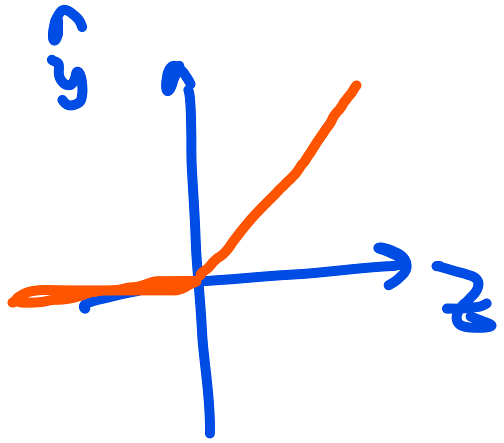
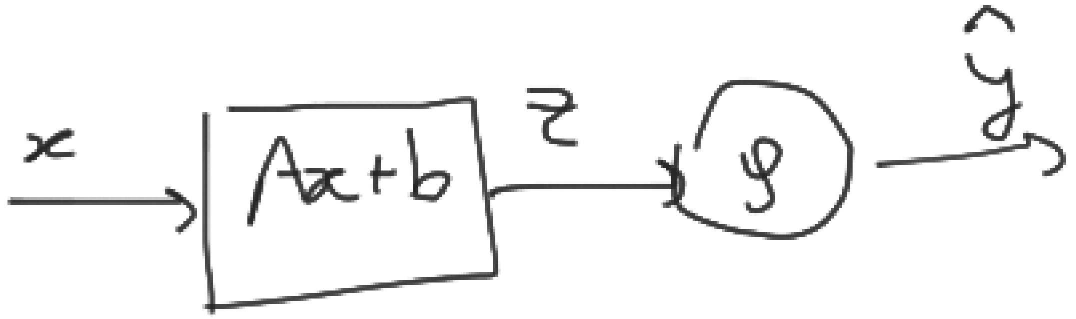
Ejemplo (1)



$$x = \begin{pmatrix}1\\0\\2\end{pmatrix} \in \mathbb{R}^3$$

Neurona



$$y = 5$$

$$z = Wx + b$$
$$\hat{y} = g(z)$$

$\hat{y} \in \mathbb{R} \to b, z \in \mathbb{R}$

si $x \in \mathbb{R}^n \Rightarrow W \in \mathbb{R}^{1\times n}$

x ej. $W = (4,2,-5)$, $b = 7$, $g(z) = \mathrm{ReLU}(z) = \mathrm{máx}(z,0)$

$$\Rightarrow \boxed{z = \underbrace{(4,2,-5)}_{W} \cdot \underbrace{\begin{pmatrix}1\\0\\2\end{pmatrix}}_{x} + \underbrace{7}_{b} = \underbrace{4\cdot 1 + 2\cdot 0 + (-5)\cdot 2}_{W\cdot x = \sum_{i=1}^{n} W_i \cdot x_i} + 7 = 1}$$

$$\hat{y} = 1$$

Ejemplo (1)

$$
\begin{tikzcd}
x \arrow[r] & \boxed{h} \arrow[r] & \overset{1}{\hat{y}} \arrow[r] & \mathcal{L}(\hat{y}, y) \\
\underset{5}{y} \arrow[urrr] & & &
\end{tikzcd}
$$

Supongamos

$$\mathcal{L}(\hat{y}, y) = \frac{1}{2}(\hat{y} - y)^2$$

$$\mathcal{L} : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$$

$$\frac{d\mathcal{L}}{d\hat{y}} = 2 \cdot (\hat{y} - y) \cdot \frac{1}{2} = \hat{y} - y$$

en el ej.

$$\mathcal{L}(1, 5) = \frac{1}{2}(1 - 5)^2 = 8$$

$$\frac{d\mathcal{L}}{d\hat{y}} = 1 - 5 = -4$$

Ejemplo (1)

$$\begin{tikzcd} x \arrow[r] & \boxed{h} \arrow[r] & \hat{y} \arrow[r] & \mathcal{L}(\hat{y}, y) \\ y \arrow[urrr] & & & \end{tikzcd}$$

Supongamos

$$\mathcal{L}(\hat{y}, y) = \frac{1}{2}(\hat{y} - y)^2$$

$$\mathcal{L} : \mathbb{R} \times \mathbb{R} \longrightarrow \mathbb{R}$$

$$\frac{d\mathcal{L}}{d\hat{y}} = 2 \cdot (\hat{y} - y) \cdot \frac{1}{2} = \hat{y} - y \in \mathbb{R}$$

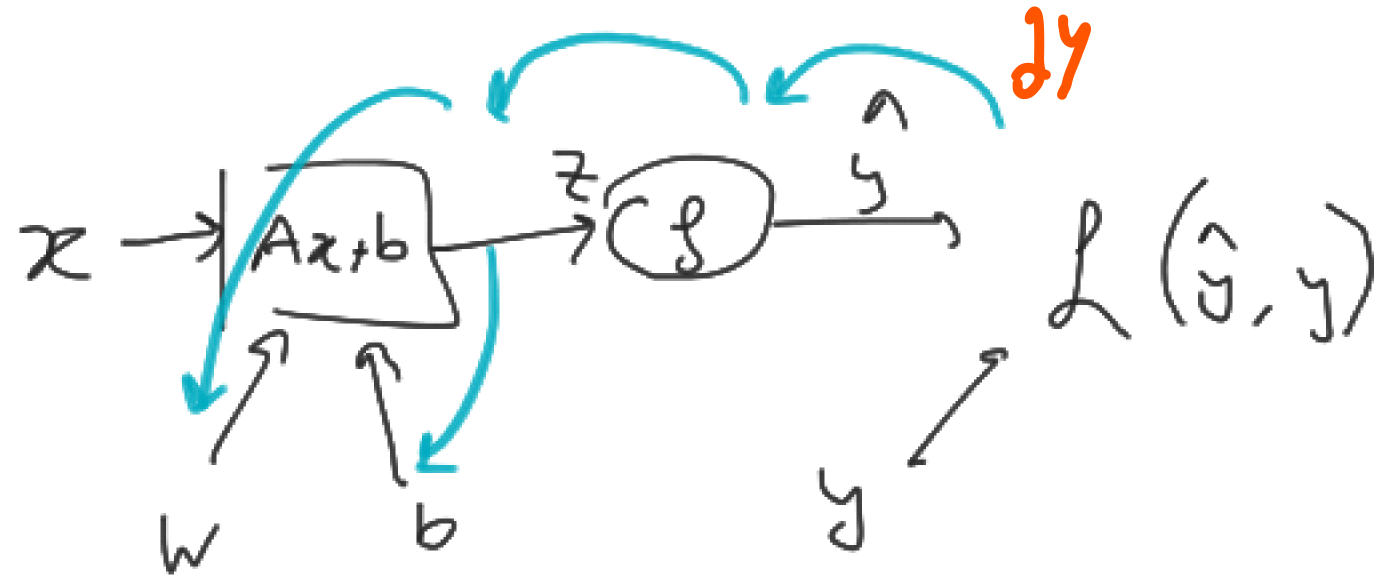
en el ej.

$$\mathcal{L}(1,5) = \frac{1}{2}(1-5)^2 = 8$$

$$\frac{d\mathcal{L}}{d\hat{y}} = 1 - 5 = -4$$ → esto va a ser dY

Volviendo a la neurona

Problema: quiero calcular las derivadas de L respecto de W y b



$$g(z) = \begin{cases} z & \text{si } z \geq 0 \\ 0 & \text{si no} \end{cases}$$

$$\Downarrow$$

$$g'(z) = \begin{cases} 1 & \text{si } z \geq 0 \\ 0 & \text{si no} \end{cases}$$

Tengo $dy = \frac{d\mathcal{L}}{d\hat{y}}$

en el ej.

$$dz = \frac{d\mathcal{L}}{dz} = \underbrace{\frac{d\mathcal{L}}{d\hat{y}}}_{dy} \cdot \underbrace{\frac{d\hat{y}}{dz}}_{g'(z)} = (-4) \cdot \underbrace{1}_{g'(1)} = -4$$

Seguimos

$$\boxed{db = \frac{d\mathcal{L}}{db} = \underbrace{\frac{d\mathcal{L}}{d\hat{y}} \cdot \frac{d\hat{y}}{dz}}_{dz} \cdot \underbrace{\frac{dz}{db}}_{1} = dz}$$

$$\boxed{dW = \frac{d\mathcal{L}}{dW} = dz \cdot \frac{dz}{dW} = dz \cdot x^T}$$

$$(= x^T \cdot dz)$$

! $z = w_1 \cdot x_1 + w_2 \cdot x_2 + \ldots + w_n \cdot x_n + b$

$$\Rightarrow \frac{dz}{dw_i} = x_i \quad \Rightarrow \nabla z(w) = (x_1, \ldots, x_n) = x^T$$

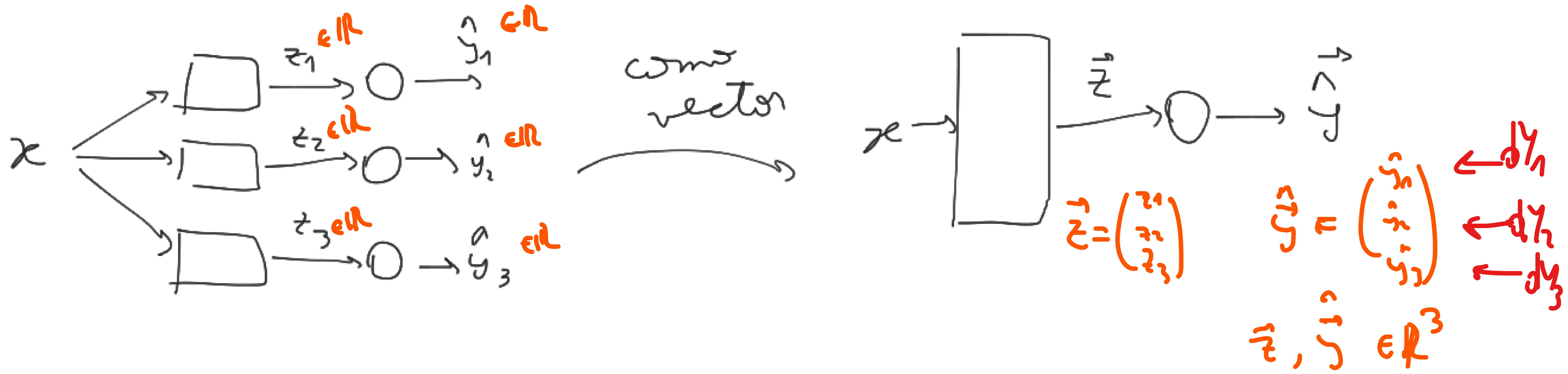
idem

$$\boxed{dx} = \frac{d\mathcal{L}}{dx} = dz \cdot \frac{dz}{dx}$$

$$= dz \cdot W^T$$

$$= \boxed{W^T \cdot dz}$$

Si son muchas neuronas



Si tengo k neuronas apiladas, me queda (recordar salida escalar de c/u):

$$\hat{y} \in \mathbb{R}^k \Rightarrow z, b \in \mathbb{R}^k$$

$$W \in \mathbb{R}^{k \times n}$$

ej: $W = \begin{pmatrix} — w^{(1)} — \\ — w^{(2)} — \end{pmatrix} = \begin{pmatrix} 4 & 2 & -5 \\ 7 & -1 & 0 \end{pmatrix}$

$k = 2$

$$b = \begin{pmatrix} b^{(1)} \\ b^{(2)} \end{pmatrix} = \begin{pmatrix} 7 \\ -2 \end{pmatrix}$$

Idea respecto de dX

$z_1 = \ldots + W_{1,i} \cdot x_i + \ldots \in \mathbb{R}$

$z_2 = \ldots + W_{2,i} \cdot x_i + \ldots \in \mathbb{R}$

$\vdots$

$z_k = \ldots + W_{k,i} \cdot x_i + \ldots \in \mathbb{R}$

$z \in \mathbb{R}^k$

li soy x_i

$(i = 1, \ldots, n)$

$W = \begin{pmatrix} W_{11} & W_{12} & W_{13} \\ W_{21} & W_{22} & W_{23} \end{pmatrix} \cdot X = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}$

$$\Rightarrow \frac{dz}{dx_i} = \sum_{j=1}^{k} W_{j,i}$$

como prod. escalar en $\mathbb{R}^k$

$$\Rightarrow \frac{dl}{dx_i} = \sum_{j=1}^{k} \underbrace{\frac{dl}{dz_j}}_{\in \mathbb{R}} \cdot \underbrace{W_{j,i}}_{\in \mathbb{R}} = \langle \overbrace{\frac{dl}{dz}}^{dz}, W_{:,i} \rangle = dX_i$$

$= W_{i,:}^T \cdot dz$