

Auto diferenciación

una forma automática de computar derivadas.

Plan para hoy:

1. ¿Por qué necesitamos derivadas?

PyTorch { 2. ¿Cómo podemos computar el gradiente de funciones más complicadas de forma automática?

3. Gratos computacionales, regla de la cadena.

4. Entropía binaria cruzada como un grato computacional.

——— / / / ———

1. ¿Por qué necesitamos derivadas?

↳ minimizar el error $\mathcal{E}(w)$

↳ computamos $\nabla_w \mathcal{E}(w)$

↳ si podemos, $\nabla_w \mathcal{E}(w) = 0$

↳ Si no podemos,
descendemos por el
gradiente.

\mathcal{E}_j : Entropía Cruzada binaria.

Reg. logística

$$f(x; w_0, w_1) = \text{Sigmoidal}(w_0 + w_1 x) \\ = 1 / (1 + \exp(-w_0 - w_1 x))$$

$$\mathcal{L}(w_0, w_1; \mathcal{D}) =$$

$$- \frac{1}{N} \sum y_i \log(f(x_i)) + (1 - y_i) \log(1 - f(x_i))$$

$$\nabla_{\vec{w}} \mathcal{L} = \left(\frac{\partial \mathcal{L}}{\partial w_0}, \frac{\partial \mathcal{L}}{\partial w_1} \right)$$

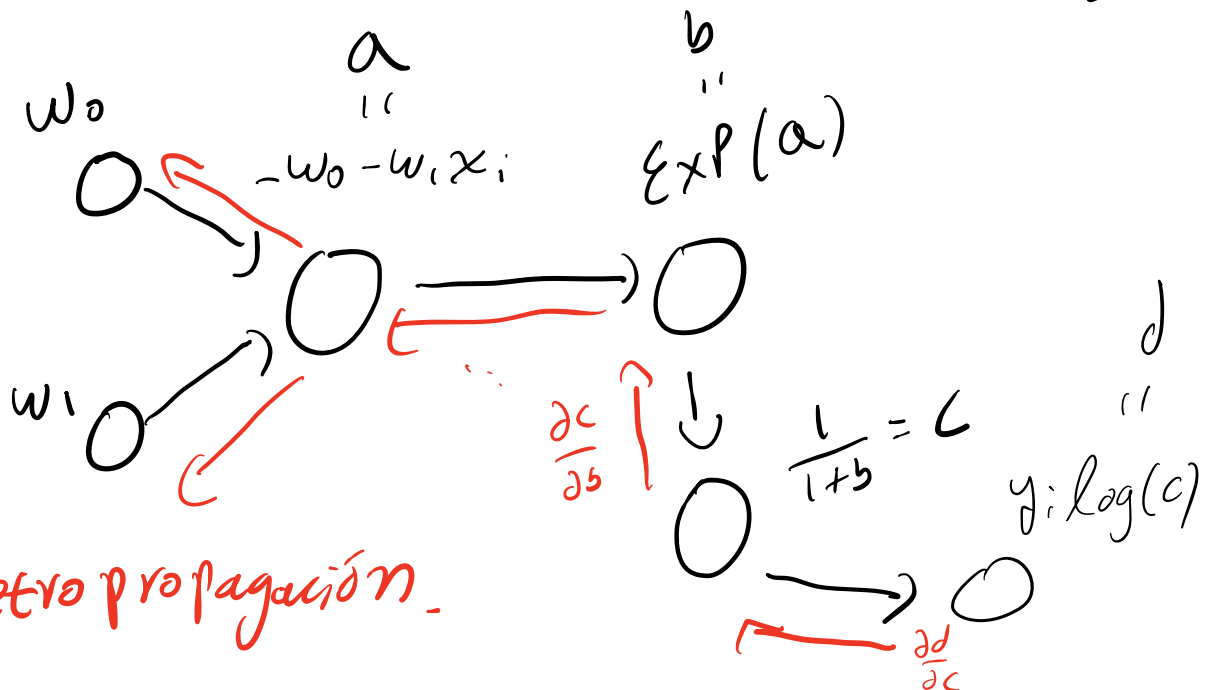
2. ¿Cómo podemos computar el gradiente de funciones más complicadas de forma automática?

↳ Representamos la función como un grafo.

dirigidos no dirigidos.

$$\mathcal{L}_j: g(w_0, w_1) = y_i \log(f(x_i; w_0, w_1))$$

$$= y_i \log\left(\frac{1}{1 + \exp(-w_0 - w_1 x_i)}\right)$$



Retro propagación.

$$\frac{\partial d}{\partial w_0} = \frac{\partial d}{\partial c} \cdot \frac{\partial c}{\partial b} \cdot \frac{\partial b}{\partial a} \cdot \frac{\partial a}{\partial w_0}$$

$$(y_i - \frac{1}{c})$$

$$c(b) = \frac{1}{1+b}$$

$$= (b+1)^{-1}$$

$$\frac{\partial c}{\partial b} = -(b+1)^{-2}$$

$$b(a) = \exp(a)$$

$$\frac{\partial b}{\partial a} = \exp(a)$$

$$a(w_0) = -w_0 - w_1 x_i$$

$$\frac{\partial a}{\partial w_0} = -1$$

$$\begin{aligned}\frac{\partial d}{\partial w_0} &= y_i \frac{1}{c} \cdot (- (b+1)^{-2}) \cdot \exp(a) \cdot -1 \\ &= y_i \frac{1}{c} (b+1)^{-2} \exp(a)\end{aligned}$$

— / / / —

Otros ejemplos de grafos
computacionales.

$$\begin{array}{l} w_0 \circ \\ w_1 \circ \\ w_2 \circ \end{array} \begin{array}{c} \diagup \\ \diagup \\ \diagdown \end{array} \circ \quad w_0 + 3w_1^2 + 5w_2^3$$

/ /

Modelado del lenguaje.

Nathalia está estudiando matemá-

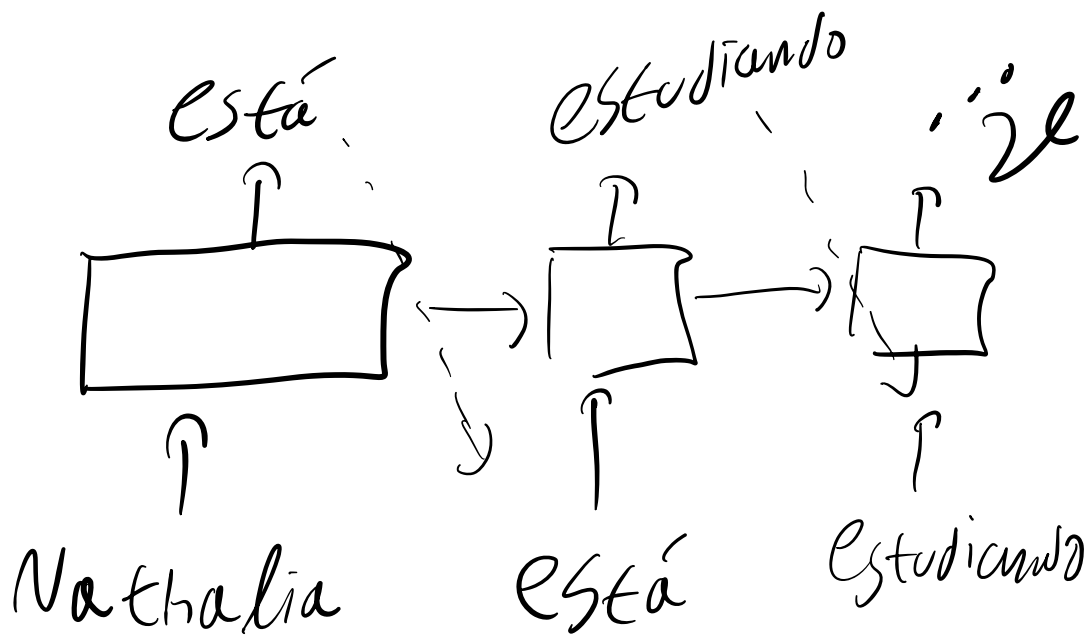
1. ¿Cómo representamos lenguaje de forma numérica?

| | | | |
|------------|---|-----|--|
| Nathalia | → | 1 # | } Vocabulario o diccionario. 300.000M |
| está | → | 2 # | |
| estudiando | → | 3 # | |
| ⋮ | | ⋮ | |

Nathalia → $[1 \ 0 \ \dots \ 0]$

está → $[0 \ 1 \ 0 \ \dots \ 0]$

estudiando → $[0 \ 0 \ 1 \ 0 \ \dots \ 0]$



$$\mathcal{V} = [0.1 \quad 0.1 \quad 0.75 \quad 0 \dots 0]$$

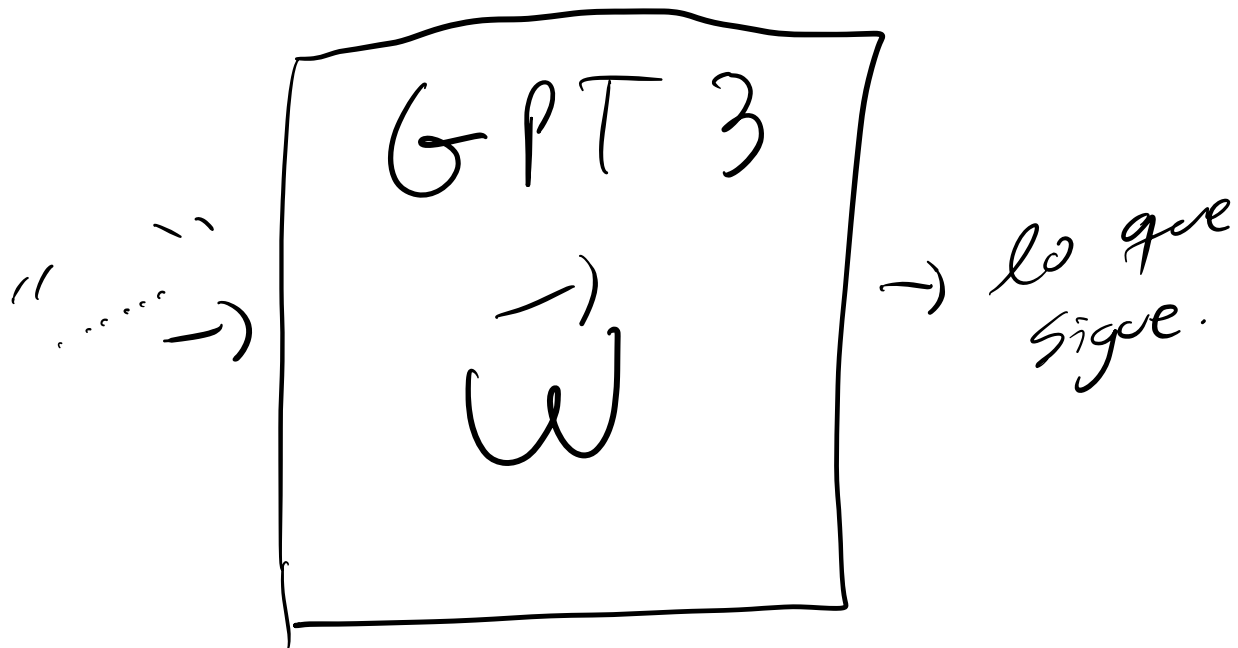
long. del lenguaje

Vec. de probabilidad.

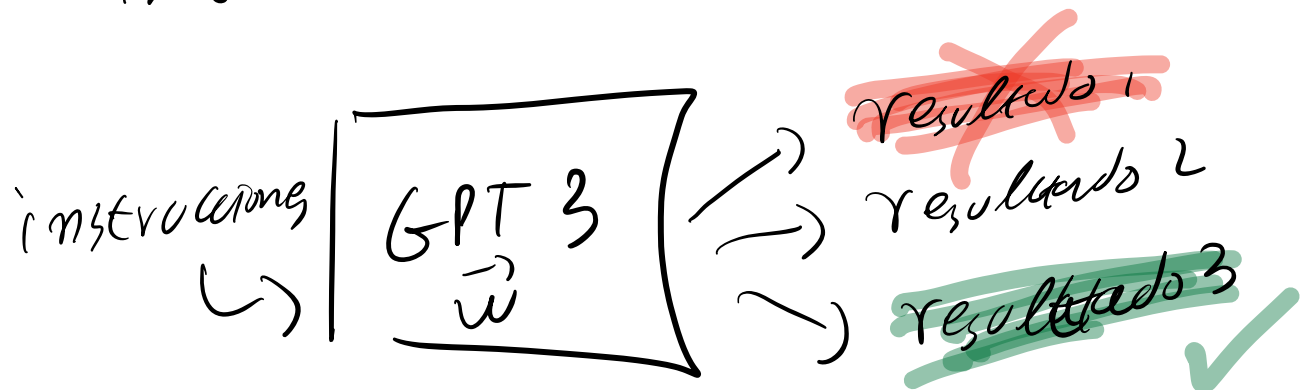
$$\sum \mathcal{V}_i = 1$$



Después del preentrenamiento



"Chat"



Usar interacción con humanos para
"mejorar" \vec{w} (Reinforcement learning
through human feedback)

$\mathbb{Z} \rightarrow \mathbb{R}$