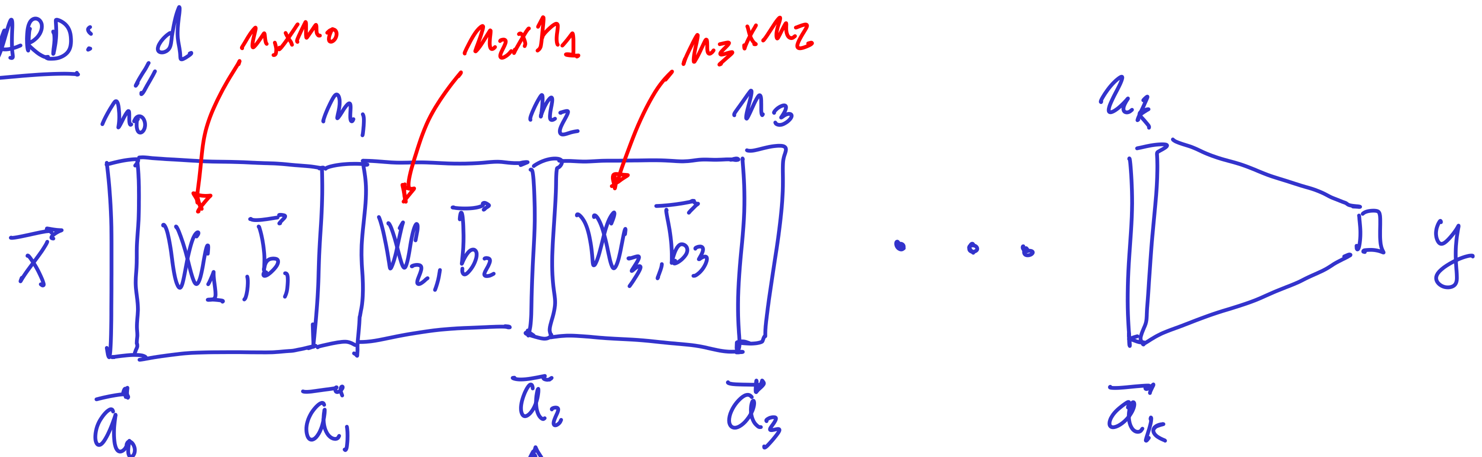
FORWARD:



k capas ocultas

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

$n_0 \times 1$

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

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

$n_1 \times 1$

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

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

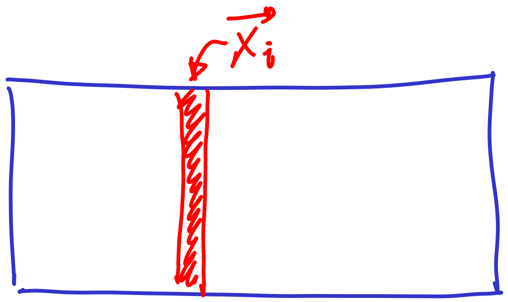
$n_2 \times 2$

$$\underset{n_1 \times 1}{z_1} = \underset{n_1 \times n_0}{W_1} \cdot \underset{n_0 \times 1}{a_0} + \underset{n_1 \times 1}{b_1}$$

$\xrightarrow{\hspace{6em}}$

FORWARD

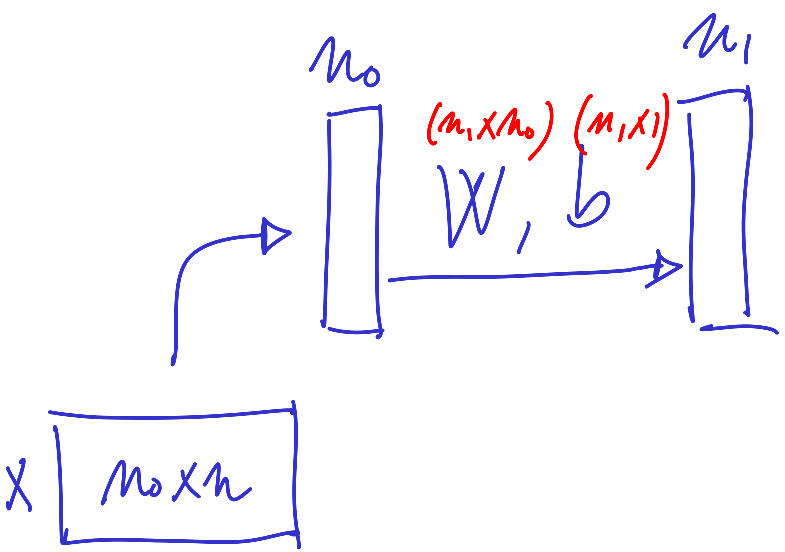
PROCESAMIENTO POR LOTES (BATCHES):

$X =$ 

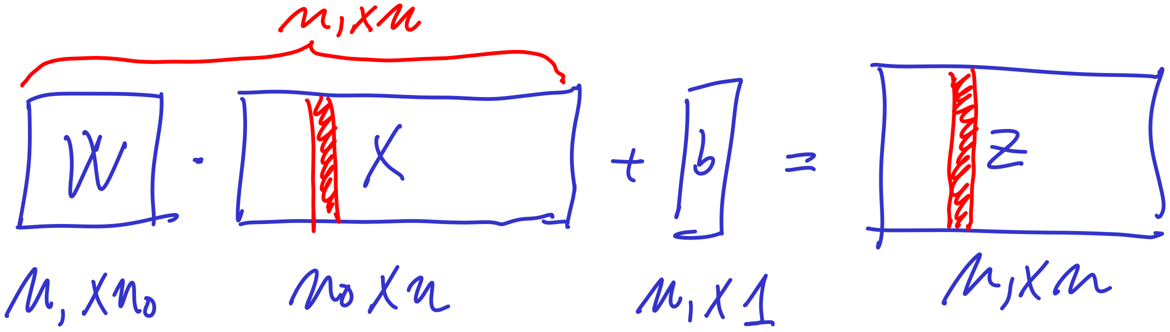
$n_0 \times n$

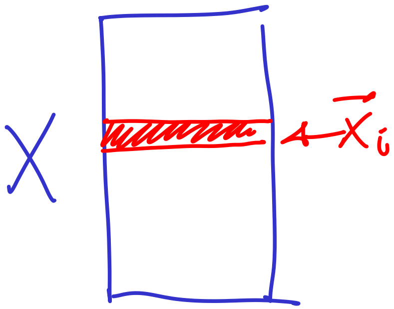
dimensión de cada vector

nº de patrones en el batch



$W \cdot X + b$



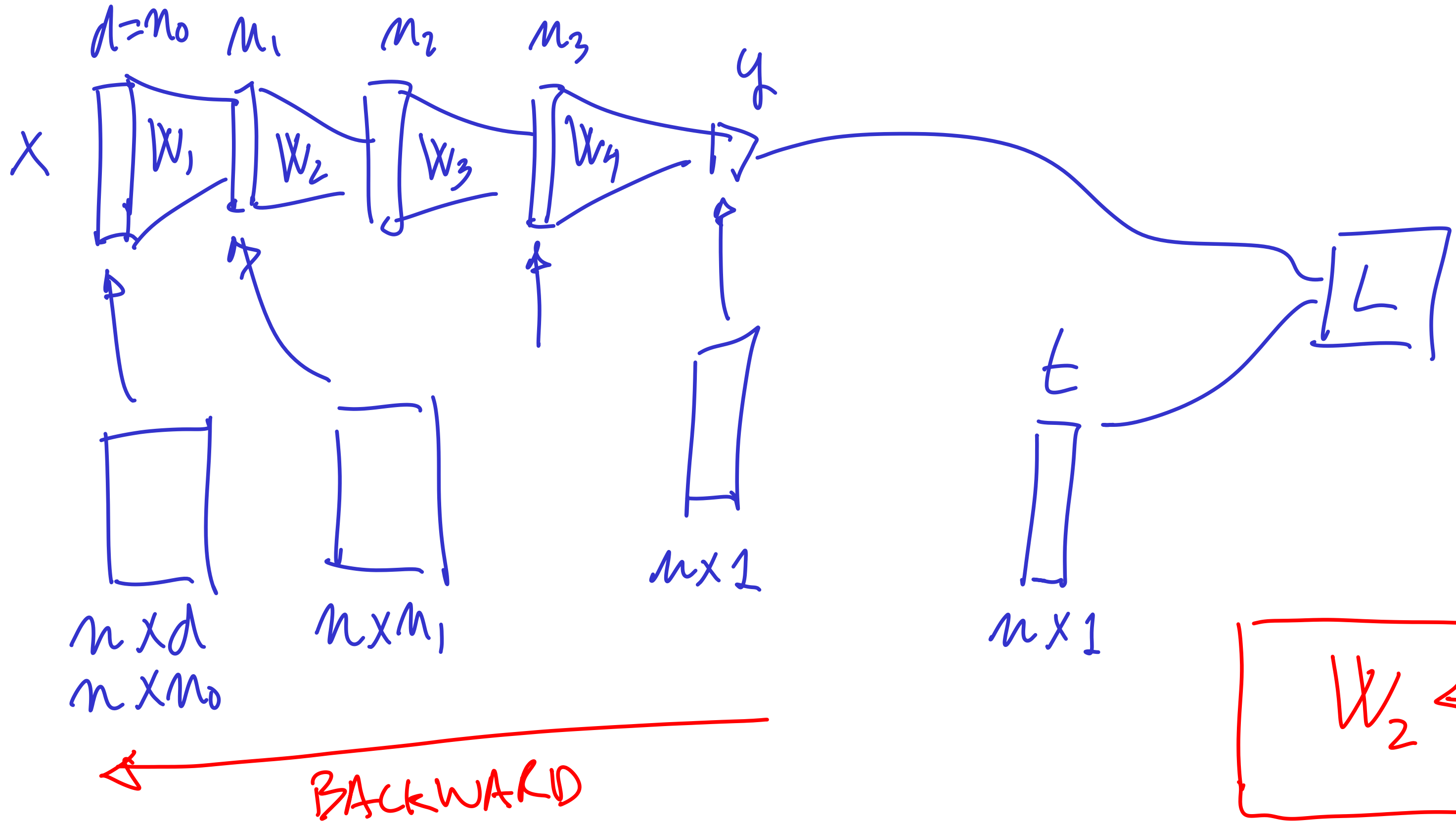


$n \times n_0$

$n \times d$

$$\underbrace{X}_{n \times n_0} \cdot \underbrace{W^T}_{n_0 \times n_1} + \underbrace{b^T}_{1 \times n_1}$$

¿CÓMO ENTRENAMOS LA RED?

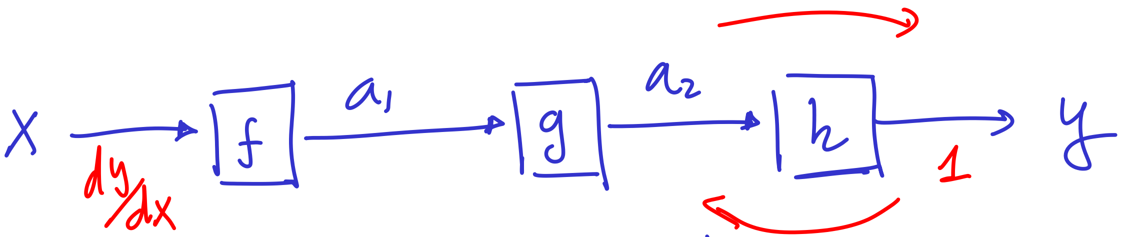


$$\frac{1}{2}\sum_{i=1}^{n}(y_i - t_i)^2 \quad \text{(MSE)}$$

$$\nabla_{W_4} L, \ \nabla_{W_3} L, \ \nabla_{W_2} L, \ \nabla_{W_1} L$$

$$\boxed{W_2 \leftarrow W_2 - \eta \cdot \nabla_{W_2} L}$$

REGLA DE LA CADENA:

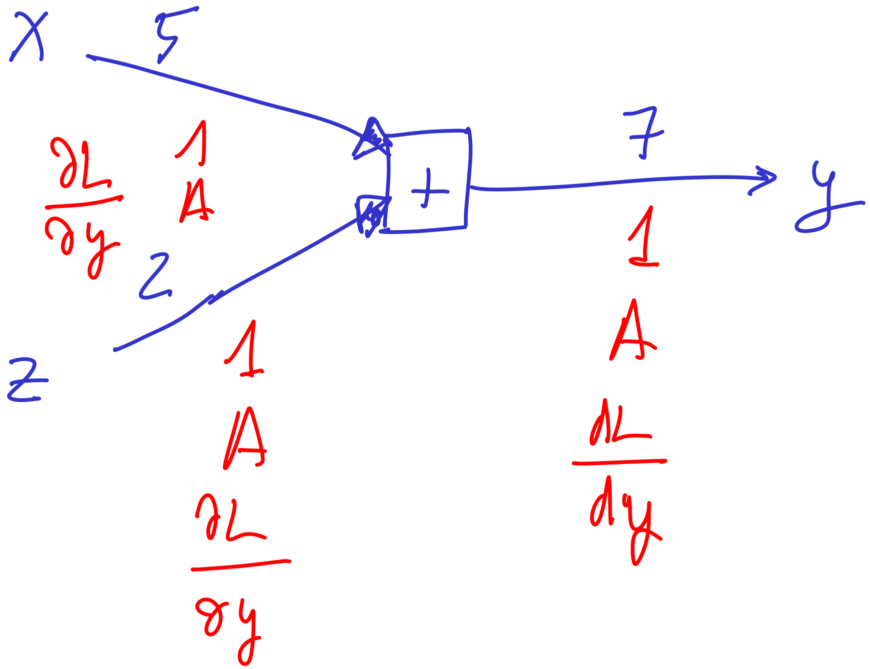


$$y = h(a_2) = h(g(f(x)))$$

$$a_2 = g(a_1) = g(f(x))$$

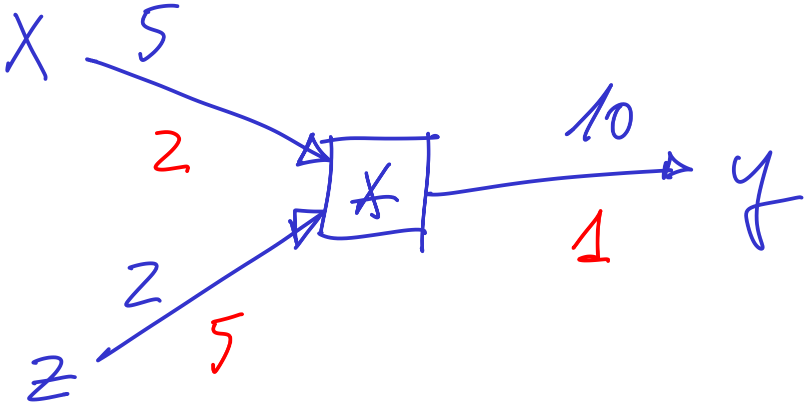
$$a_1 = f(x)$$

$$\frac{dy}{dx} = \underbrace{\boxed{\frac{dy}{dy}}}_{1} \cdot \frac{dy}{da_2} \cdot \frac{da_2}{da_1} \cdot \frac{da_1}{dx}$$



$$y = x + z$$

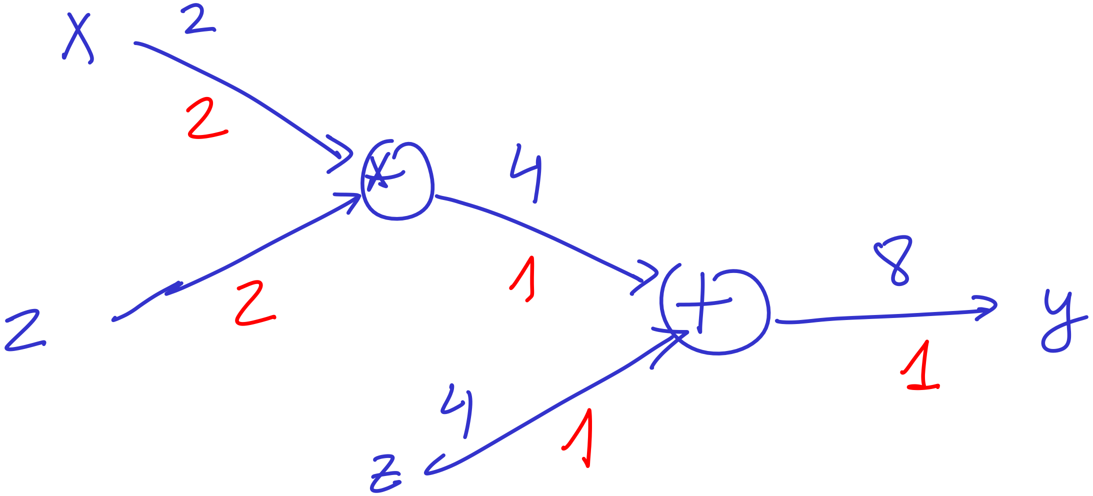
$$\frac{dy}{dx} = 1 \qquad \frac{dy}{dz} = 1$$



$$y = x \cdot z$$

$$\frac{dy}{dx} = z \quad , \quad \frac{dy}{dz} = x$$

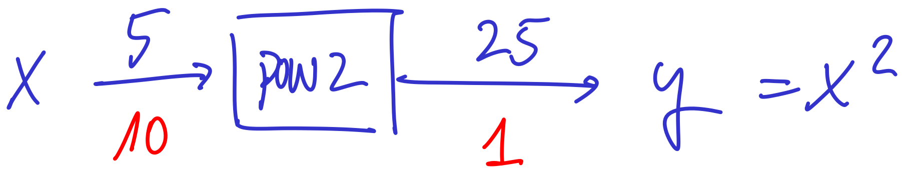
$$y = 2x + z$$



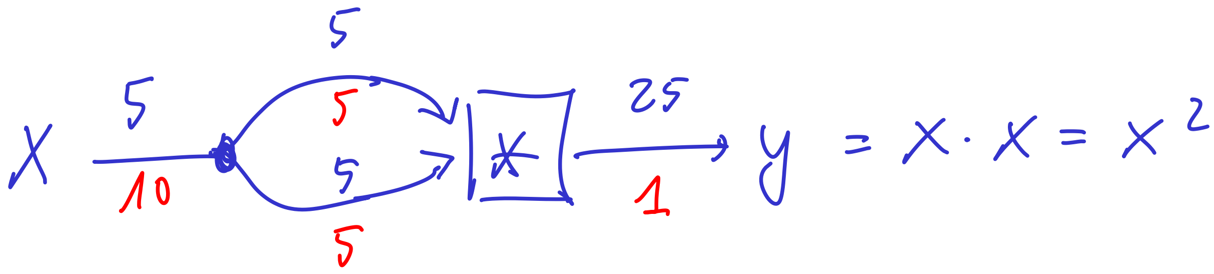
$$\frac{dy}{dx} = 2$$

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

$y = x^2$

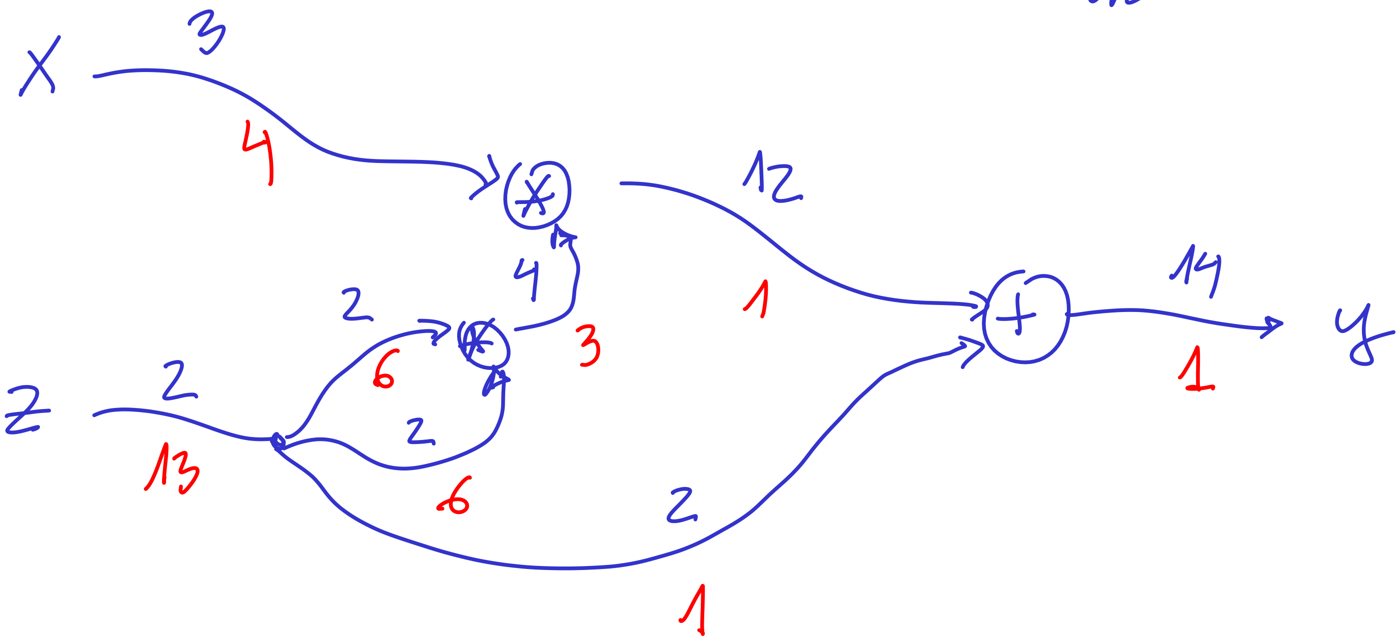


$$\frac{dy}{dx} = 2x$$

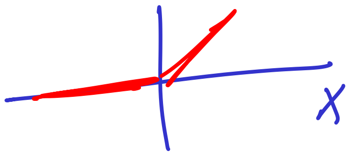


$y = (x \cdot z^2 + z)$

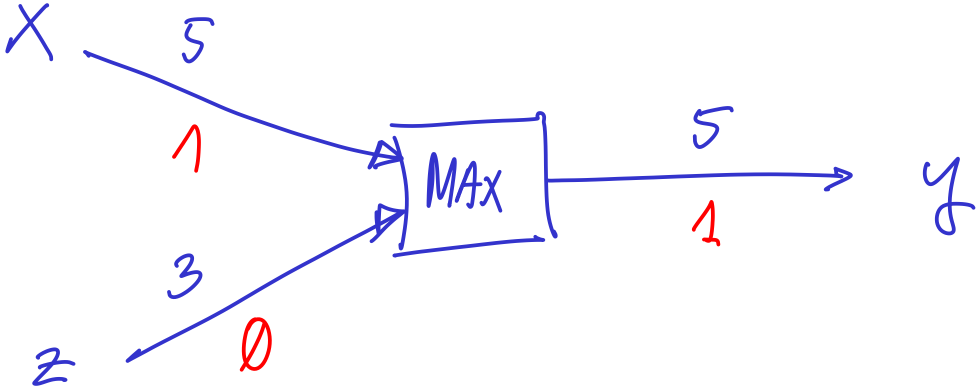
$$\frac{dy}{dz} = 2xz + 1$$



ReLU →

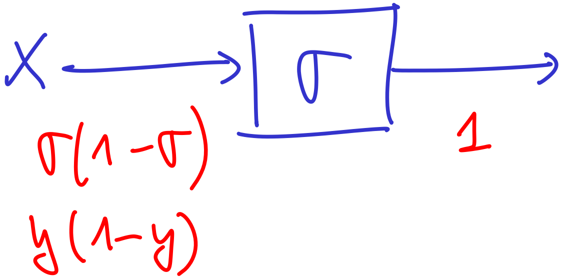


$$ReLU(x) = max(0, x)$$



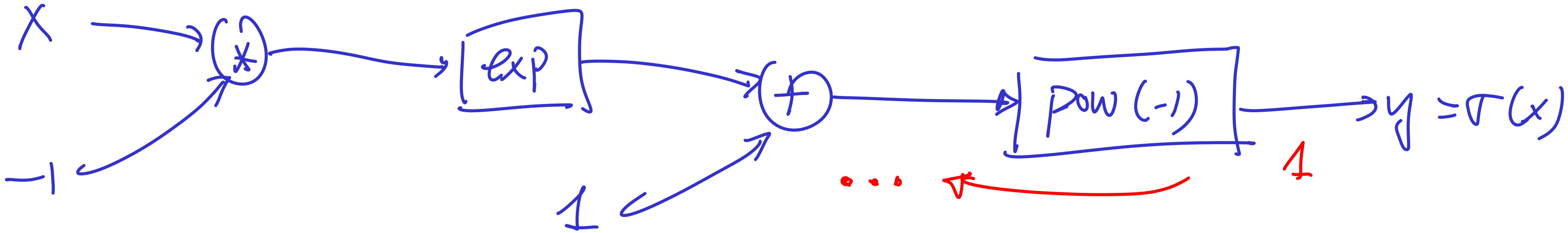
$$max(x,z) = \begin{cases} x & si \quad x \geq z \\ z & si \quad x < z \end{cases}$$

$$\frac{dy}{dx} = \begin{cases} 1 & si \quad x \geq z \\ 0 & si \quad < z \end{cases}$$

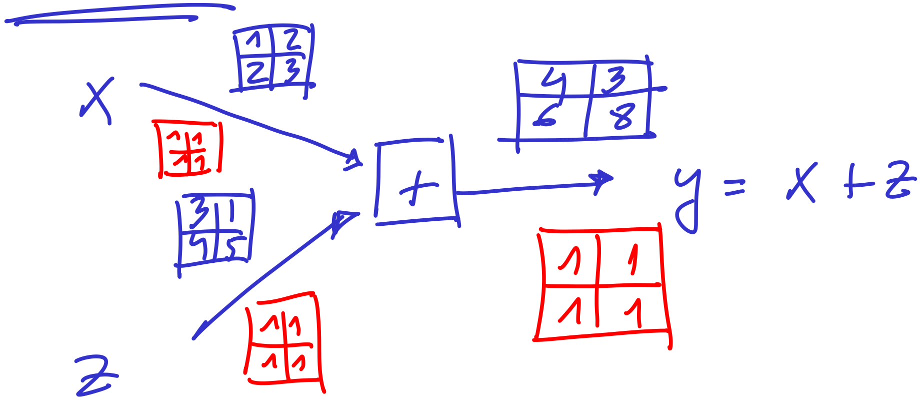


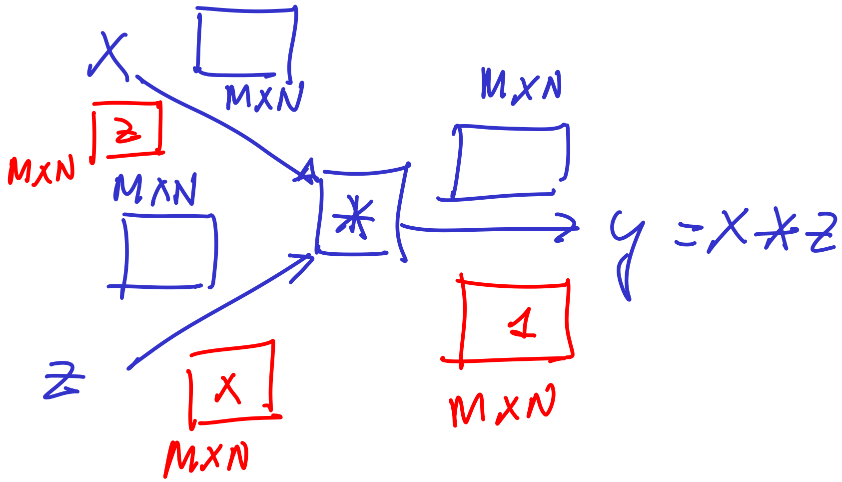
$$y = \sigma(x)$$

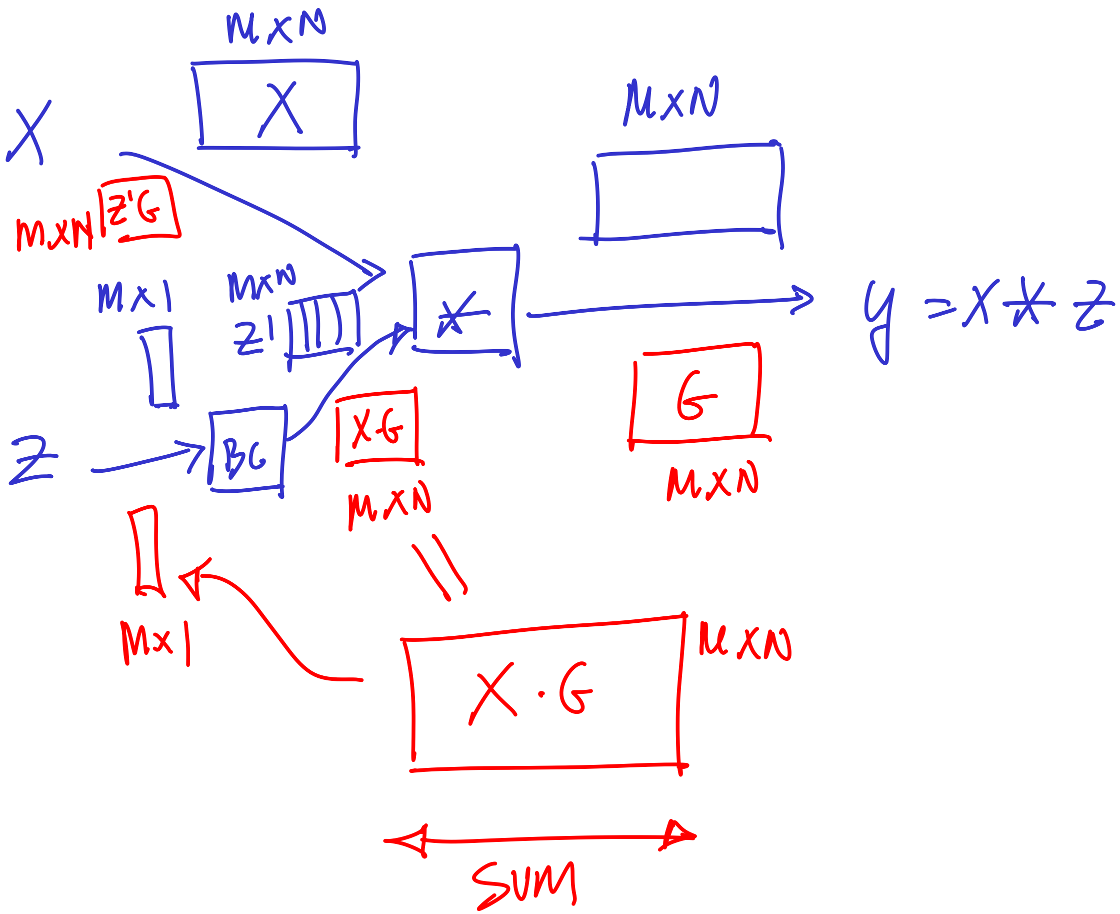
$$\sigma(x) = \frac{1}{1+e^{-x}}$$

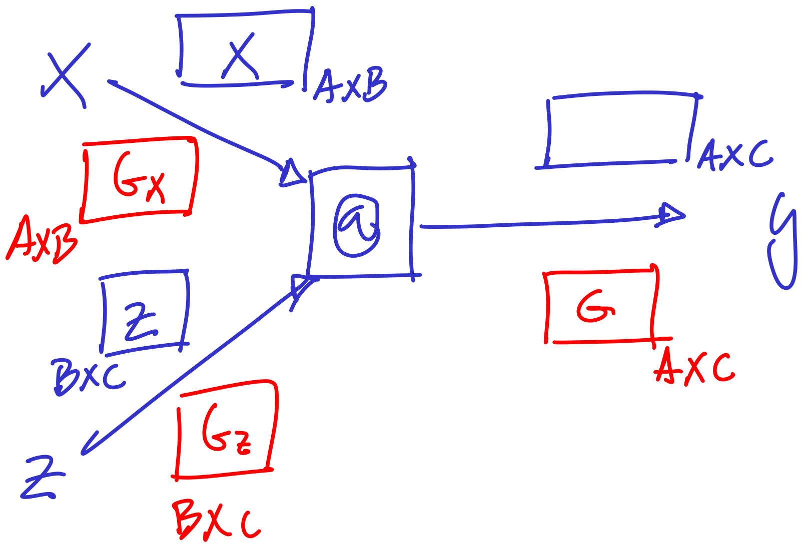


ARRAYS:









$$y = X @ Z$$

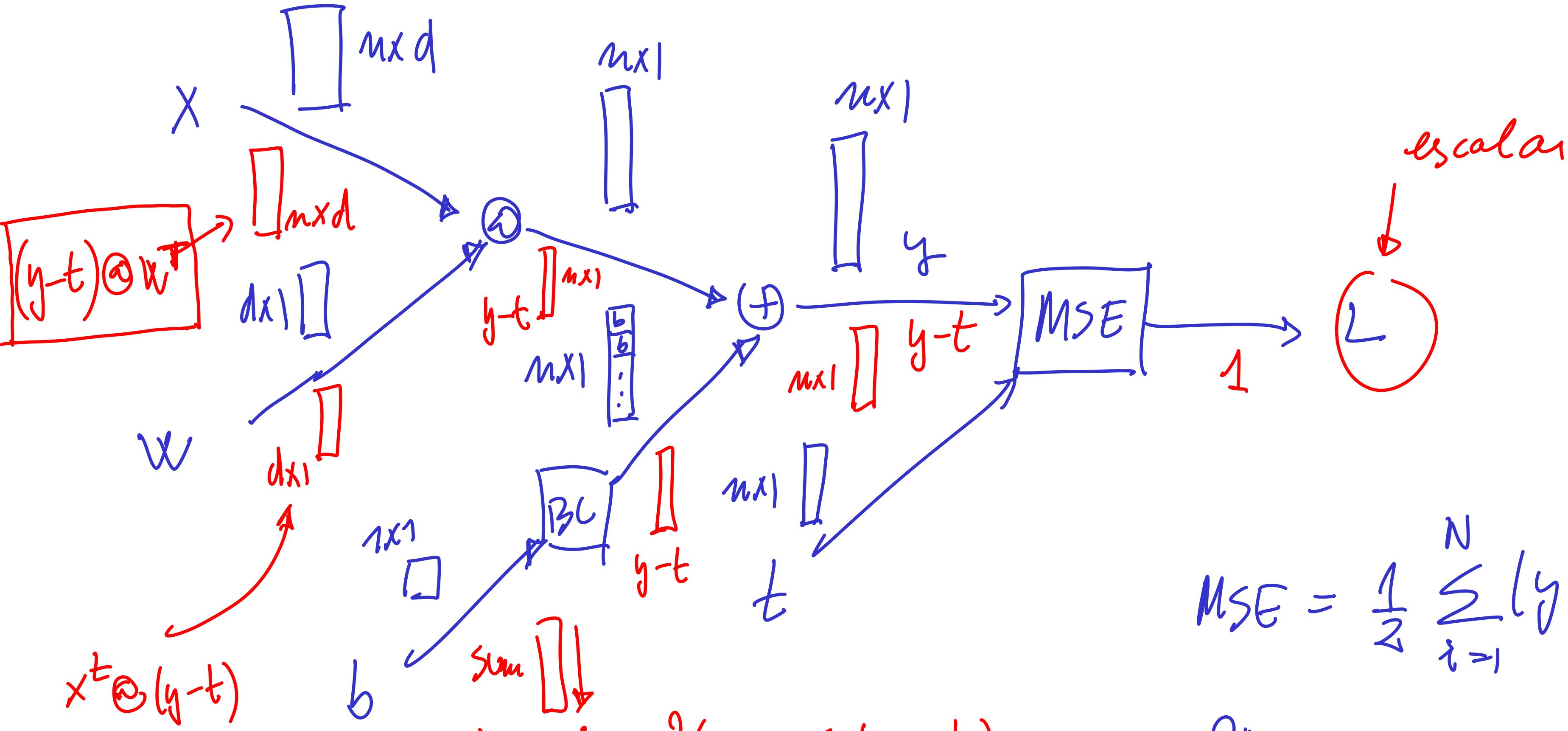
$$y_{ij} = \sum_k X_{ik} \cdot Z_{kj}$$

$$\underbrace{\boxed{G_X}}_{A\times B} = \underbrace{\boxed{G}}_{A\times C} @ \underbrace{\boxed{G_Z^T}}_{C\times B}$$

$$\underbrace{\boxed{G_Z}}_{B\times C} = \underbrace{\boxed{X^T}}_{B\times A} @ \underbrace{\boxed{G}}_{A\times C}$$

REGRESIÓN LINEAL:

$$\underbrace{y}_{} = \underbrace{X}_{m\times d} \cdot \underbrace{W}_{d\times 1} + \underbrace{b}_{1\times 1} = \underbrace{[\ ]}_{m\times 1}$$

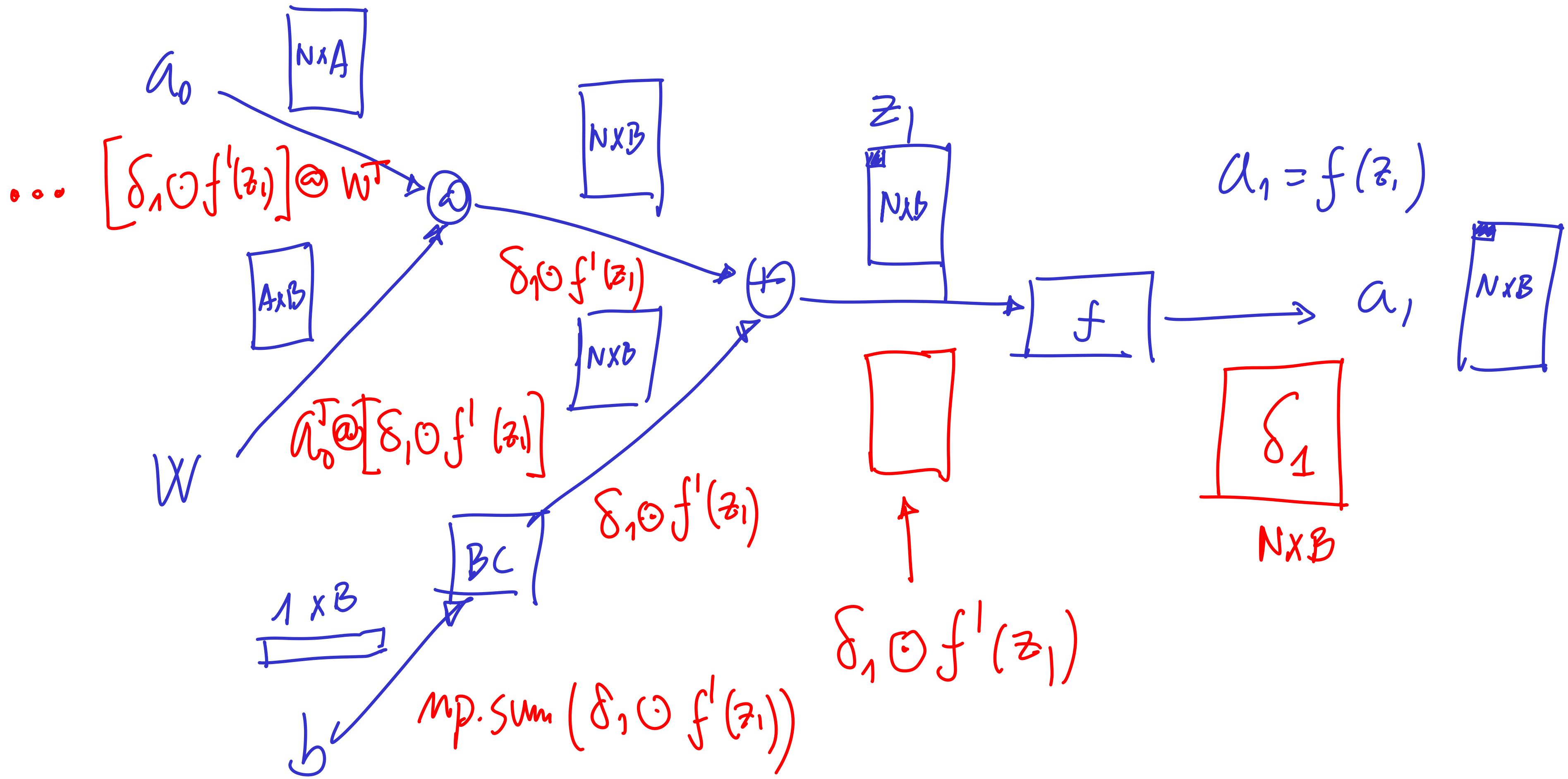


$$\frac{\partial L}{\partial b} = \sum_i (y_i - t_i)$$

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

$$\frac{\partial MSE}{\partial y_i} = (y_i - t_i)$$

$$\nabla_W L = X^t @ (y-t) \Rightarrow (\nabla_W L)_j = \sum_{i=1}^{N} X^T_{ij} \cdot (y_i - t_i)$$



a_0
N x A
N x B
z_1
N x B
$a_1 = f(z_1)$
N x B
$\dots [\delta_1 \odot f'(z_1)] @ W^T$
A x B
$\delta_1 \odot f'(z_1)$
N x B
$a_0^T @ [\delta_1 \odot f'(z_1)]$
W
f
a_1
δ_1
N x B
BC
1 x B
$\delta_1 \odot f'(z_1)$
$\delta_1 \odot f'(z_1)$
np.sum$(\delta_1 \odot f'(z_1))$
b