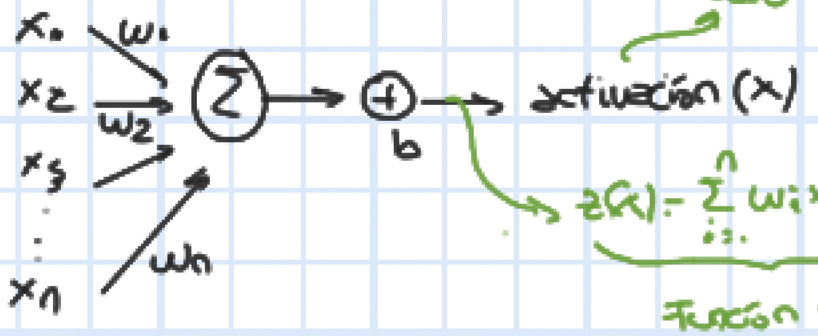
FF + ReLU



ReLu $\max(0, z(x))$

$z(x) = \sum_{i=1}^{n} w_i x_i + b = \bar{w}\cdot\bar{x} + b$

Función afín

Función de Costo (MSE)

$$\frac{1}{N}\sum_{x}(\text{salida}(x) - \text{activación}(x))^2 = \frac{1}{N}\sum_{x}\left(\overset{y}{\widetilde{\text{salida}}}(x) - \max\left(0, \sum_{i}^{n} w_i x_i + b\right)\right)^2$$

AM
$$\begin{cases} f(x,y) = 3x^2 y \\ \nabla f(x,y) = \left[\dfrac{\partial f(x,y)}{\partial x}, \dfrac{\partial f(x,y)}{\partial y}\right] \\ \nabla f(x,y) = [6yx, 3x^2] \end{cases}$$

- Gradiente escalar
- Gradiente de operaciones elementwise vectores } Gradiente producto escalar
- Gradientes suma
- Gradiente ReLu ⟶ piecewise (por partes)

Vectores Gradientes (funciones escalares)

$f, g \rightarrow x, y$ n

$$J = \begin{bmatrix} \nabla f(x,y) \\ \nabla g(x,y) \end{bmatrix} = \begin{bmatrix} \dfrac{\partial f(x,y)}{\partial x} & \dfrac{\partial f(x,y)}{\partial y} \\ \dfrac{\partial g(x,y)}{\partial x} & \dfrac{\partial g(x,y)}{\partial y} \end{bmatrix}$$

m $\quad$ $m \times n$

$n = |x| \longrightarrow$ cardinalidad

$n \times 1$

$$x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \quad y = f(x) \text{ escalar} \quad \begin{bmatrix} y_1 = f_1(x) \\ y_2 = f_2(x) \\ y_3 = f_3(x) \end{bmatrix} \text{ vector}$$

$$\frac{\partial y}{\partial x} = \begin{bmatrix} \nabla f_1(x) \\ \nabla f_2(x) \\ \vdots \\ \nabla f_m(x) \end{bmatrix} = \begin{bmatrix} \frac{\partial}{\partial x} f_1(x) \\ \frac{\partial}{\partial x} f_2(x) \\ \vdots \end{bmatrix} = \overbrace{\begin{bmatrix} \frac{\partial}{\partial x_1} f_1(x) & \frac{\partial}{\partial x_2} f_1(x) & \cdots \\ & \frac{\partial}{\partial x_2} f_m(x) & \frac{\partial}{\partial x_n} f_m(x) \end{bmatrix}}^{m \times n}$$

	escalar x	vector x
escalar f	$\frac{\partial f}{\partial x}$	$\frac{\partial f}{\partial x}$
vector f	$\frac{\partial f}{\partial x}$	$\frac{\partial f}{\partial x}$ ($x_1 \dots x_n$; $f_1 \dots f_m$)

Operaciones Element-wise

$\bar{w}$ $\bar{x}$ operación ew

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} f_1(w) \circ g_1(x) \\ f_2(w) \circ g_2(x) \\ \vdots \\ f_n(w) \circ g_n(w) \end{bmatrix}$$

n

$\bar{w}, \bar{x} \in \mathbb{R}^n$

$y \in \mathbb{R}^n$

$$\bar{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \quad \bar{w} = \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \end{bmatrix} \quad \bar{x} + \bar{w} = \begin{bmatrix} x_1 + w_1 \\ x_2 + w_2 \\ x_3 + w_3 \end{bmatrix}$$

$\frac{\partial}{\partial w_j} g_i(x) = 0$

$$J_w = \frac{\partial y}{\partial w} = \operatorname{diag}\left(\frac{\partial}{\partial w_1}(f_1(w) \circ g_1(x)), \frac{\partial}{\partial w}()\right)$$

$$J_w = \underbrace{\frac{\partial y}{\partial w}}_{n \times n} = \overbrace{\begin{bmatrix} \frac{\partial}{\partial w_1}(f_1(\bar{w}) \circ g_1(\bar{x})) & \frac{\partial}{\partial w_2}(f_1(w) \circ g_1(x)) & \cdots & \frac{\partial}{\partial w_n}(f_1(\bar{w}) \circ g_1(\bar{x})) \\ \vdots & & & \\ \frac{\partial}{\partial w_1}(f_n(\bar{w}) \circ g_n(\bar{x})) & & \cdots & \frac{\partial}{\partial w_n}(f_n(\bar{w}) \circ g_n(\bar{x})) \end{bmatrix}}^{n \times n}$$

0

Expansión escalar

$\bar{x}$ vector, z escalar $\qquad \overbrace{\vec{1}\,z}$ $\qquad \frac{\partial z}{\partial x_1} = 0 \rightarrow$ simplifica la operación

$$y = \bar{x} + z \rightarrow y = f(x) + g(z)$$

$x+z$

$$\frac{\partial}{\partial x_1}\left(f_1(x) + g_1(z)\right) = \frac{\partial (x_1 + z)}{\partial x_1} = \overset{1}{\frac{\partial x_1}{\partial x_1}} + \overset{0}{\frac{\partial z}{\partial x_1}} = 1$$

$$\frac{\partial}{\partial x}(\bar{x} + z) = diag(1) = I$$

$$\frac{\partial}{\partial z}\left(f_1(x_1) + g_1(z)\right) = \frac{\partial (x_1 + z)}{\partial z} = 0 + 1 = 1$$

$$\frac{\partial}{\partial z}(\bar{x} + z) = \vec{1}$$

$$\frac{\partial}{\partial x}(\bar{x} z) = \overset{n \times n}{\widetilde{I z}}$$

$$\frac{\partial}{\partial z}(\bar{x} z) = \overset{n \times 1}{\widetilde{\bar{x}}}$$

Ley de las derivadas totales

$y = x + x^2 \qquad \frac{dy}{dx}$

$u_1(x) = x^2$

$u_2(x, u_1) = x + u_1$

$y = f(x) = u_2(x, u_1)$

$$\frac{dy}{dx} = \frac{\partial f(x)}{\partial x} = \frac{\partial u_2(x, u_1)}{\partial x} = \frac{\partial u_2}{\partial x}\overset{1}{\frac{\partial x}{\partial x}} + \frac{\partial u_2}{\partial u_1}\frac{du_1}{dx} = \overbrace{\frac{\partial u_2}{\partial x} + \frac{\partial u_2}{\partial u_1}\frac{du_1}{dx}}^{1 + 2x}$$

$$\frac{\partial f(x, u_1, u_2, \dots u_n)}{\partial x} = \frac{\partial f}{\partial x} + \sum_{i=1}^{n} \frac{\partial f}{\partial u_i}\frac{du_i}{dx}$$

Reducción vectorial por suma

$f_i \rightarrow x_i, x_2, x_3$

$f_i(x) = x_i$

$$y = sum(f(x)) = \sum_{i=1}^{n} f_i(x) \qquad y = sum(x)$$

$$\frac{\partial y}{\partial x} = \left[\frac{\partial y}{\partial x_1}, \frac{\partial y}{\partial x_2} \dots \frac{\partial y}{\partial x_n}\right] \qquad \frac{\partial}{\partial x_j} x_i = 0, \; j \neq i$$

$$= \left[\sum_i \frac{\partial f_i(x)}{\partial x_1} \dots \sum_i \frac{\partial f_i(x)}{\partial x_n}\right]$$

$$\nabla y = \left[\frac{\partial x_1}{\partial x_1}, \frac{\partial x_2}{\partial x_2} \dots \frac{\partial x_n}{\partial x_n}\right] = \vec{1}^T$$

$f_i(\bar{x}, z) = x_i z$

$$\boxed{\begin{aligned} \nabla_x f &= [z, z, \dots z] = \vec{z}^T \\ \nabla_z f &= \frac{\partial}{\partial z} \sum_{i=1}^{n} x_i z = sum(x) \end{aligned}}$$

Regla de la Cadena

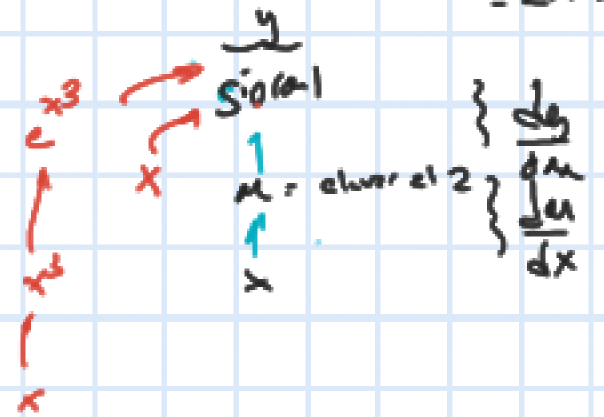
- Regla de la cadena Univariada
- Regla de la cadena por derivada total
- Regla de la cadena vectorial

Univariada $\qquad y = f(x)$

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

$$y = f(g(x)) = \sin(x^2)$$

1. $u = x^2$, $y = \sin(u)$
2. $\frac{du}{dx} = 2x$, $\frac{dy}{du} = \cos(u)$
3. $\frac{dy}{dx} = \frac{dy}{du}\frac{du}{dx} = \cos(u) 2x$
4. Sustitución: $\frac{dy}{dx} = \frac{dy}{du}\frac{du}{dx} = 2x\cos(x^2)$



Caso Multivariable

$\bar{y}, \bar{x}$

$$y = \begin{vmatrix} y_1(x) \\ y_2(x) \end{vmatrix} = \begin{bmatrix} \ln(x^2) \\ \sin(3x) \end{bmatrix}$$

$$g = \begin{vmatrix} g_1(x) \\ g_2(x) \end{vmatrix} = \begin{vmatrix} x^2 \\ 3x \end{vmatrix}$$

$$\begin{vmatrix} f_1(g) \\ f_2(g) \end{vmatrix} = \begin{vmatrix} \ln(g_1) \\ \sin(g_2) \end{vmatrix}$$

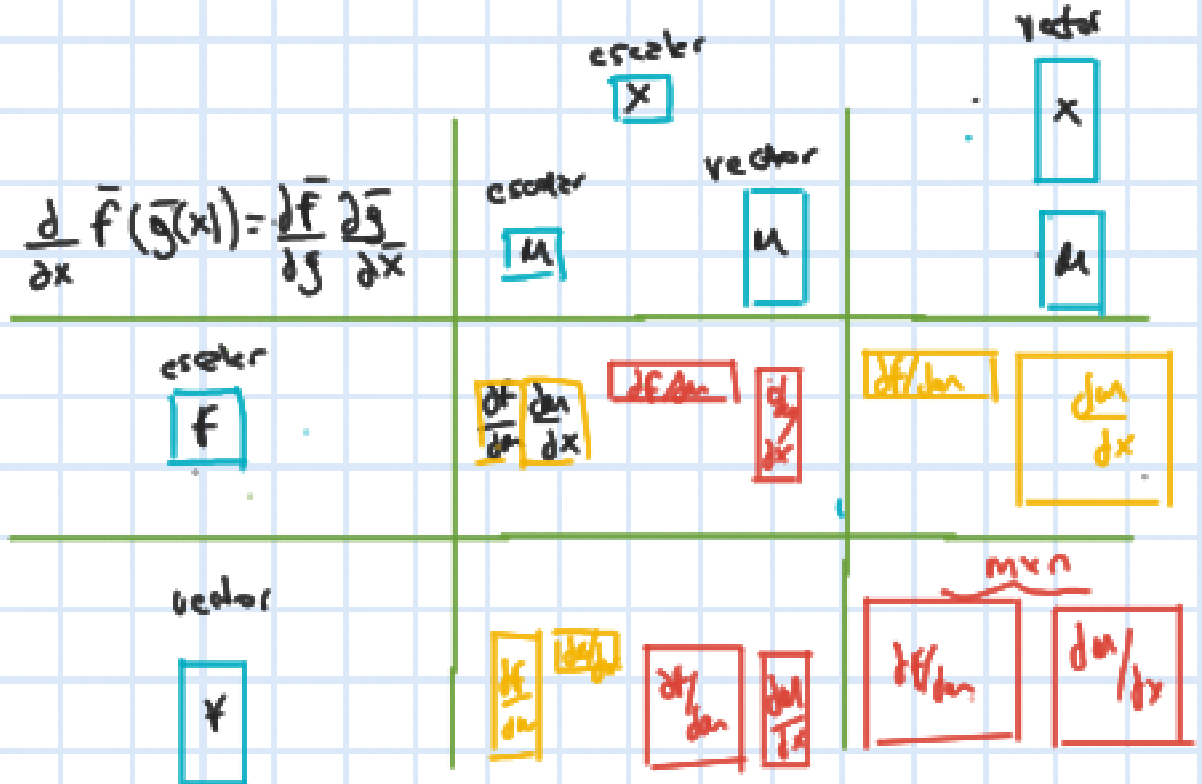
$$\frac{\partial y}{\partial x} = \begin{vmatrix} \frac{\partial f_1(g)}{\partial x} \\ \frac{\partial f_2(g)}{\partial x} \end{vmatrix} = \begin{vmatrix} \frac{\partial f_1}{\partial g_1}\frac{\partial g_1}{\partial x} + \overbrace{\frac{\partial f_1}{\partial g_2}}^{0}\frac{\partial g_2}{\partial x} \\ \underbrace{\frac{\partial f_2}{\partial g_1}}_{0}\frac{\partial g_1}{\partial x} + \frac{\partial f_2}{\partial g_2}\frac{\partial g_2}{\partial x} \end{vmatrix} = \begin{vmatrix} \frac{1}{g_1} 2x + 0 \\ 0 + \cos(g_2)3 \end{vmatrix} = \begin{vmatrix} 2/x \\ 3\cos(3x) \end{vmatrix}$$

$$\underbrace{\begin{vmatrix} \frac{\partial f_1}{\partial g_1} & \frac{\partial f_1}{\partial g_2} \\ \frac{\partial f_2}{\partial g_1} & \frac{\partial f_2}{\partial g_2} \end{vmatrix}}_{m \times k} \underbrace{\begin{vmatrix} \frac{\partial g_1}{\partial x} \\ \frac{\partial g_2}{\partial x} \end{vmatrix}}_{k \times 1} = \underbrace{\frac{\partial \bar{f}}{\partial \bar{g}}\frac{\partial \bar{g}}{\partial x}}_{m \times 1}$$

$$\frac{\partial}{\partial x}\bar{f}(\bar{g}(\bar{x})) = \frac{\partial \bar{F}}{\partial \bar{g}} \cdot \frac{\partial \bar{g}}{\partial \bar{x}} \qquad |g| = k$$

$$\underbrace{\frac{\partial}{\partial x}\bar{f}(\bar{g}(\bar{x}))}_{\substack{|x| = n \\ m \times n}} = \underbrace{\begin{bmatrix} \frac{\partial f_1}{\partial g_1} & \cdots & \frac{\partial f_1}{\partial g_k} \\ \vdots & & \\ \frac{\partial f_m}{\partial g_1} & \cdots & \frac{\partial f_m}{\partial g_k} \end{bmatrix}}_{m \times k} \underbrace{\begin{bmatrix} \frac{\partial g_1}{\partial x_1} & \cdots & \frac{\partial g_1}{\partial x_n} \\ & & \\ \frac{\partial g_k}{\partial x_1} & & \frac{\partial g_k}{\partial x_n} \end{bmatrix}}_{k \times n}$$

$m = n$



1 Neurona

$\bar{w}, \bar{x}$

$\bar{w}\cdot\bar{x} = w_1x_1 + w_2x_2 + w_3x_3 \dots$

$\sum_i^n (w_i x_i) = sum(w \odot x)$

$u = w \odot x$

$y = sum(u)$

$\frac{\partial y}{\partial \bar{w}} = \frac{\partial y}{\partial \bar{u}} \cdot \frac{\partial \bar{u}}{\partial \bar{w}} = diag(x) \cdot \vec{1}^T = x^T$

$\frac{\partial u}{\partial w} = diag(x)$

$\frac{\partial y}{\partial u} = \vec{1}^T$

$y = wx + b$

$\frac{\partial y}{\partial w} = \frac{\partial}{\partial w} w\cdot x + \frac{\partial}{\partial w} b = x^T + \vec{0}^T = x^T$

$\frac{\partial y}{\partial b} = \frac{\partial}{\partial b} w\cdot x + \frac{\partial}{\partial b} b = 0 + 1 = 1$

$max(0, \underbrace{\bar{w}\cdot\bar{x}+b}_{z})$

$\frac{\partial}{\partial z} max(0,z) = \begin{cases} 0 & z \le 0 \\ \frac{dz}{dz} = 1 & z > 0 \end{cases}$

$max(0,\bar{x}) = \begin{bmatrix} max(0, x_1) \\ max(0, x_2) \\ max(0, x_n) \end{bmatrix}$ $\frac{\partial}{\partial \bar{x}} max(0,\bar{x}) = \begin{bmatrix} \frac{\partial}{\partial x_1} max(0, x_1) \\ \dots \\ \frac{\partial}{\partial x_n} max(0, x_n) \end{bmatrix}$

$\frac{\partial\, activación}{\partial w} = \frac{\partial\, activación}{\partial z} \cdot \frac{\partial z}{\partial w}$

$\frac{\partial\, activación}{\partial \bar{w}} = \begin{cases} \vec{0}^T & w\cdot x + b \le 0 \\ \bar{x}^T & w\cdot x + b > 0 \end{cases}$

$\frac{\partial\, activación}{\partial b} = \begin{cases} 0 \cdot \frac{\partial z}{\partial b} = 0 \\ 1 \cdot \frac{\partial z}{\partial b} = 1 \end{cases}$

$\frac{\partial C}{\partial b} = \frac{2}{N} \sum_{i=1}^{N} e_i$

Función de costo

$C(w,b,X,y) = \frac{1}{N} \sum_{i=1}^{N} (y_i - \overbrace{max(0, w\cdot x_i + b)}^{\text{cantidad actuales}})^2$

$u(w,b,x) = max(0, w\cdot x + b)$

$v(y,u) = y - u$

$c(v) = \frac{1}{N} \sum_{i=1}^{N} v^2$

$\frac{\partial}{\partial w} u(w,b,x) = \begin{cases} \vec{0}^T & wx+b \le 0 \\ x^T & wx+b > 0 \end{cases}$

$\frac{\partial v(y,u)}{\partial w} = \frac{\partial}{\partial w}(y - u) = 0^T - \frac{\partial u}{\partial w} = \begin{cases} \vec{0}^T \\ -\bar{x}^T \end{cases}$

$\frac{\partial C(v)}{\partial w} = \frac{1}{N} \sum_{i=1}^{N} \frac{\partial}{\partial w} v^2$

$= \frac{1}{N} \sum_{i=1}^{N} \frac{\partial v^2}{\partial v} \frac{\partial v}{\partial w} = \frac{1}{N} \sum_{i=1}^{N} 2v \frac{\partial v}{\partial w}$

$= \frac{1}{N} \sum_{i=1}^{N} \begin{cases} 2v\vec{0}^T & wx_i + b \le 0 \\ -2v x_i^T & wx_i + b > 0 \end{cases}$

$= \frac{1}{N} \sum_{i=1}^{N} \begin{cases} 2(y_i - u) 0^T \\ -2(y_i - u) x_i^T \end{cases}$

$= \begin{cases} \vec{0}^T & w\cdot x_i + b \le 0 \\ \frac{2}{N} \sum_{i=1}^{N} (\underbrace{w\cdot x_i + b - y_i}_{e}) x_i^T & w\cdot x_i + b > 0 \end{cases}$

$\frac{\partial C}{\partial w} = \frac{2}{N} \sum_{i=1}^{N} e_i x_i^T$

$N=1 \quad 2e_i x_i^T$

$e_i = 0$

$w_{t+1} = w_t \ominus \eta \frac{\partial C}{\partial w}$