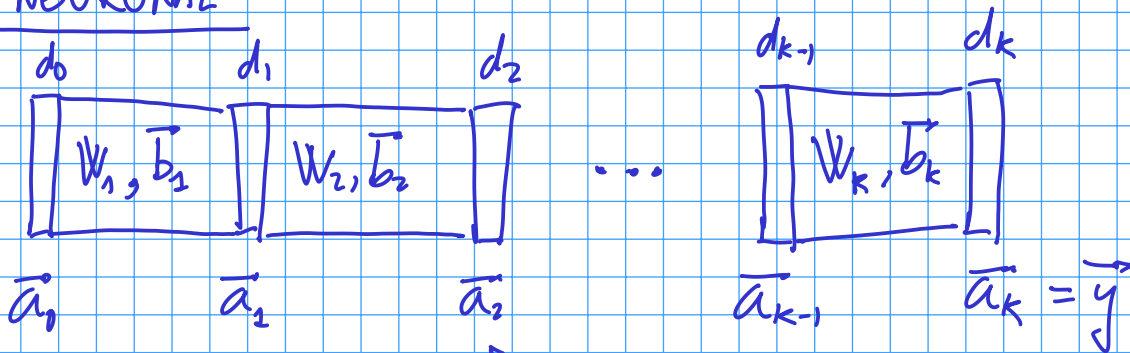


RED NEURONAL:



\$k+1\$ capas

\$d_j \equiv N^\circ\$ neuronas capa \$j\$

\$\vec{a}_j \equiv\$ activación capa \$j\$

$$\vec{a}_j = \begin{pmatrix} a_{j1} \\ a_{j2} \\ \vdots \\ a_{jd_j} \end{pmatrix}$$

$$\vec{a}_0 = \vec{x}$$

\$(d_0 \times 1)\$

$$\vec{z}_1 = W_1 \vec{a}_0 + \vec{b}_1$$

$$\vec{a}_1 = f(\vec{z}_1)$$

$$(d_1 \times 1)$$

$$W_1 (d_1 \times d_0)$$

$$\vec{b}_1 (d_1 \times 1)$$

$$\vec{z}_2 = W_2 \vec{a}_1 + \vec{b}_2$$

$$\vec{a}_2 = f(\vec{z}_2)$$

\$\dots\$

$$d_0 \begin{bmatrix} \vdots \\ 1 \end{bmatrix}$$

$$\begin{matrix} \vec{z}_1 \\ d_1 \times 1 \end{matrix} = \underbrace{\begin{matrix} W_1 & @ & \vec{a}_0 \\ d_1 \times d_0 & & d_0 \times 1 \end{matrix}}_{d_1 \times 1} + \begin{matrix} \vec{b}_1 \\ d_1 \times 1 \end{matrix}$$

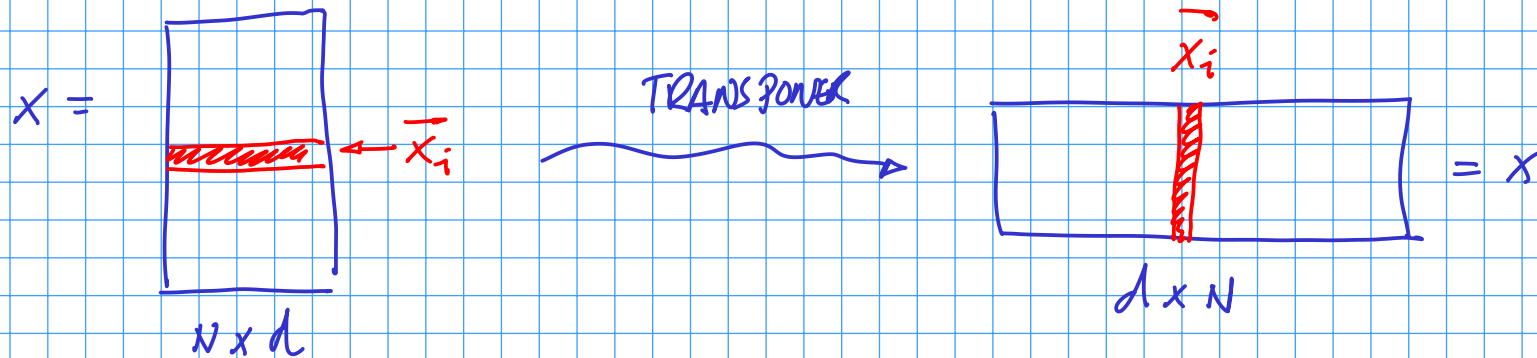
¿Número de parámetros?

capa 1 : \$W_1 (d_1 \times d_0) \ b_1 (d_1 \times 1) \rightarrow d_1 \cdot d_0 + d_1\$

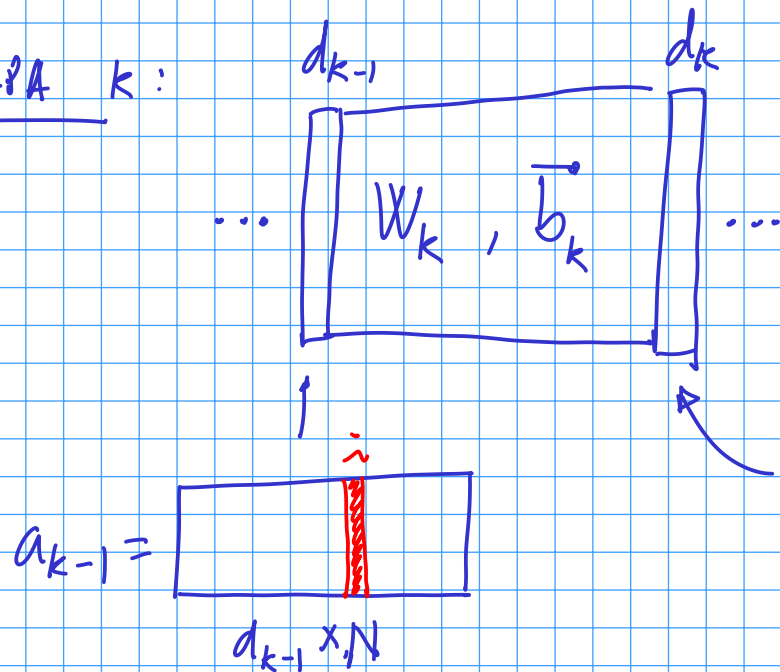
capa 2 : \$W_2 (d_2 \times d_1) \ b_2 (d_2 \times 1) \rightarrow d_2 \cdot d_1 + d_2\$

\$\vdots\$

PROCESAMIENTO POR LOTES (BATCHES):



CAPA k:



$$W_k (d_k \times d_{k-1})$$

$$\vec{b}_k (d_k \times 1)$$

$$a_k = f(W_k a_{k-1} + b_k)$$

$$\underbrace{\begin{bmatrix} W_k \end{bmatrix}}_{d_k \times d_{k-1}} \otimes \begin{bmatrix} \phantom{a_{k-1}} \end{bmatrix}_{d_{k-1} \times N} + \begin{bmatrix} \phantom{a_{k-1}} \end{bmatrix}_{d_k \times 1} = \begin{bmatrix} a_k \end{bmatrix}_{d_k \times N}$$

FORMATO MÁS HABITUAL:

$$X = \begin{array}{|c|} \hline \square \\ \hline \end{array} \quad N \times d$$

$$a_k = \begin{array}{|c|} \hline \square \\ \hline \end{array} \quad N \times d_k$$

$$a_k = f(a_{k-1} W_k^T + b_k^T)$$

$N \times d_{k-1}$ $d_{k-1} \times d_k$ $1 \times d_k$

$N \times d_k$

¿POR QUÉ ES IMPORTANTE LA F. ACTIVACIÓN?

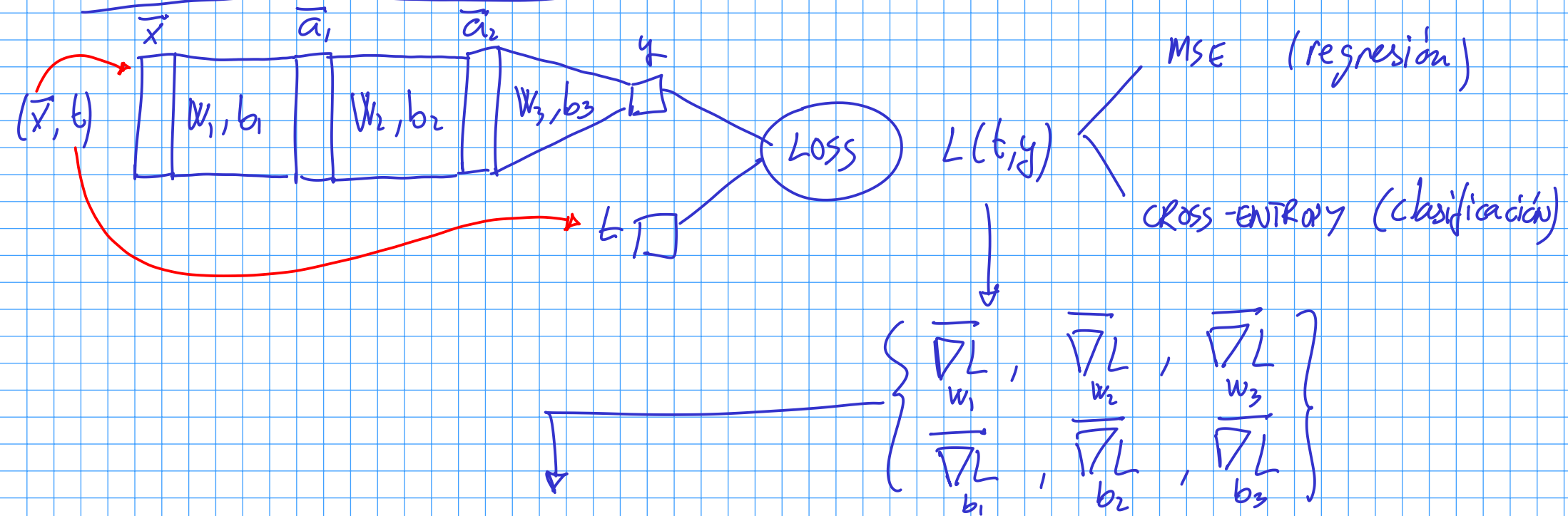
$$\vec{x} \rightarrow \vec{a}_1 = f(W_1 \vec{x} + \vec{b}_1) \rightarrow \vec{a}_2 = f(W_2 \vec{a}_1 + \vec{b}_2) \rightarrow \dots$$

(Si No)
AA7 f): $\vec{x} \rightarrow \vec{a}_1 = W_1 \vec{x} + b_1 \rightarrow \vec{a}_2 = W_2 \vec{a}_1 + b_2 = W_2 (W_1 \vec{x} + \vec{b}_1) + \vec{b}_2 =$

$$= \underbrace{W_2 W_1}_{\vec{W}} \vec{x} + \underbrace{W_2 \vec{b}_1 + \vec{b}_2}_{\vec{b}} = \vec{W} \vec{x} + \vec{b}$$

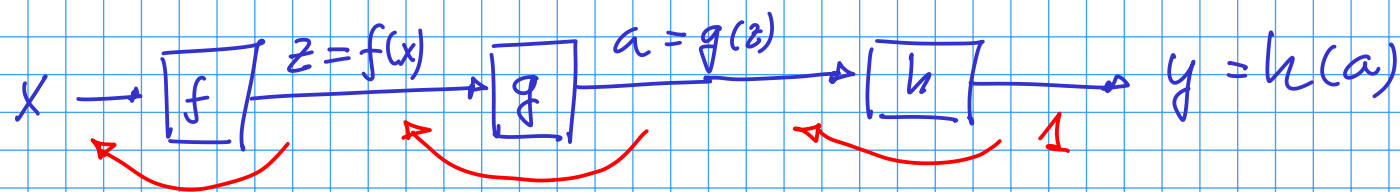
LINEAL!!!

¿CÓMO ENTRENAMOS LA RED?



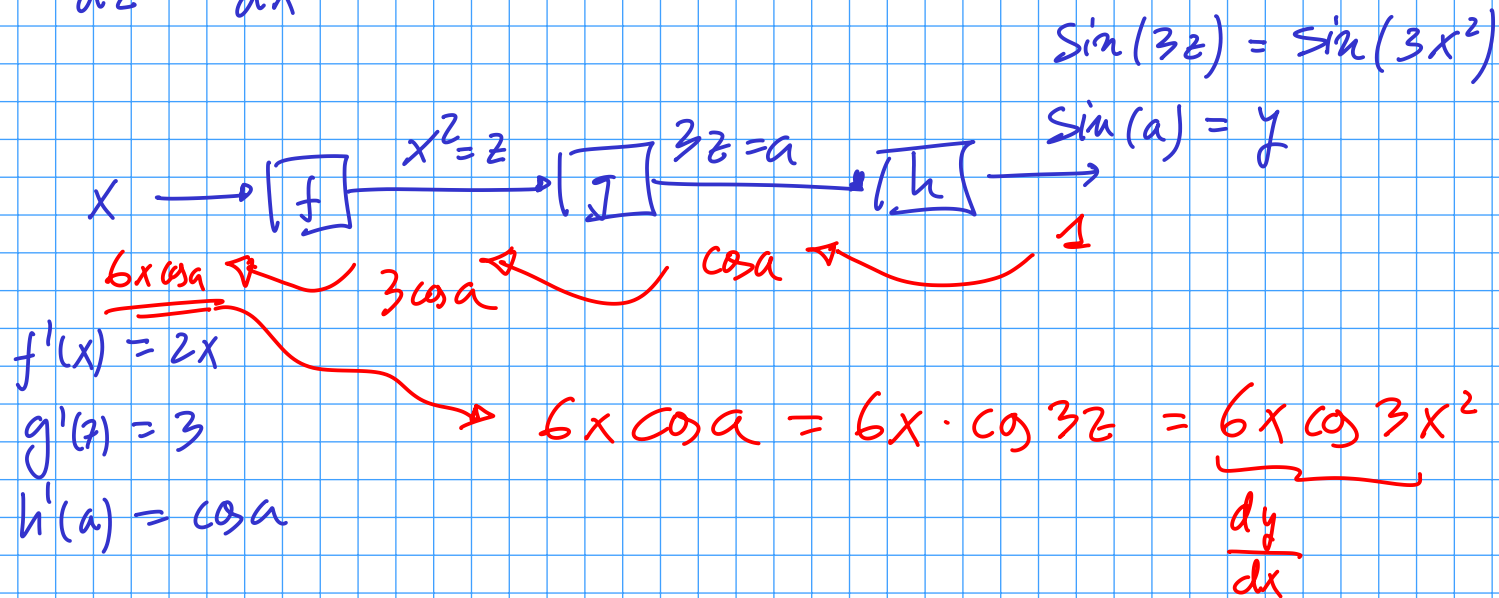
$$w_1 \leftarrow w_1 - \eta \cdot \frac{\partial L}{\partial w_1}$$
$$w_2 \leftarrow w_2 - \eta \cdot \frac{\partial L}{\partial w_2}$$
$$\vdots$$

REGLA DE LA CADENA DE LA DERIVADA:

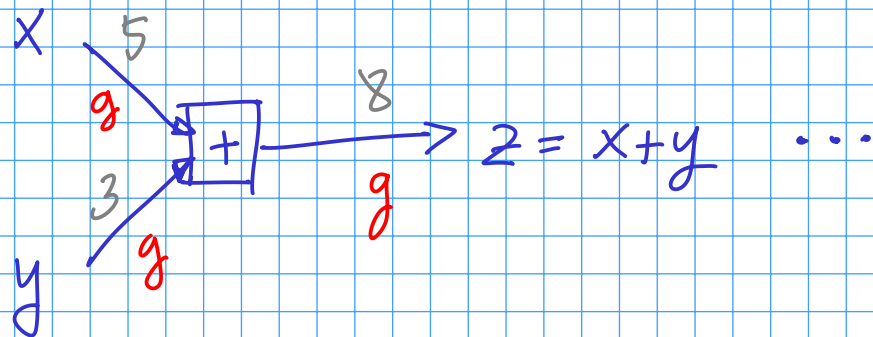


$$\frac{dy}{dx} = \underbrace{\frac{dy}{dy}}_1 \cdot \frac{dy}{da} \cdot \frac{da}{dz} \cdot \frac{dz}{dx}$$

$$\left. \begin{aligned} f(x) &= x^2 \\ g(z) &= 3z \\ h(a) &= \sin(a) \end{aligned} \right\}$$



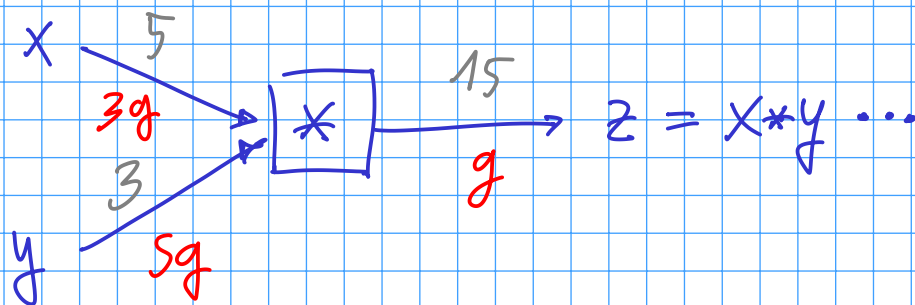
SUMA



$$\frac{dz}{dx} = 1$$

$$\frac{dz}{dy} = 1$$

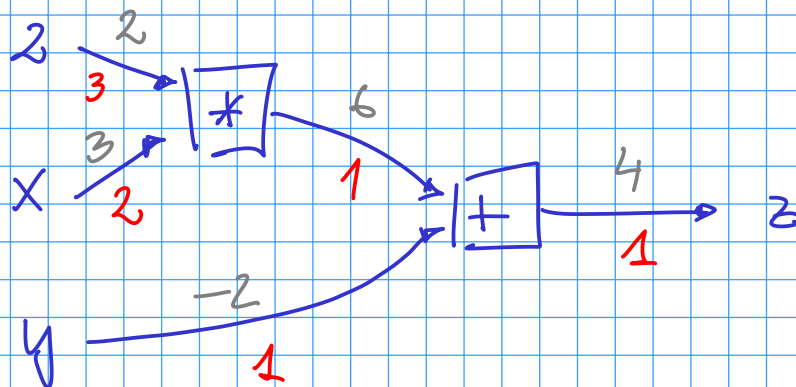
PRODUCTO



$$\frac{dz}{dx} = y$$

$$\frac{dz}{dy} = x$$

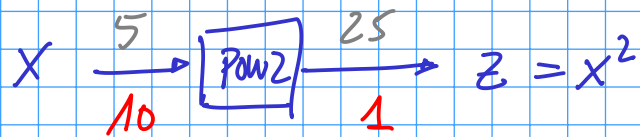
$$z = 2 * x + y$$



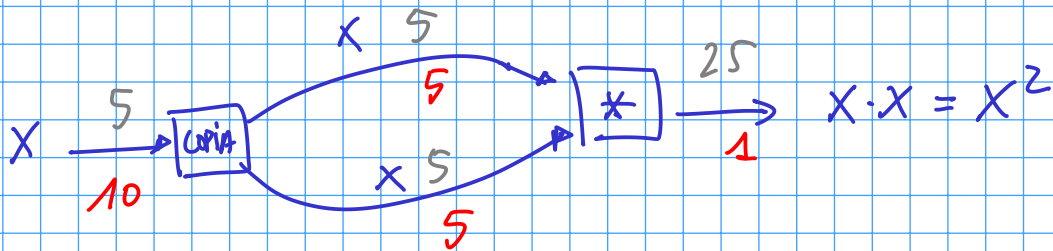
$$\frac{dz}{dx} = 2$$

$$\frac{dz}{dy} = 1$$

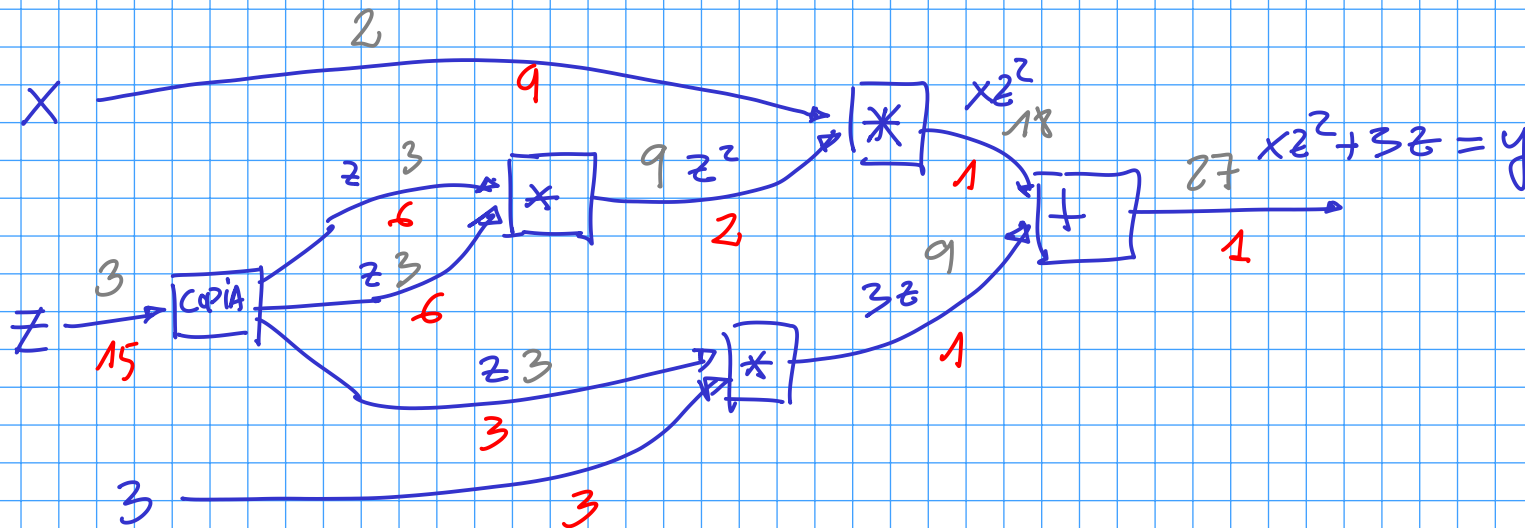
$$z = x^2$$



$$\frac{dx^2}{dx} = 2x$$



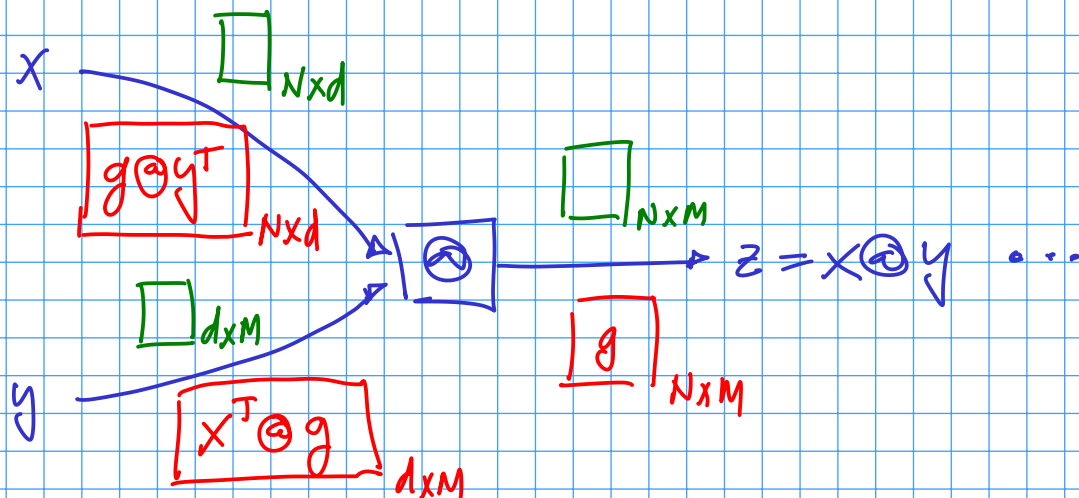
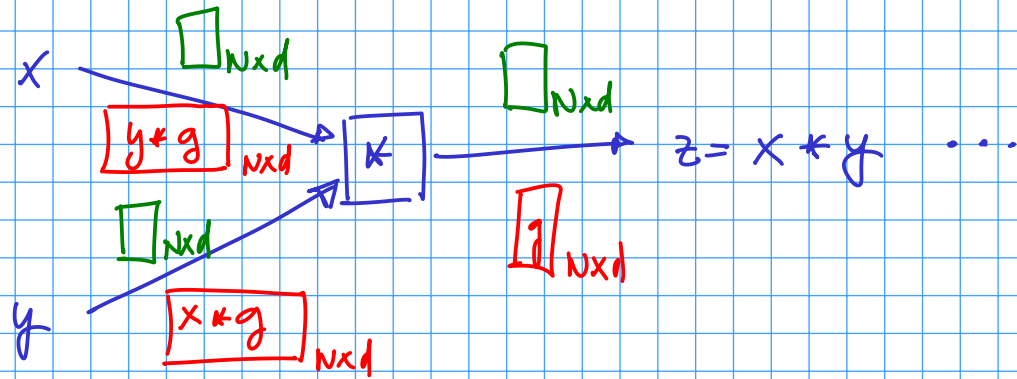
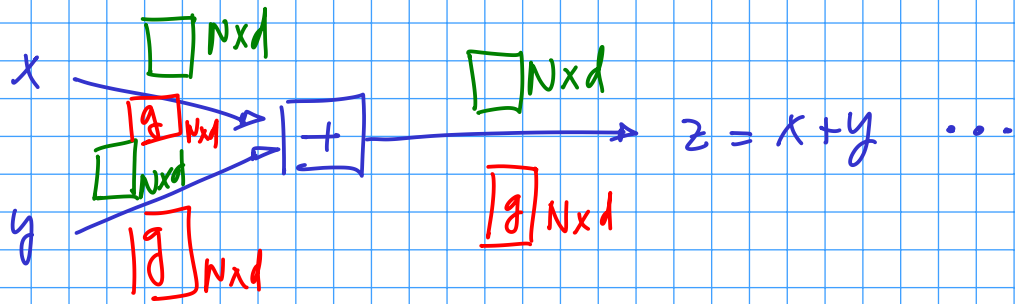
$$y = (xz^2 + 3z)$$



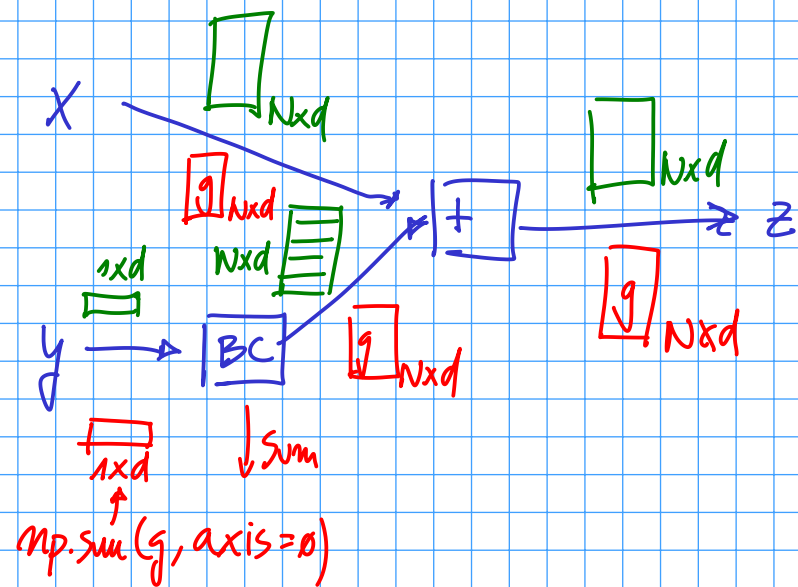
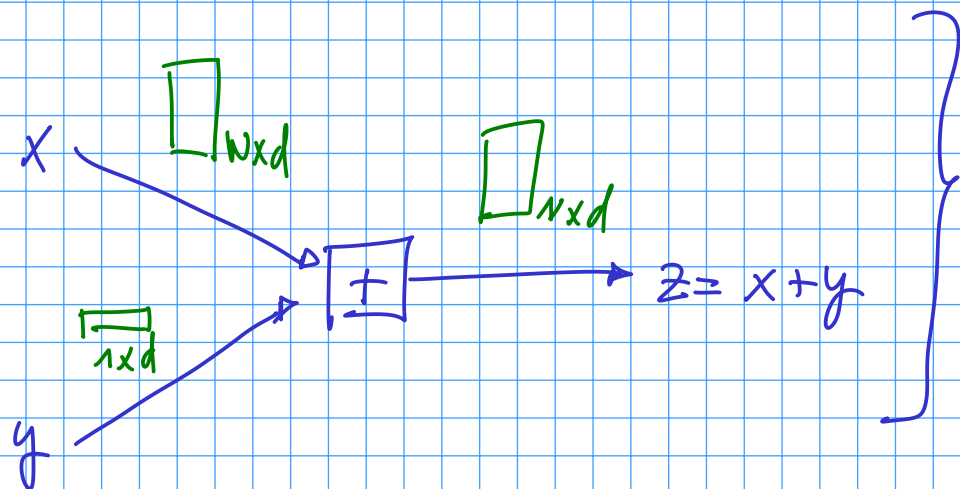
$$\frac{dy}{dx} = 9$$

$$\frac{dy}{dz} = 15$$

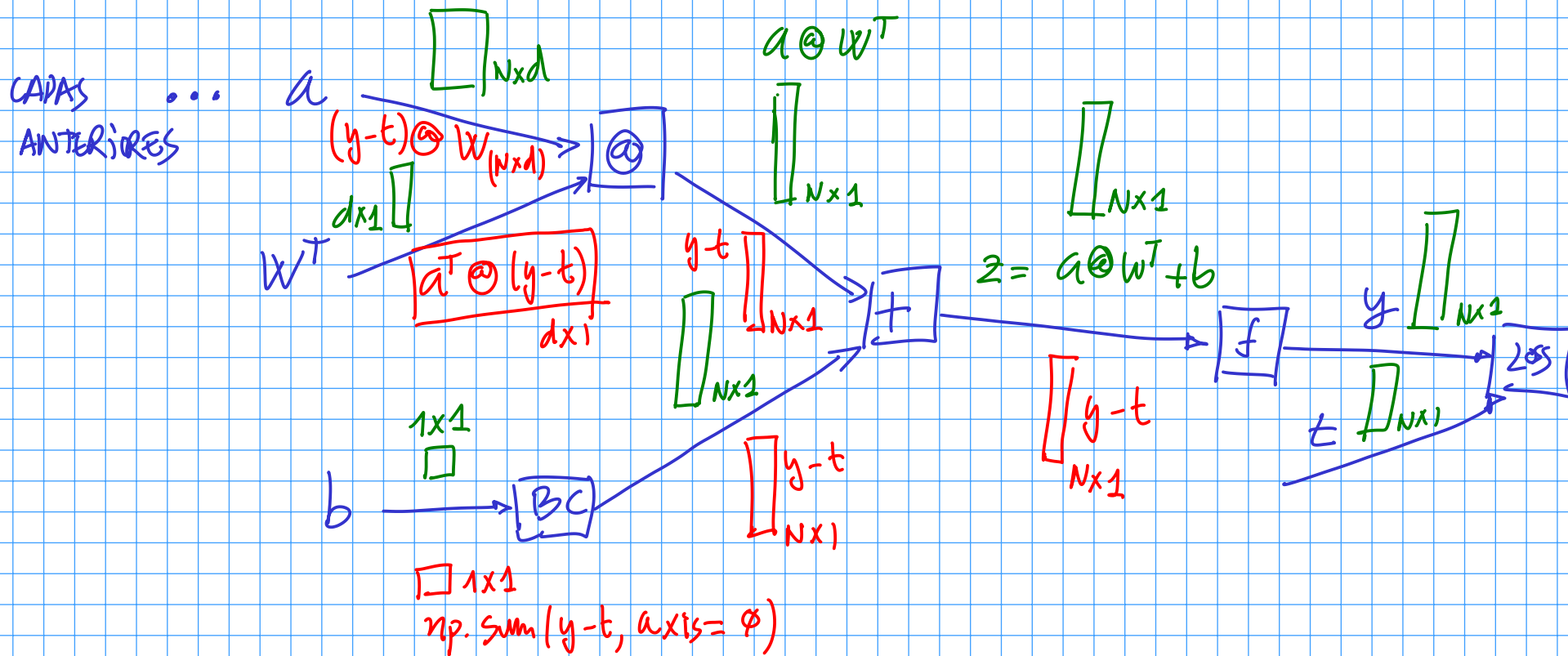
¿QUÉ PASA CON ARRAYS?



BROADCASTING:



RED NEURONAL ÚLTIMA CAPA (REGR. LINEAL / LOGÍSTICA)



RED NEURONAL CAPA INTERMEDIA

