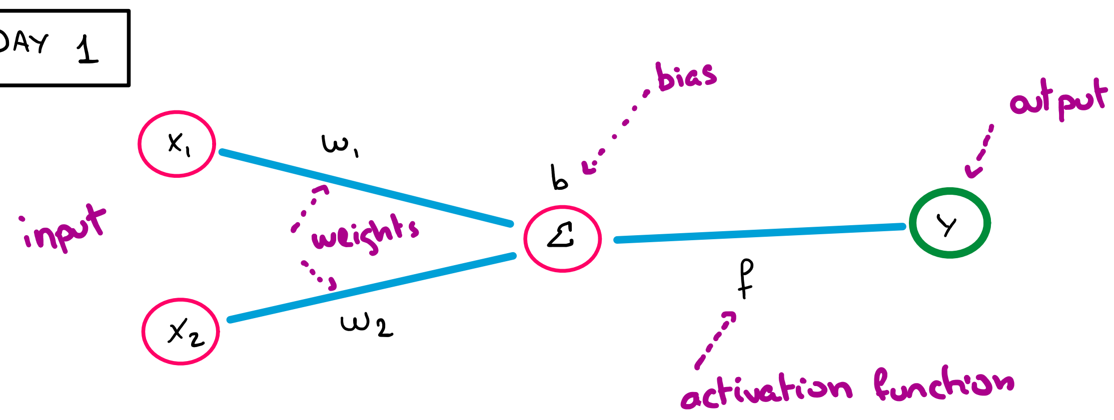
Easy Nets (Day 2)

dimecres, 25 de novembre de 2020 4:16 p. m.

DAY 1



Example:

$$f(x) = \frac{1}{1+e^{-x}}, \quad \frac{\partial f}{\partial x} = \frac{e^{-x}}{(1+e^{-x})^2} = \frac{1-f(x)}{f(x)}$$

then

$$\text{input } (x_1, x_2) \xrightarrow{\text{weights \& bias}} \underbrace{x_1 \cdot w_1 + x_2 w_2 + b}_{q} \xrightarrow{\text{activation}} \frac{1}{1+e^{-(x_1 \cdot w_1 + x_2 w_2 + b)}} = f(x_1, w_1, x_2, w_2, b)$$

- to compute partial derivatives, use chain rule:
- suppose:

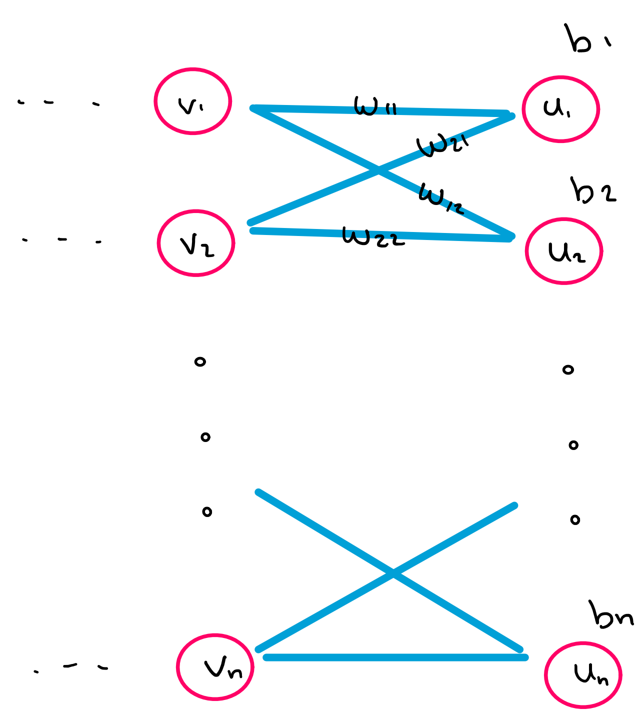
$$\begin{cases} x_0 = -1 \\ x_1 = -2 \\ w_0 = 4 \\ w_1 = -4 \\ b = -3 \end{cases}, \quad \text{then we have} \quad \begin{array}{l} q = x_1 w_1 + x_2 w_2 + b \\ q = (-2)(-4) + (-1) \cdot 4 - 3 = 1 \end{array}$$

- $$\frac{\partial f}{\partial x_1} = \frac{\partial f}{\partial q} \cdot \frac{\partial q}{\partial x_1} = \frac{1-f(q)}{f(q)} \cdot w_1 \Bigg|_{\substack{q=1 \\ w_1=-4}} = \frac{1-f(1)}{f(1)} \cdot (-4) = -4/e \simeq -1 \ldots$$

- $$\frac{\partial f}{\partial b} = \frac{\partial f}{\partial q} \cdot \frac{\partial q}{\partial b} = \frac{1-f(q)}{f(q)} \cdot 1 \Bigg|_{q=1} = \frac{1}{e}$$ (and so on ...)

$$\frac{\partial f}{\partial w_0} = \frac{\partial f}{\partial q} \cdot \frac{\partial q}{\partial w_0} = \frac{1-f(q)}{f(q)} \, x_0 \Bigg|_{\substack{x_0=-1 \\ q=1}} = -1/e$$

Linear Layer: $\mathcal{L}$



$u_1 = v_1 \cdot w_{11} + v_2 w_{21} + \cdots v_n w_{n1} + b_1$

$u_2 = v_1 \cdot w_{12} + v_2 w_{22} + \cdots + v_n w_{n2} + b_2$

$u_n = v_1 w_{1n} + v_2 w_{2n} + \cdots + v_n w_{nn} + b_n$

$\longleftrightarrow$

forward

$$\underbrace{\begin{bmatrix} w_{11} & w_{21} & \cdots & w_{n1} \\ w_{12} & & & \\ \vdots & & & \\ w_{1n} & \cdots & & w_{nn} \end{bmatrix}}_{W} \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix} + \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix} = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix}$$

$$\boxed{W \cdot x_t + b = x_{t+1}}$$

backward

$$\boxed{\frac{\partial \mathcal{L}}{\partial v_j} = \sum \frac{\partial \mathcal{L}}{\partial u_i} \cdot \frac{\partial u_i}{\partial v} = \sum \frac{\partial \mathcal{L}}{\partial u_i} \cdot w_{ji}}$$

chain rule

$$\begin{bmatrix} w_{11} & w_{12} & \cdots & w_{1n} \\ w_{21} & & & \\ \vdots & & & \\ w_{n1} & \cdots & & w_{nn} \end{bmatrix} \begin{bmatrix} \partial \mathcal{L}/\partial u_1 \\ \partial \mathcal{L}/\partial u_2 \\ \vdots \\ \partial \mathcal{L}/\partial u_n \end{bmatrix} = \begin{bmatrix} \partial \mathcal{L}/\partial v_1 \\ \partial \mathcal{L}/\partial v_2 \\ \vdots \\ \partial \mathcal{L}/\partial v_n \end{bmatrix}$$

$$\boxed{W^T \delta_{t+1} = \delta_t}$$