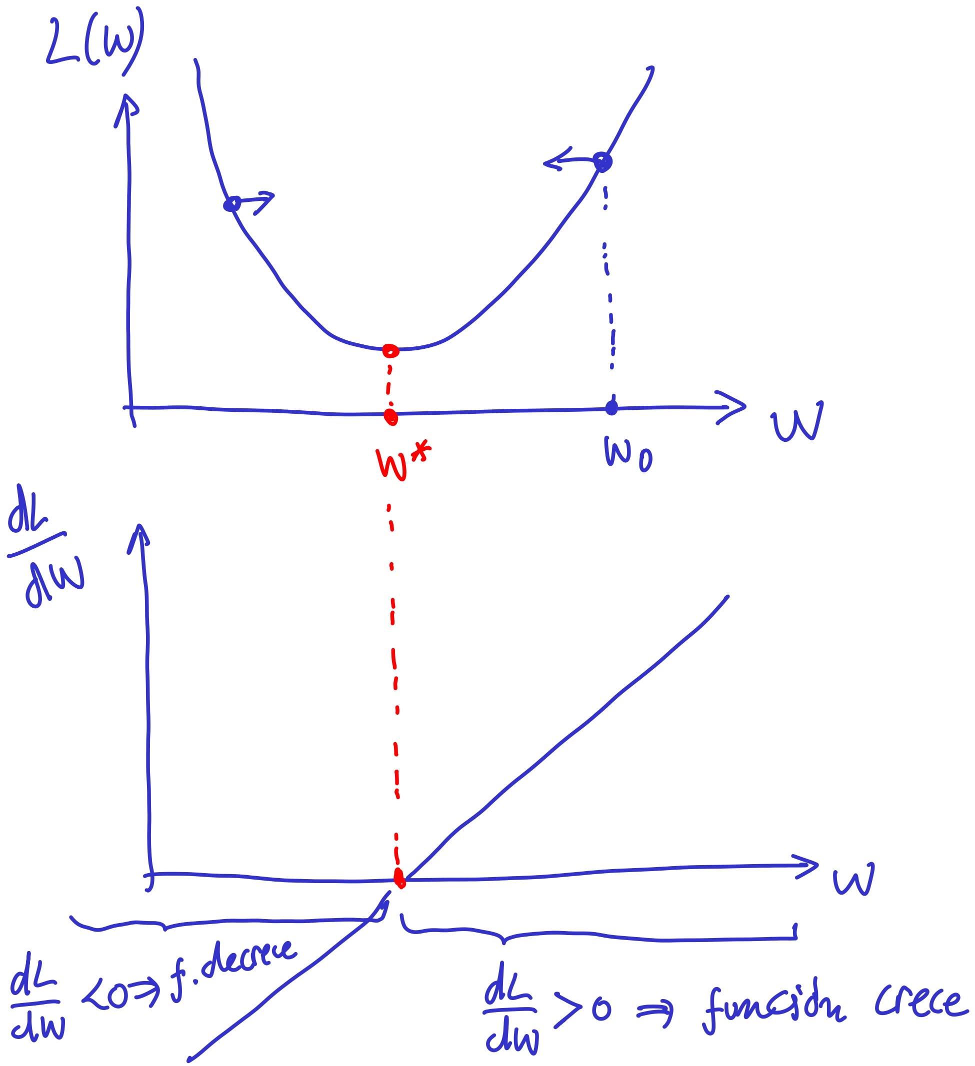
EN 1D:

$$x_i \in \mathbb{R} \qquad f(x_i) = w x_i + b = y_i$$

$$w \in \mathbb{R} \qquad L = \frac{1}{2} \sum_{i=1}^{N} (y_i - t_i)^2$$

$$\frac{\partial L}{\partial w} \qquad \frac{\partial L}{\partial b} \qquad *$$

DESCENSO POR GRADIENTE:



Empezamos en w_0 aleatorio

- if $\left.\frac{dL}{dw}\right|_{w_0} > 0 \Rightarrow$ reducir w_0

  $w_0 \leftarrow w_0 - k$

- if $\left.\frac{dL}{dw}\right|_{w_0} < 0 \Rightarrow$ aumentar w_0

  $w_0 \leftarrow w_0 + k$

- if $\left.\frac{dL}{dw}\right|_{w_0} = 0 \Rightarrow$ mínimo!!!

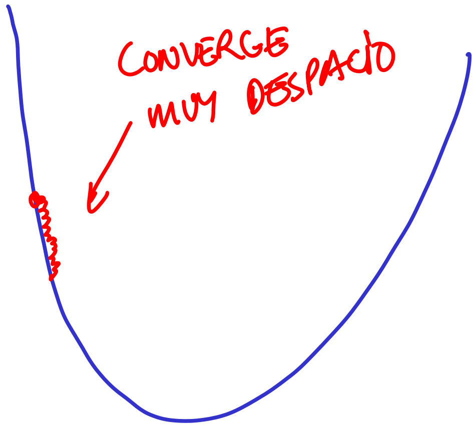
$k > 0$

$$\boxed{w_1 = w_0 - \boxed{k} \cdot \left.\frac{dL}{dw}\right|_{w_0}}$$

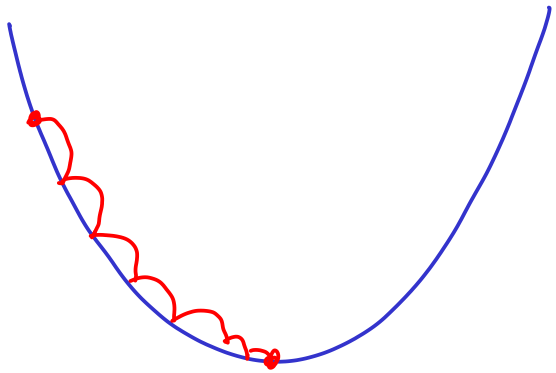
LEARNING RATE

$$\boxed{w_t = w_{t-1} - \eta \cdot \frac{dL}{dw}\Big|_{w_{t-1}}}$$

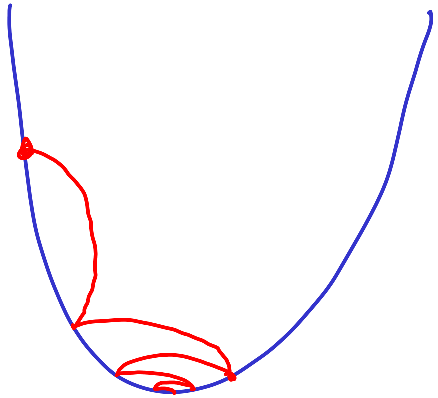
$\eta = k =$ LEARNING RATE



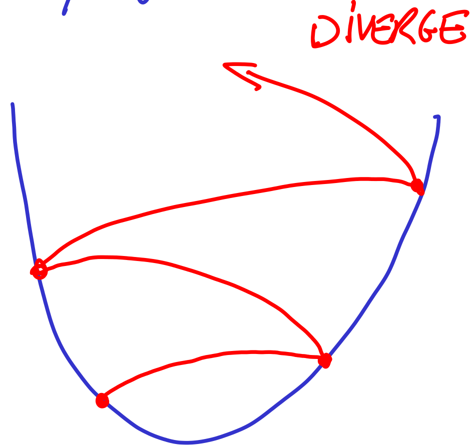
η pequeño



η adecuado



η grande



EN 1D:

$x_i \in \mathbb{R}$

$w \in \mathbb{R}$

$$\left.\begin{aligned} f(x_i) &= w x_i + b = y_i \\ L &= \frac{1}{2}\sum_{i=1}^{N}(y_i - t_i)^2 \end{aligned}\right\}$$

$$\frac{\partial L}{\partial w} \qquad \frac{\partial L}{\partial b}$$

$$\frac{\partial L}{\partial b} = \sum_{i=1}^{N}(y_i - t_i)\cdot\frac{\partial y_i}{\partial b} = \sum_{i=1}^{N}(y_i - t_i)$$

$$\frac{\partial L}{\partial w} = \sum_{i=1}^{W}(y_i - t_i)\cdot\frac{\partial y_i}{\partial w} = \sum_{i=1}^{N}(y_i - t_i)\cdot x_i$$

$$\boxed{\begin{aligned} b_t &= b_{t-1} - \eta\cdot\sum_{i=1}^{N}(y_i - t_i) \\ w_t &= w_{t-1} - \eta\cdot\sum_{i=1}^{N}(y_i - t_i)x_i \end{aligned}}$$

```
num_iters = 8 # number of iterations
eta = 0.0001  # learning rate
for i in range(num_iters):
    y = w*x + b
    y_minus_t = y - t
    dw = np.sum(y_minus_t*x)
    db = np.sum(y_minus_t)
    w -= eta*dw
    b -= eta*db
```

w = w - eta * dw

+= *=

EN d dimensiones:

$$(w_1, w_2, \dots, w_d)$$

$$\vec{\nabla}_{\vec{w}} L = \left(\frac{\partial L}{\partial w_1}, \frac{\partial L}{\partial w_2}, \dots, \frac{\partial L}{\partial w_d}\right)$$

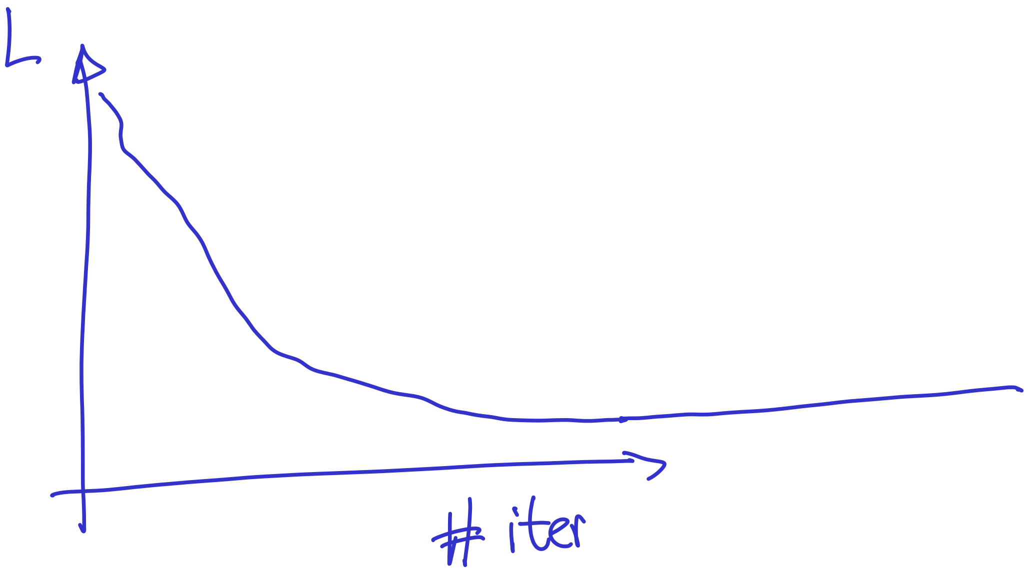
$$\vec{\nabla}_{\vec{w}} L = \sum_{i=1}^{N} (y_i - t_i) \cdot \vec{x}_i$$

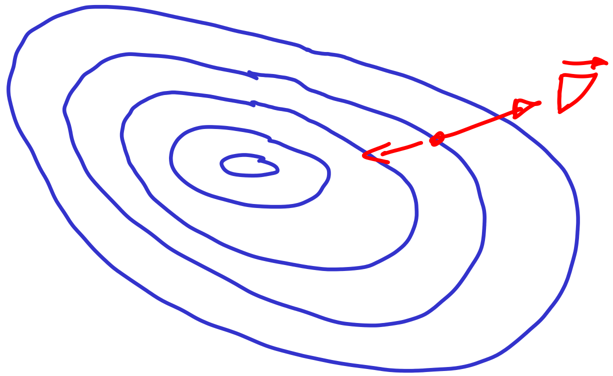
$$\vec{x}_i = (x_{i1}, x_{i2}, \dots, x_{id})$$

$$\boxed{\vec{w}_t = \vec{w}_{t-1} - \eta \cdot \sum_{i=1}^{N} (y_i - t_i)\vec{x}_i}$$

$$\mathbf{w}_{t+1} = \mathbf{w}_t - \eta \nabla_{\mathbf{w}} E, \qquad \nabla_{\mathbf{w}} E = \sum_i (y_i - t_i)\mathbf{x}_i,$$

$$b_{t+1} = b_t - \eta \frac{\partial E}{\partial b}, \qquad \frac{\partial E}{\partial b} = \sum_i (y_i - t_i).$$





REGRESIÓN LOGÍSTICA:

$$\{(\vec{x}_1, t_1), (\vec{x}_2, t_2), \dots, (\vec{x}_N, t_N)\} \qquad \vec{w} = (w_1, w_2, \dots, w_d)^T$$

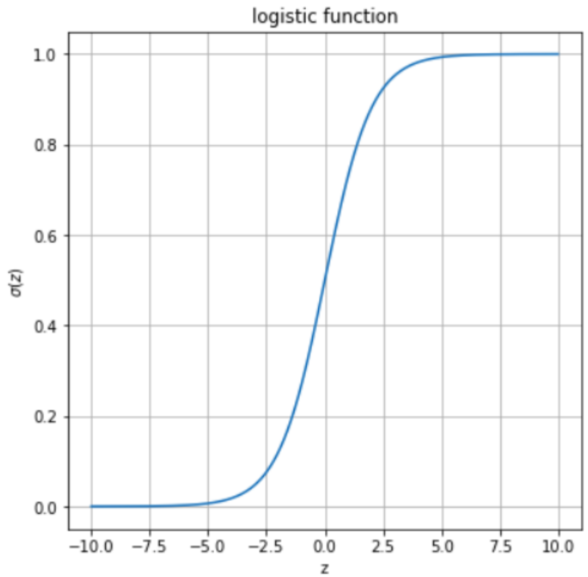
$$\vec{x}_i = (x_{i1}, x_{i2}, \dots, x_{id})^T \in \mathbb{R}^d$$

$$t_i \in \{0, 1\}$$

FUNCIÓN SIGMOIDE / LOGÍSTICA

Modelo: $y_i = f(\vec{x}_i) = \sigma(\underline{\vec{w}}^T \cdot \vec{x}_i + \underline{b})$

PARÁMETROS

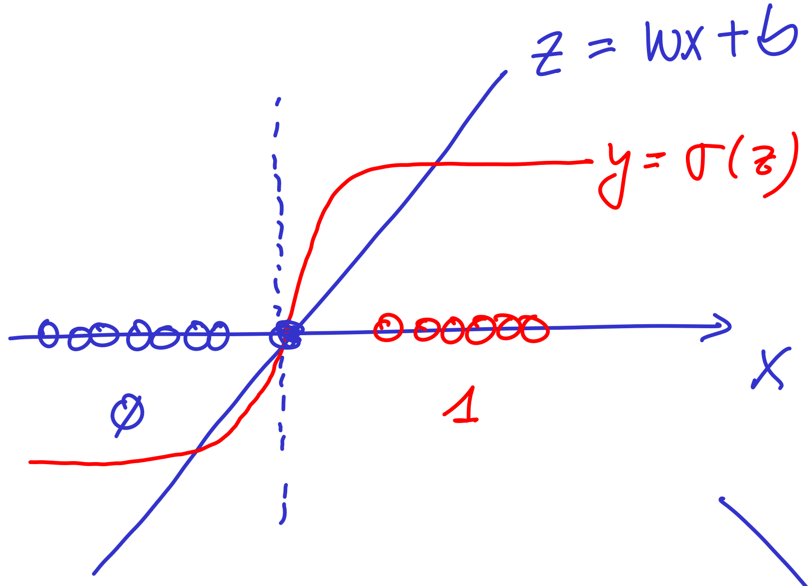


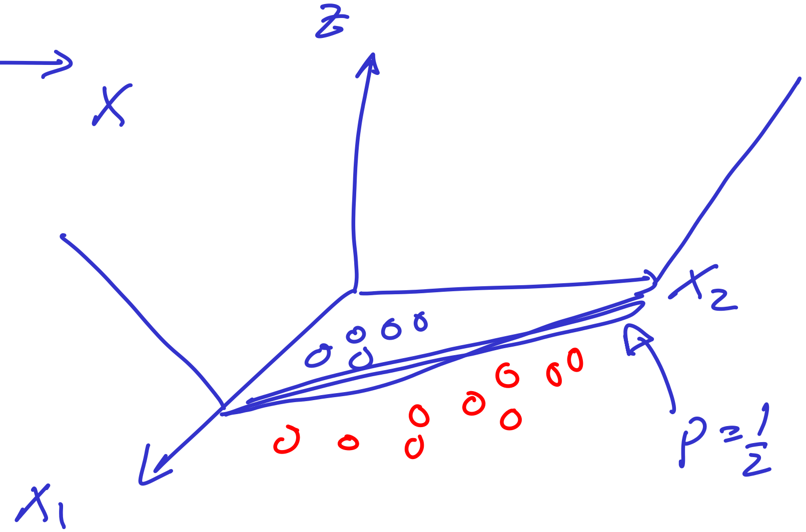
$y_i \equiv$ PROB. QUE ASIGNA EL MODELO A LA CLASE 1

$$y_i = P(t_i = 1 \mid \vec{x}_i) \qquad 1 - y_i = P(t_i = 0 \mid \vec{x}_i)$$

MODELO LINEAL:

$$y = \sigma(z) = \sigma(wx+b)$$





FUNCIÓN DE COSTE:

MAXIMIZAR LA PROB. QUE EL MODELO ASIGNA A LAS OBSERVACIONES

VEROSIMILITUD: $$l = \prod_{i=1}^{N} P(t = t_i \mid \vec{x}_i) = \prod_{i=1}^{N} \begin{bmatrix} y_i & \text{si } t_i = 1 \\ 1 - y_i & \text{si } t_i = \emptyset \end{bmatrix}$$

LOG-VEROSIMILITUD: $$L = \log l = \sum_{i=1}^{N} \log P(t = t_i \mid \vec{x}_i) = \sum_{i=1}^{N} \begin{bmatrix} \log y_i & \text{si } t_i = 1 \\ \log(1 - y_i) & \text{si } t_i = \emptyset \end{bmatrix}$$

$$L = \sum_{i=1}^{N} \left[t_i \cdot \log y_i + (1-t_i) \cdot \log(1-y_i) \right]$$

$$\boxed{\text{CROSS-ENTROPY} = -L = -\sum_{i=1}^{N} \Big[\underbrace{t_i \log y_i + (1-t_i) \log(1-y_i)}_{XE_i} \Big] = XE}$$

$$y_i = \sigma(z_i) \quad , \quad z_i = \vec{w}^T \cdot \vec{x}_i + b \qquad\qquad XE = -\sum_{i=1}^{N} XE_i$$

$$\frac{\partial XE}{\partial b} = -\sum_{i=1}^{N} \boxed{\frac{\partial XE_i}{\partial y_i}} \cdot \boxed{\frac{\partial y_i}{\partial z_i}} \cdot \overset{1}{\boxed{\frac{\partial z_i}{\partial b}}}$$

$$\frac{\partial XE}{\partial w_j} = -\sum_{i=1}^{N} \boxed{\frac{\partial XE_i}{\partial y_i}} \cdot \boxed{\frac{\partial y_i}{\partial z_i}} \cdot \underset{x_{ij}}{\boxed{\frac{\partial z_i}{\partial w_j}}}$$

$$\vec{\nabla}_{\vec{w}} XE = \left(\frac{\partial XE}{\partial w_1}, \frac{\partial XE}{\partial w_2}, \dots, \frac{\partial XE}{\partial w_d} \right)$$

$$XE_i = t_i \cdot \log y_i + (1-t_i) \log(1-y_i)$$

$$\frac{\partial XE_i}{\partial y_i} = \frac{t_i}{y_i} - \frac{1-t_i}{1-y_i} = \frac{t_i - \cancel{t_i y_i} - y_i + \cancel{t_i y_i}}{y_i(1-y_i)} = \boxed{\frac{t_i - y_i}{y_i(1-y_i)}}$$

$$y_i = \sigma(z_i) \qquad \sigma(z) = \frac{e^z}{e^z+1} = \frac{1}{1+e^{-z}}.$$

$$\frac{\partial y_i}{\partial z_i} = \frac{+1}{(1+e^{-z_i})^2} \cdot e^{-z_i} = \frac{1 + e^{-z_i} - 1}{(1+e^{-z_i})(1+e^{-z_i})} = \underbrace{\frac{1}{1+e^{-z_i}}}_{\sigma(z_i)} - \underbrace{\frac{1}{(1+e^{-z_i})^2}}_{\sigma(z_i)^2} =$$

$$= \sigma(z_i) - \sigma(z_i)^2 = \sigma(z_i)(1-\sigma(z_i)) = \boxed{y_i(1-y_i)}$$

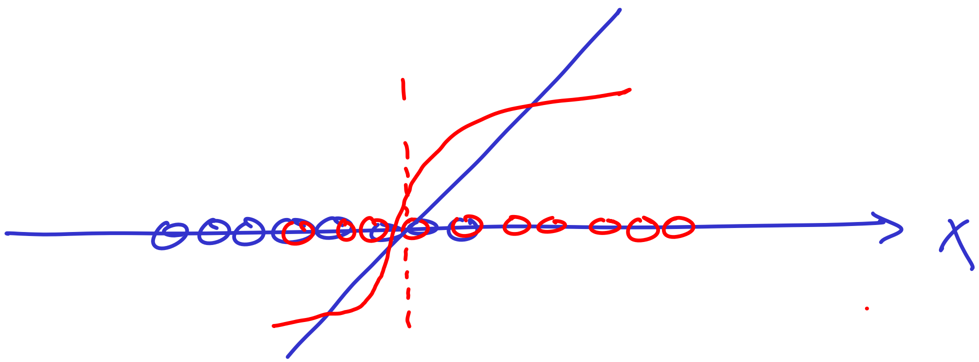
$$\frac{\partial XE}{\partial b} = -\sum_{i=1}^{N} \frac{t_i - y_i}{\cancel{y_i(1-y_i)}} \cdot \cancel{y_i(1-y_i)} \cdot 1 = \underline{\sum_{i=1}^{N} (y_i - t_i)}$$

$$\frac{\partial XE}{\partial w_j} = \underline{\sum_{i=1}^{N} (y_i - t_i) \cdot x_{ij}}$$

$$\vec{\nabla}_{\vec{w}} XE = \left(\frac{\partial XE}{\partial w_1}, \frac{\partial XE}{\partial w_2}, \dots, \frac{\partial XE}{\partial w_d}\right) = \sum_{i=1}^{N}(y_i - t_i)\cdot \underbrace{(x_{i1}, x_{i2}, \dots, x_{id})}_{\vec{x}_i} =$$

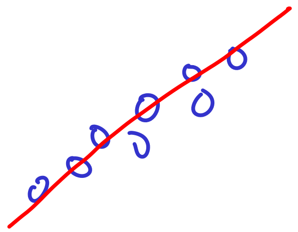
$$= \sum_{i=1}^{N}(y_i - t_i)\cdot \vec{x}_i$$

$$\boxed{\begin{aligned} b_t &= b_{t-1} - \eta \cdot \frac{\partial XE}{\partial b} = b_{t-1} - \eta \cdot \sum_{i=1}^{N}(y_i - t_i) \\ \vec{w}_t &= \vec{w}_{t-1} - \eta \cdot \vec{\nabla}_{\vec{w}} XE = \vec{w}_{t-1} - \eta \cdot \sum_{i=1}^{N}(y_i - t_i)\cdot \vec{x}_i \end{aligned}}$$



REGR. LINEAL

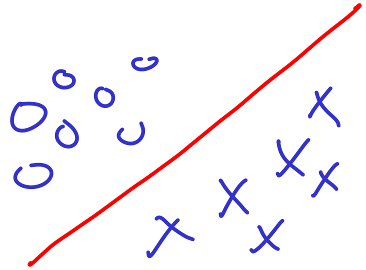
```
num_iters = 8 # number of iterations
eta = 0.0001  # learning rate
for i in range(num_iters):
    y = w*x + b
    y_minus_t = y - t
    dw = np.sum(y_minus_t*x)
    db = np.sum(y_minus_t)
    w -= eta*dw
    b -= eta*db
```

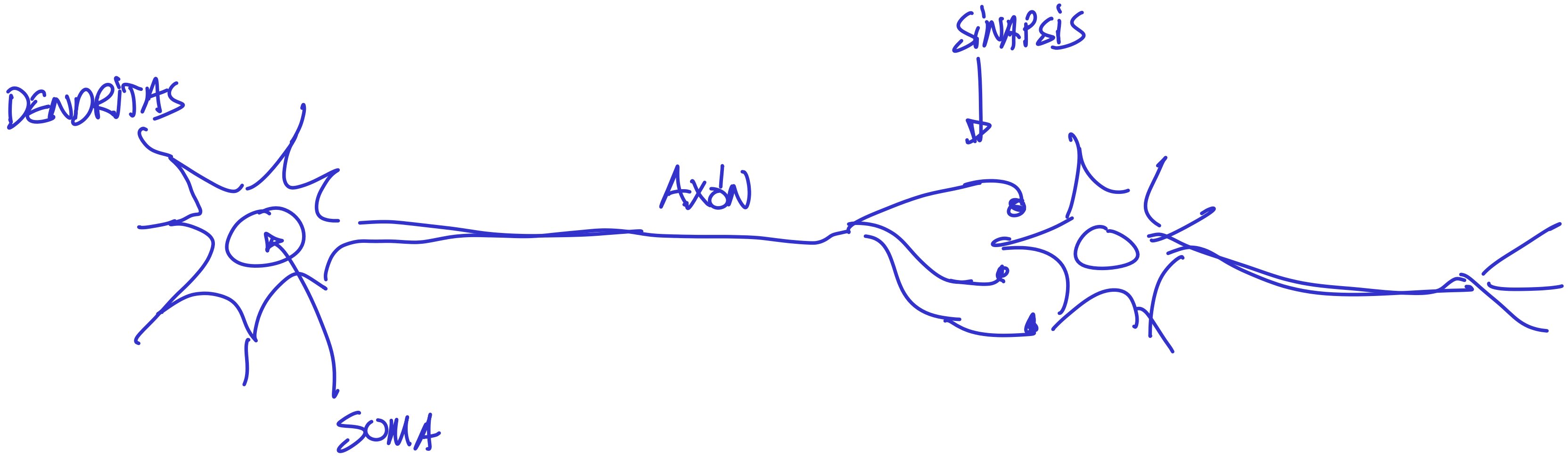


REGR. LOGISTICA

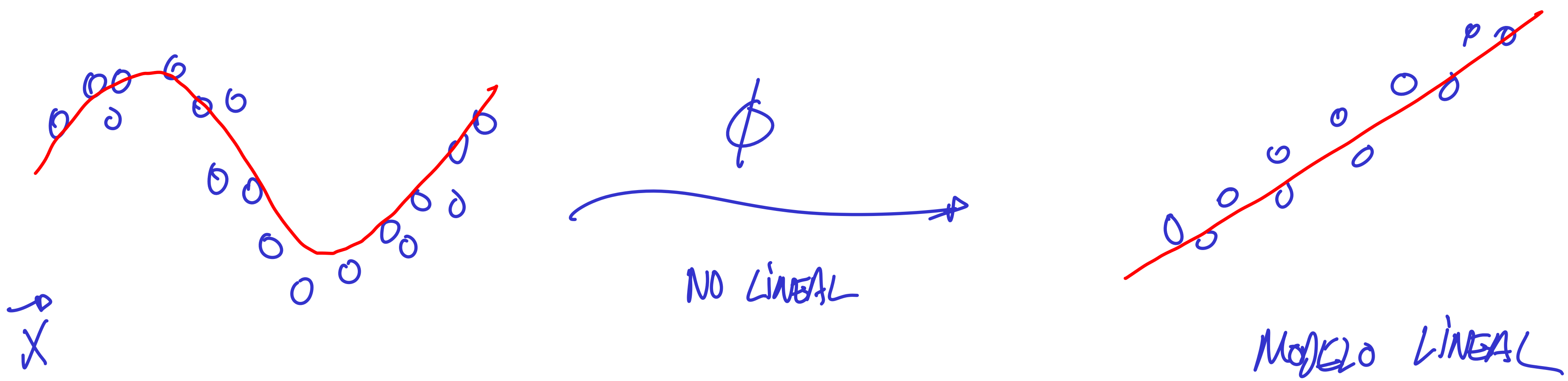
```
num_iters = 1000 # number of iterations
eta = 0.0002     # learning rate
for i in range(num_iters):
    z = w*x + b
    y = 1.0/(1.0 + np.exp(-z))
    y_minus_t = y - t
    dw = np.sum(y_minus_t*x)
    db = np.sum(y_minus_t)
    w -= eta*dw
    b -= eta*db
```

(LINEALES)



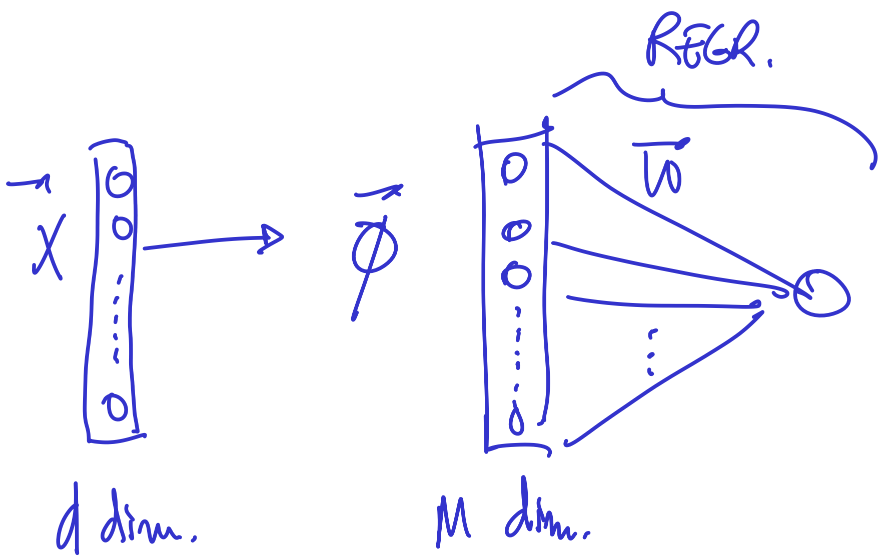


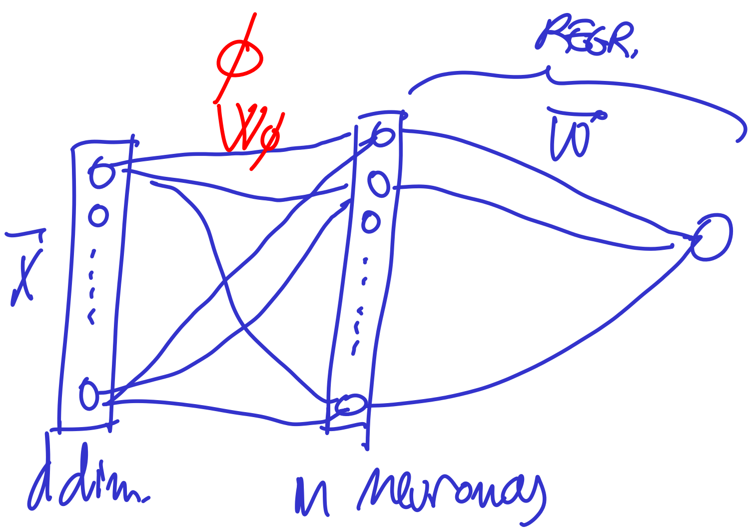
PROBLEMA NO LINEAL:

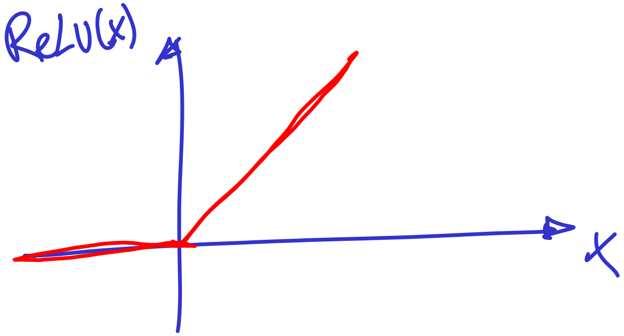


MODELO LINEAL EN ϕ

$$y = \vec{w}^t \cdot \vec{\phi} + b$$

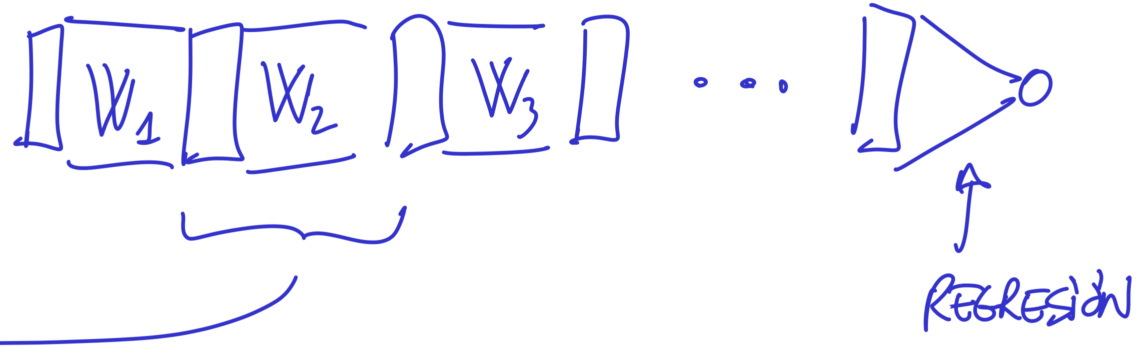


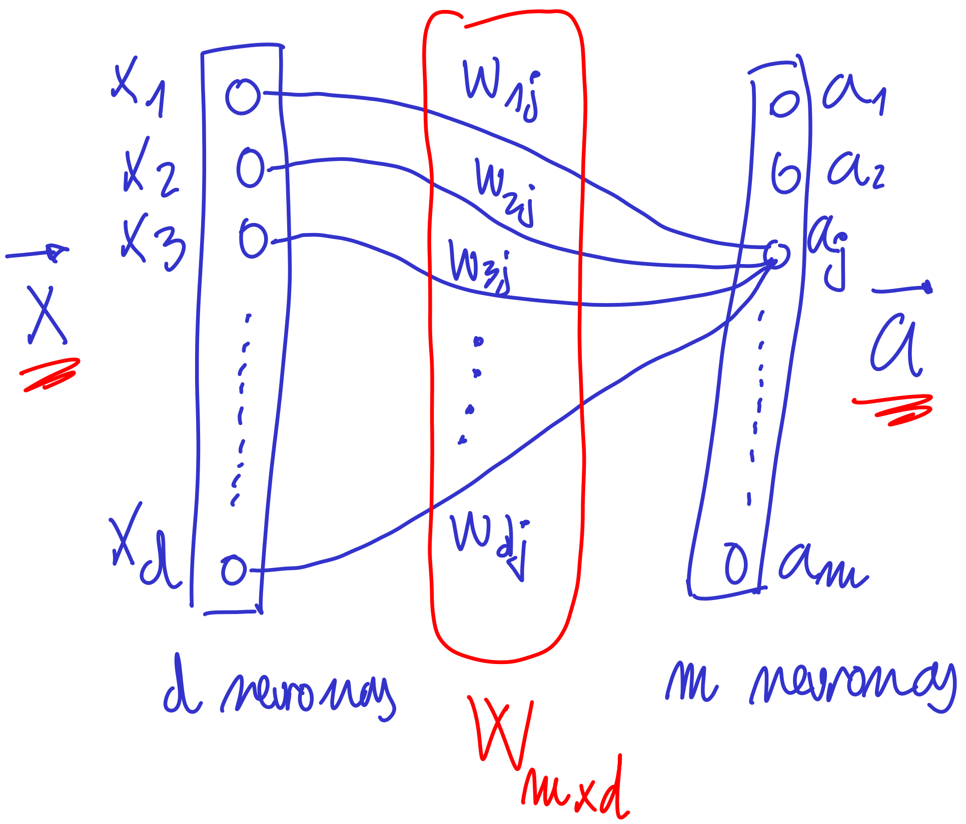




	REGRESIÓN LINEAL	REGR. LOGÍSTICA (CLASIFICACIÓN)
MODELO	$y = \vec{w}^t \cdot \vec{x} + b$	$z = \vec{w}^t \vec{x} + b$ $y = \sigma(z)$
COSTE	$MSE = \frac{1}{2} \sum_{i=1}^{N} (y_i - t_i)^2$	$XE = -\sum_{i=1}^{N} \left[t_i \log(y_i) + (1 - t_i) \log(1 - y_i) \right]$

RED NEURONAL:





$$\left.\begin{aligned} a_j &= f(z_j) \\ z_j &= \sum_{i=1}^{d} w_{ij}\cdot x_i + b_j \end{aligned}\right\}$$

$$\boxed{\begin{aligned} \underset{m\times 1}{\vec{z}} &= \underset{m\times d}{W}\cdot \underset{d\times 1}{\vec{x}} + \underset{m\times 1}{\vec{b}} \\ \vec{a} &= f(\vec{z}) \end{aligned}}$$