Intro a las redes neuronales.

Resumen:

Supervisado | no supervisado.

Supervisado → regresión, clasificación

(x_n, t_n)

$t_n \in \mathbb{R}$ (regresión), $t_n \in \{0, 1, \dots, M\}$ (clasificación)

1. Definir un modelo $y(x; \vec{w})$.
2. Definir una función de error $\mathcal{E}(\vec{w})$
3. Minimizar $\mathcal{E}(\vec{w})$ para encontrar los mejores pesos $\vec{w}$.

Ej: regresión Polinomial

1. $y(x;\vec{w}) = w_0 + w_1 x + \dots + w_M x^M$
2. $\mathcal{E}(\vec{w}) = \frac{1}{2} \sum \left(y(x_n;\vec{w}) - t_n \right)^2$
3. minimizar $\mathcal{E}$.

Ej: regresión logística

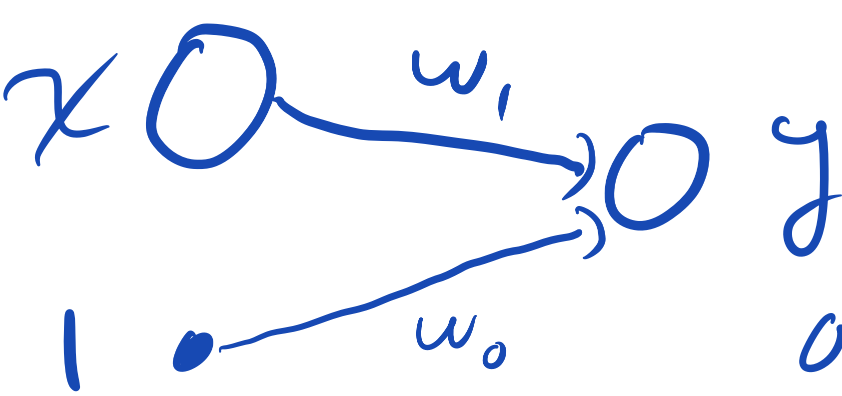
1. $y(x; w_0, w_1) = \dfrac{1}{\underbrace{1 + \exp(-(w_0 + w_1 x))}_{\text{Sigmoidal.}}}$
2. $\mathcal{E}(w_0, w_1) \underbrace{= \cdots}_{\text{entropía binaria cruzada.}}$

binary cross-entropy.

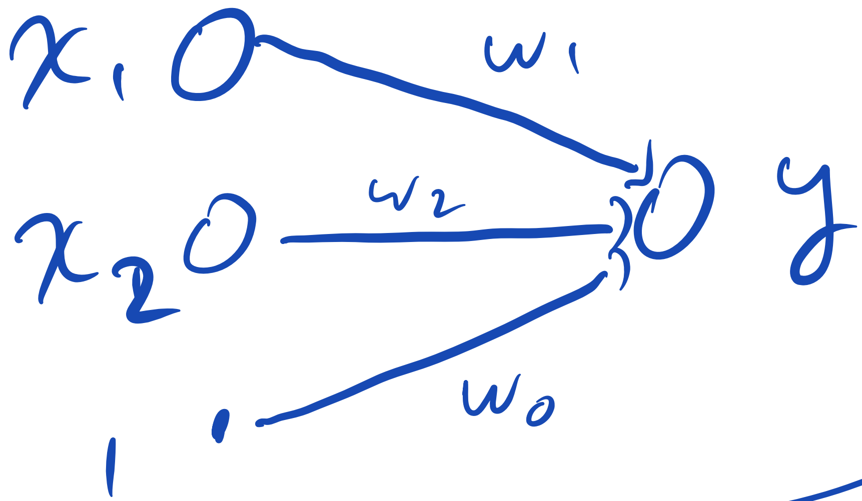
3. $\underbrace{\text{minimizar } E(\vec{w})?}_{\text{no lo podemos hacer analíticamente, entonces descendemos por el gradiente.}}$

¡Pensemos más general!

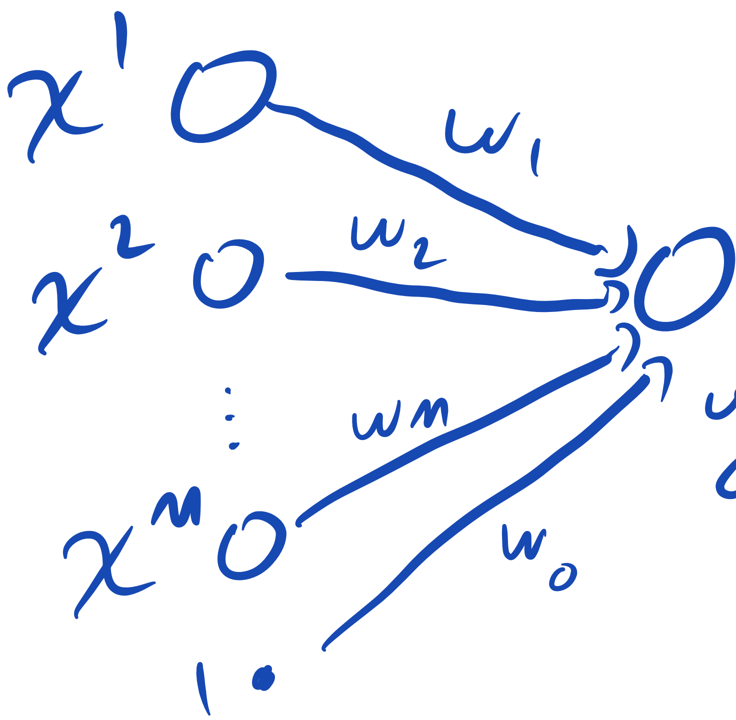
$$y(x;\vec{w}) = \underbrace{w_0 + w_1 x}_{\text{regresión lineal.}}$$



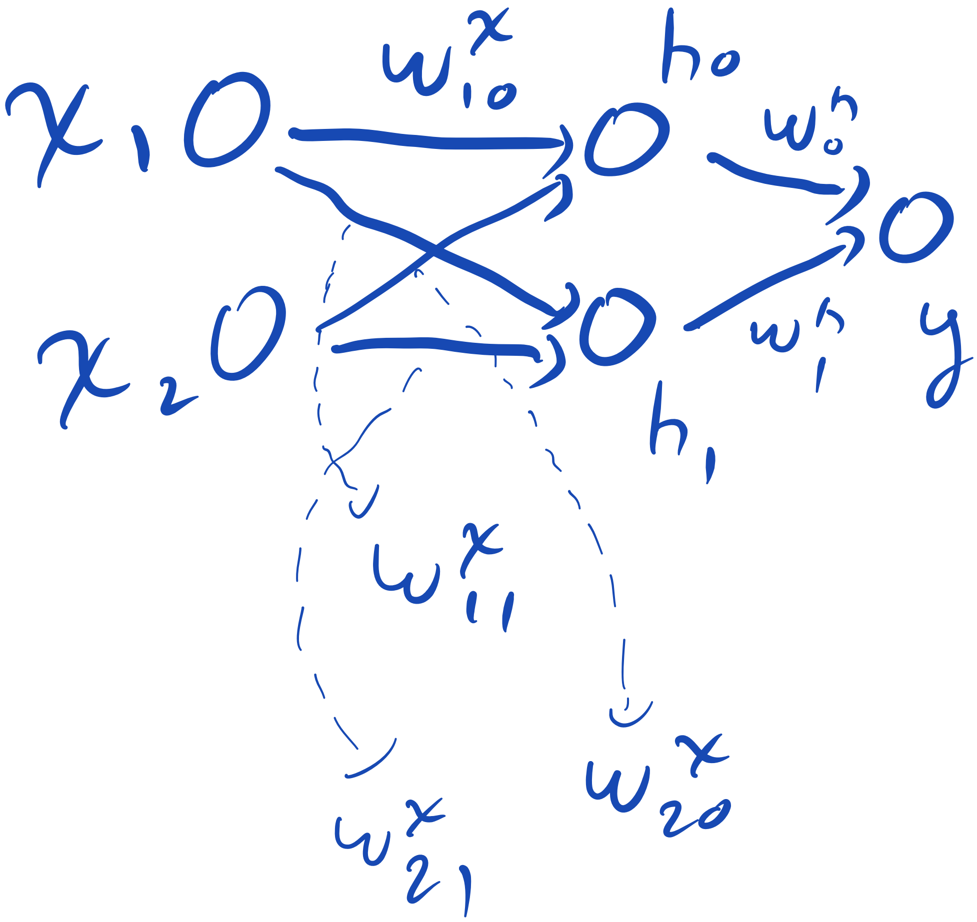
grafo.



$$y = x_1 w_1 + x_2 w_2 + w_0$$



$$y = w_0 + w_1 x + \dots + w_n x^n$$



$$h_0 = x_1 w^{x}_{10} + x_2 w^{x}_{20}$$

$$h_1 = x_1 w^{x}_{11} + x_2 w^{x}_{21}$$

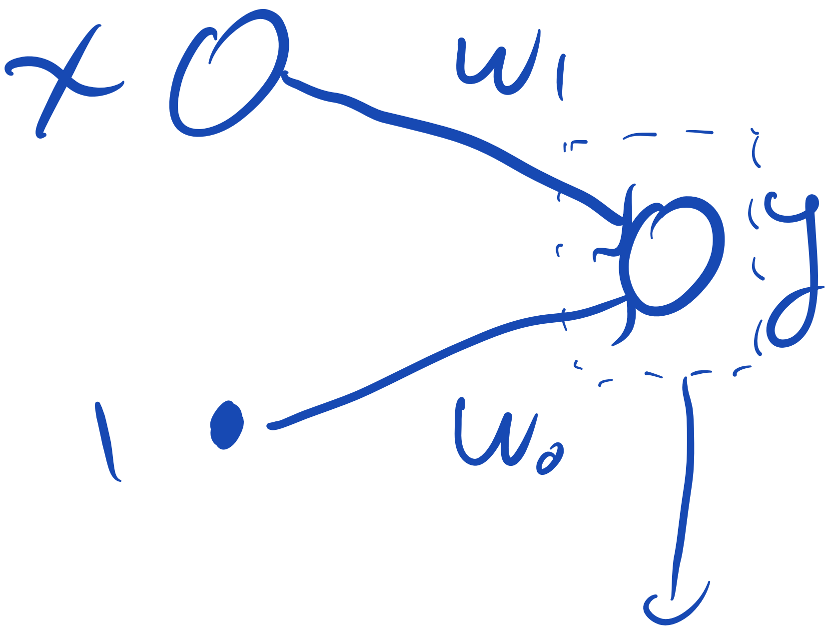
$$y = h_0 w^{h}_{0} + h_1 w^{h}_{1}$$

$$\vec{h} = \begin{bmatrix} h_0 \\ h_1 \end{bmatrix}, \quad W^x = \begin{bmatrix} w^x_{10} & w^x_{20} \\ w^x_{11} & w^x_{21} \end{bmatrix}$$

$$\vec{h} = {W^x}^T \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

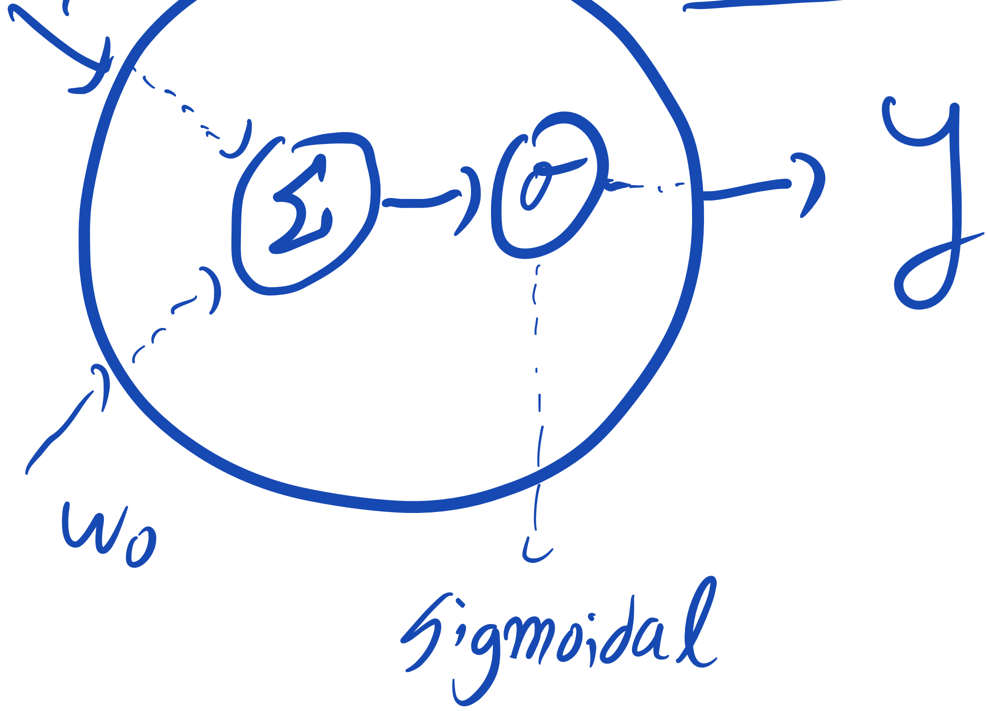
$$y = \begin{bmatrix} w^h_0 \\ w^h_1 \end{bmatrix}^T \vec{h}$$

- Podemos representar una regresión lineal/polin. como un grafo donde los pesos son las aristas y la información fluye de derecha a izquierda.



$$y = \frac{1}{1+\exp(-w_0 - w_1 x)}$$

Neuronas



Def: Una neurona es una función que toma la siguiente forma:

$$h(\vec{x};\vec{w}) = f(\vec{w}^T\vec{x})$$

Ej: logística

$$\sigma(x;w_0,w_1) = \frac{1}{1+\exp(-w_0-w_1x)}$$

Ej: regresión lineal

$$f(x;w_0,w_1) = w_0+w_1x$$

Notación: a f la llamamos función de activación.

- en reg. lineal, la función de activación es la identidad

- en reg. logística, la fun. de activación es la sigmoidal.

Ej: tanh, ReLU.

¿Cómo computamos el gradiente de una neurona respecto a los pesos?

$$h(t(\vec{w}))$$

$$\nabla_{\vec{w}} h(\vec{w}^T \vec{x})$$

$$\frac{\partial}{\partial w_j} h(w_1 x_1 + \dots + w_n x_n)$$

$$x_j \cdot \frac{\partial h}{\partial (w_1 x_1 + \dots + w_n x_n)}$$