Regresión Lineal

$$x \in \mathbb{R}^n \qquad x = (x^{(1)}, x^{(2)}, \dots x^{(n)})$$

$$\begin{bmatrix} x^{(1)} \\ x^{(2)} \\ \vdots \\ x^{(m)} \end{bmatrix} \underbrace{\begin{bmatrix} x_1^{(1)}, & x_2^{(1)} & \dots & x_n^{(1)} \\ x_1^{(2)}, & x_2^{(2)}, & \dots & x_n^{(2)} \\ x_1^{(m)}, & x_2^{(m)}, & \dots & x_n^{(m)} \end{bmatrix}}_{X} \begin{bmatrix} y^{(1)} \\ y^{(2)} \\ \vdots \\ y^{(m)} \end{bmatrix}$$

Notación

$w = [w_1, w_2]$ feature extractor $\phi(x) = [1, x]$

feature vector

$f(w) = w \cdot \phi(x)$ score

$$\phi(x) = (\phi(x)_1, \phi(x)_2 \dots \phi(x)_m)$$

Función

$$\phi : \mathbb{R}^n \to \mathbb{R}^m$$

características del vector x

Nuevas características del vector

$$\phi(x) = (1, x_1, x_2 \dots, x_m)$$

Asumimos que existe

$$h_w(x) = w^T \phi(x) \quad w \in \mathbb{R}^{n+1}$$

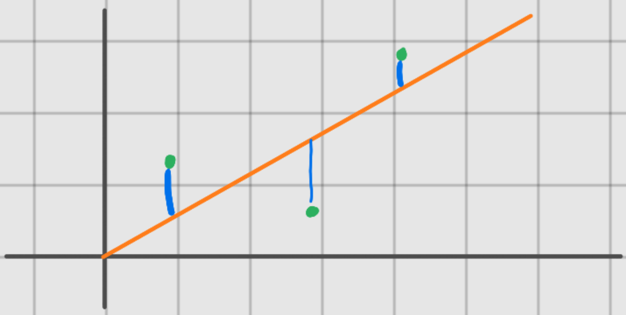
$$x = x_1 \qquad \phi(x) = [1, x_1]$$

$$w = [w_0 \ \ w_1] \qquad h_w(x) = [w_0 \ \ w_1] \begin{bmatrix} 1 \\ x_1 \end{bmatrix} = w_0 + w_1 x_1$$

$$h_w(x) = \sum_{i=1}^{n} w_i x_i + w_0 = w^T x + b$$

$$\mathcal{F} = \{ f_w : w \in \mathbb{R}^2 \}$$

Función de Pérdida



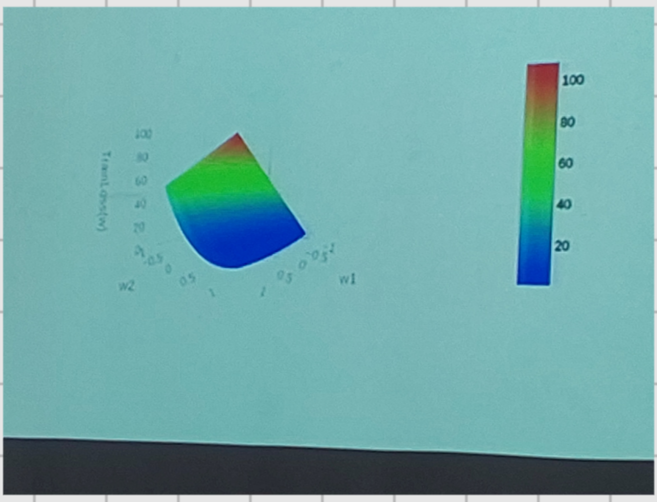
$$\text{loss}(y, \hat{y}) = \frac{1}{2}(y - \hat{y})^2$$

$$E_{in}(h_\theta) = \frac{1}{n}\sum_{i=1}^{n} \text{loss}(y_i, h_\theta(x^{(i)}))$$

$$E_{in}(h_\theta) = \frac{1}{m}\sum_{i=1}^{1} \frac{1}{2}\left(y^{(i)} - w^T \phi(x^{(i)})\right)^2$$

Min Square Error

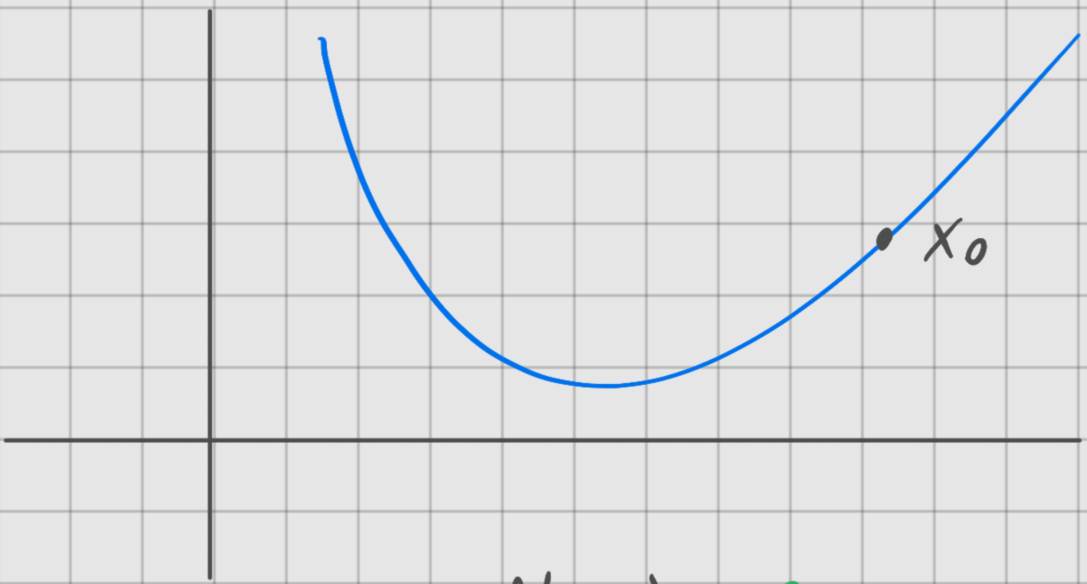
$$W^* = \arg\min_{w \in \mathbb{R}^{n+1}} \frac{1}{2m} \sum_{i=1}^{m} \left(y^{(i)} - \sum_{i=1}^{n} W_i x_i - W_2\right)^2$$



Dado que es cuadrático, es una función convexa.

Tiene un solo mínimo global

Gradiente: Es la dirección que <u>más</u> incrementa la pérdida en el entrenamiento



$f'(x_0) > 0$

$x_0 \quad f(x_0) \quad f'(x_0)$ ↙ Paso / Tasa de aprendizaje

$x_1 \rightarrow x_0 - \alpha f'(x_0)$

$x_2 \rightarrow x_1 - \alpha f'(x_1)$

$\vdots$

$x_n \rightarrow x_{n-1} - \alpha f'(x_{n-1})$

$$\nabla f(X_1, X_2 \ldots X_m) = \begin{bmatrix} \frac{\partial f}{X_1} \\ \vdots \\ \frac{\partial f}{X_m} \end{bmatrix}$$

$$\frac{\partial f(w)}{\partial w_k} = -\frac{1}{m} \sum_{i=1}^{m} (y^{(i)} - \hat{y}^{(i)}) x_k$$

$$\nabla E_{out}(w) = -\frac{1}{m} \begin{bmatrix} \sum_{i=1}^{m} (y^{(i)} - \hat{y}^{(i)}) \cdot 1 \\ \sum_{i=1}^{m} (y^{(i)} - \hat{y}^{(i)}) \cdot x_1^{(i)} \\ \vdots \\ \sum_{i=1}^{m} (y^{(i)} - \hat{y}^{(i)}) X_n^{(i)} \end{bmatrix}$$

$$\nabla E_{out}(w) = -\frac{1}{m} (y - \hat{y})$$

$w \leftarrow w\_ins$

Para epoch de 1 a max_epoch

$\hat{y}^{(i)} \leftarrow w^T \phi(x^{(i)}) \ \forall i \ 1 \ldots n$

$e^{(i)} \leftarrow y^{(i)} - \hat{y}^{(i)} \quad \forall i \ 1 \ldots n$

$E_{in} \leftarrow 1/m \sum_{i=1}^{m} e^{(i)^2}$ MSE

$\nabla_w E_{in} \leftarrow$

Si $\|\nabla_w E_{in}\| < tol$
return w

$w \leftarrow w - \eta \nabla_w E_{in}$

fin para
return w ← Después de todas las iteraciones

$$\phi(X) = \begin{bmatrix} 1, & X_1^{(1)}, & \dots & X_n^{(1)} \\ 1, & X_1^{(2)}, & \dots & X_n^{(2)} \\ \vdots & & \ddots & \vdots \\ 1, & X_1^{(n)} & \dots & X_n^{(n)} \end{bmatrix} \quad y = \begin{bmatrix} y^{(1)} \\ y^{(2)} \\ \vdots \\ y^{(n)} \end{bmatrix}$$

$$\hat{y} = w^T \phi(x^{(i)}) = \phi(x^{(i)})^T w$$

$w \leftarrow w\_ins$

Para epoch de 1 a max_epoch

$\tilde{y}^{(i)} \leftarrow \phi(X) w$

$E = y - \hat{y}$

$E_{in} = \frac{1}{N} E^T E$

$\nabla_w E_{in} \leftarrow \frac{1}{N} \phi(x)^T E$

si $\|\nabla_w E_{in}\| < Tol$:

return w

$w = w - \eta \nabla_w E_{in}$

fin para

return w $\leftarrow$ Después de todas las iteraciones

$y^{(i)} \in \{-1, 1\}$

$x^{(i)} = \{x_1^{(i)}, \dots x_n^{(i)}\} \in \mathbb{R}^n$
$\phi(x^{(i)}) = \{1, x_1^{(i)}, \dots x_n^{(i)}\} \in \mathbb{R}^{n+1}$
$W = (W_0, W_1 \dots W_n) \in \mathbb{R}^{n+1}$

$$\hat{y}^{(i)} = h(x^{(i)}) = \text{sign}(\phi(x^{(i)})^T W)$$

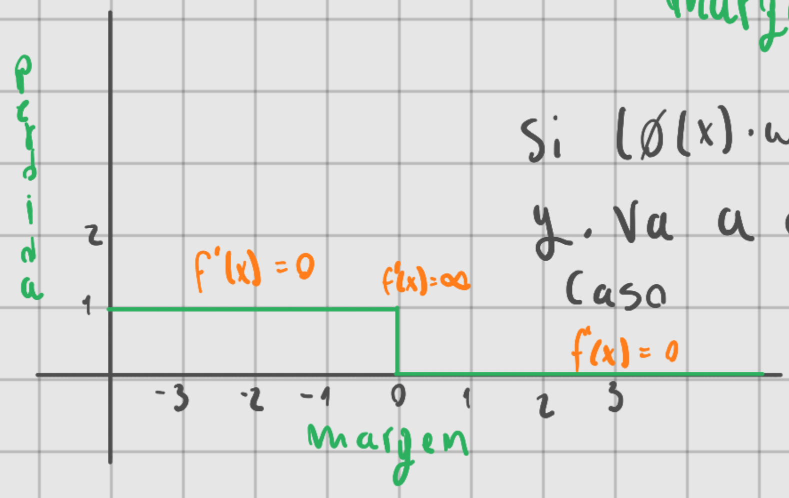
$$W_1 X + W_0 = 0$$

Cuando una f de lrning es lineal
nos referimos a la relación de los parámetros

$$\text{Loss}_{0-1}(x, y, w) = 1[H_w(x) \neq y] \quad \text{Zero - one loss}$$

$$= 1[\underbrace{(\phi(x) \cdot w) y}_{\text{Margen}} \leq 0]$$

Si $(\phi(x) \cdot w)$ difiere del signo de y, va a dar -1, 1 en otro caso



$$E_{out}(W) = \frac{1}{M} \sum_{i=1}^{m} 1[W^T \phi(x^{(i)}) y^{(i)} \leq 0]$$

$$\text{Loss}_{\text{Hinge}}(x, y, w) = \max\{1 - (W \cdot \phi(x))y, 0\}$$

Para compensar datos mal clasificados o bien clasificados por muy poco

$$\nabla \text{Loss}_{\text{Hinge}}(x, y, w) = \begin{cases} -\phi(x)y & \text{if } \{1-(w \cdot \phi(x)y)\} > 0 \\ 0 & \text{en otro caso} \end{cases}$$

$w \leftarrow w\_ins$

Para epoch de 1 a max_epoch

$z^{(i)} = \phi(x^{(i)})^T w \ \forall i = 1 \dots n$

$a^{(i)} = z^{(i)}$ if $\text{sign}(z^{(i)}) == y^{(i)} \ \forall i = 1 \dots n$ si no 0

$\nabla_w E_{in} = \frac{1}{M} \sum_{i=1}^{m} -a^{(i)} y^{(i)}$

$w = w - \eta \nabla_w E_{in}$

fin para

return w