3 feb 2026

$$\frac{\partial Ein(w,b)}{\partial w_j} = \frac{1}{M}\sum_{i=1}^{m} -(a^{(i)} - \hat{a}^{(i)})\, x_j^{(i)}$$

$$= \frac{1}{M}\sum_{i=1}^{m}\left(-\frac{a^{(i)}}{\hat{a}^{(i)}} + \frac{1-a^{(i)}}{1-\hat{a}^{(i)}}\right)(a^{(i)} - \hat{a}^{(i)})\, x_j^{(i)}$$

$$\frac{\partial Ein(w,b)}{\partial w_j} = \frac{\partial}{\partial w_j}\frac{1}{M}\sum_{i=1}^{m}\left[-a^{(i)}\ln(\hat{a}^{(i)}) - (1-a^{(i)})\ln(1-\hat{a}^{(i)})\right]$$

$$= \frac{1}{M}\sum_{i=1}^{m} -\frac{a^{(i)}}{\hat{a}^{(i)}}\frac{\partial \hat{a}^{(i)}}{\partial w_j} + \frac{1-a^{(i)}}{1-\hat{a}^{(i)}}\frac{\partial \hat{a}^{(i)}}{\partial w_j}$$

$$= \frac{1}{M}\sum_{i=1}^{m}\left[-a^{(i)}(1-\hat{a}^{(i)}) + (1-a^{(i)})\,\hat{a}^{(i)}\right] x_j^{(i)}$$

$$= \frac{1}{M}\sum_{i=1}^{m}\left[-a^{(i)} + \cancel{a^{(i)}\hat{a}^{(i)}} + \hat{a}^{(i)} - \cancel{a^{(i)}\hat{a}^{(i)}}\right] x_j^{(i)}$$

$$= -\frac{1}{M}\sum_{i=1}^{m}(a^{(i)} - \hat{a}^{(i)})\, x_j^{(i)}$$

3 feb 2026

$$W \leftarrow W - lr \; X^T (A - \hat{A})$$

$$b \leftarrow b - lr \; \frac{1}{M} \sum_{i=1}^{m} (a^{(i)} - \hat{a}^{(i)})$$

Regularización

Multiplicador de la Grange

$1/c$

> $W^*_{reg}, b^*_{reg} = \arg\min \left[E_{in}(w,b) + \frac{\lambda}{M} \sum_{j=1}^{n} W_j^2 \right]$

C tiene que ser ≥ 0,

Podemos agregarle esto y se puede regular cualquier problema.

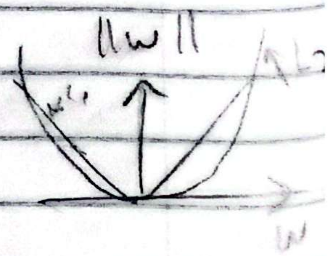
> $\lambda = 0$ no regularices

> $\lambda \Rightarrow$ factor de regularización

$$\frac{\partial}{\partial W_j} \left[E_{in}(w,b) + \frac{\lambda}{M} \sum_{j=1}^{n} \right]$$

$$= -\frac{1}{M} \sum_{j=1}^{n} (a^{(i)} - \hat{a}^{(i)}) X_j^{(i)} + \underbrace{\frac{2\lambda}{M} W_j}_{\text{regularización}}$$

$$W^*_{reg}, b^*_{reg} = \arg\min \left[E_{in}(w,b) + \frac{\lambda}{M} \, regu(w) \right]$$



$regu(w) \rightarrow \sum_{i=1}^{n} W_j^2 = ||W|| \; L_2$

$\rightarrow \sum_{i=1}^{n} |W_j| \quad L_1$

3 feb

reg. lineal + $\to$ L_2 Ridge
$\to$ L_1 Lasso
$\to$ $L_1 + L_2$ ElasticNet

L_1 funciona mejor en frente de dimensiones

4 feb 2026

$$w^*, b^* = \arg\min_{(w,b)} E_{in}(w,b)$$

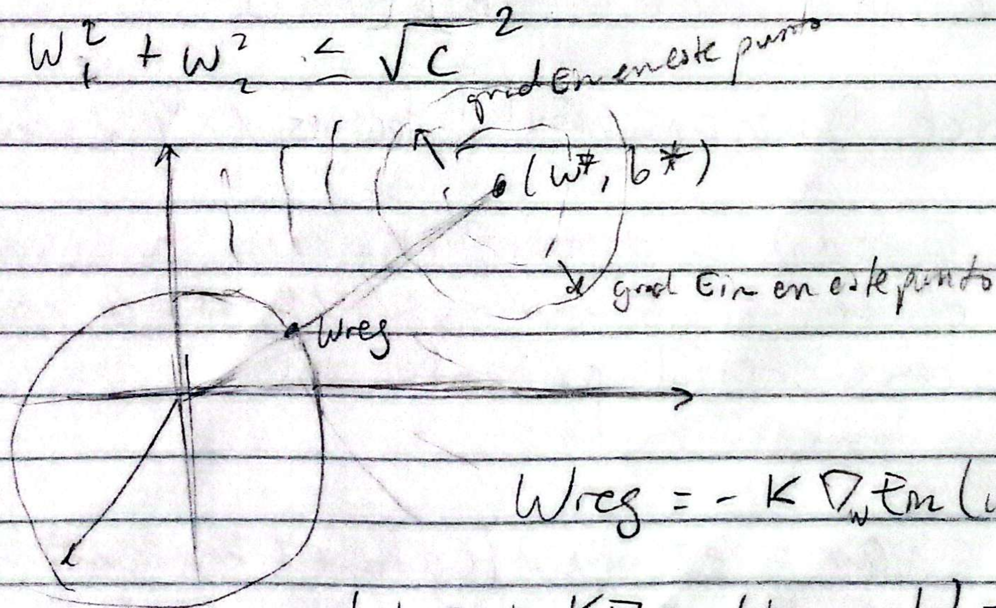
bajo $\sum_{j=1}^{n} w_j^2 \leq C$

$w^T w \leq C$

$\|w\|_2^2 \leq C$

Estos son lo mismo

$$w_1^2 + w_2^2 \leq \sqrt{C}^2$$



$$W_{reg} = -K \nabla_w E_{in}(w,b)$$

$$W_{reg} + K \nabla E_{in}(W_{reg}, b) = 0$$

lambda es positivo

$$2\lambda W_{reg} + \nabla E_{in}(W_{reg}, b) = 0$$

$\nabla_w (E_{in}(w, b) + \lambda w^T w)$

$\nabla_w E_{in}(w, b) + 2\lambda w$

Multiplicadores de Lagrange

5 Feb 2026

Arboles de decision

$f: X \to Y$

y quiero ajustar $h(\theta): X \to Y$ a partir de un conjunto de datos (X^n, Y^n)

$$\begin{array}{cc} & \begin{array}{cccc} f_1 & f_2 & \cdots & f_n \end{array} \leftarrow \text{feature} \\ \begin{array}{c} X^{(1)} \\ X^{(2)} \\ \vdots \\ X^{(n)} \end{array} & \begin{pmatrix} X^1_1 & X^1_2 & & X^1_n \\ X^2_1 & X^2_2 & & X^2_n \\ & & & \\ X^n_1 & X^n_2 & & X^n_n \end{pmatrix} \begin{pmatrix} Y^{(1)} \\ Y^{(2)} \\ \\ Y^{(n)} \end{pmatrix} \end{array}$$

def Generar_arbol (features, X, Y, nodo)

Si todos los datos pertenecen a misma clase, si no hay datos o si no hay características entonces
regresa nodo;

var = escoge_feature(features, X, Y)
# Separo los datos de las características
Quitar var de features
por valor en valores(var)
X_h, Y_h = separa_datos(X, Y, var, valor)

nn = crear_hijo(n, yh)

>Cuando no pueda tomar decisión me quedo con la clase del cuadrado.

nh = genera_arbol(features, Xn, Yn, nn)