Fundamentos de Aprendizaje Automático y Reconocimiento de Patrones

Actividades en clase: Práctico 5

Graciana Castro, Martín Schmidt, Federico Lecumberry, Guillermo Carbajal

Instituto de Ingeniería Eléctrica

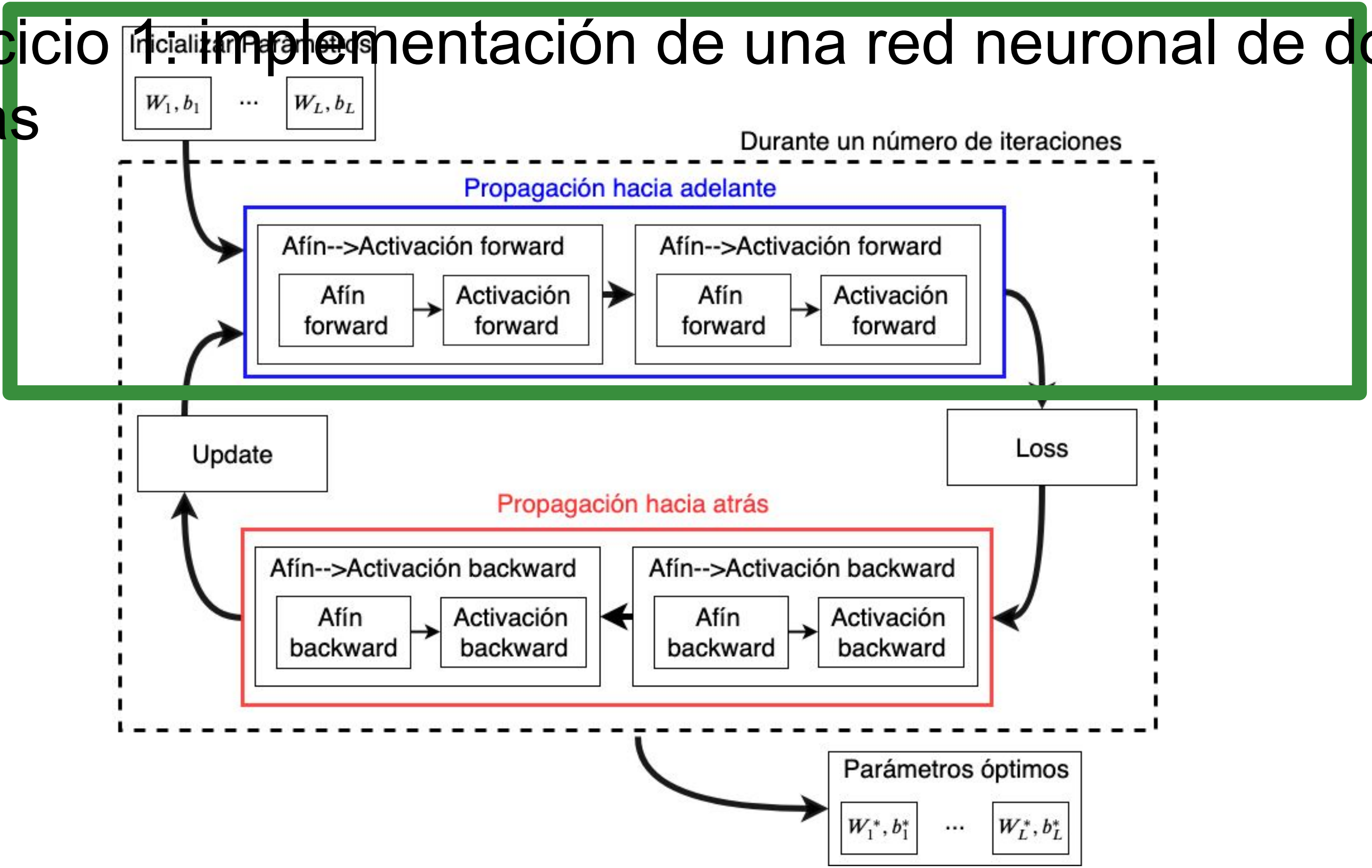
2024

Tabla de contenido

- Ejercicio 1: implementación de una red neuronal de dos capas
  - Semana pasada: inicialización de pesos y feed forward
  - **Para hoy: funciones de costo y backpropagation**

- Ejercicio 2: Jugando con Tensorflow playground

Ejercicio 1: implementación de una red neuronal de dos capas



Funciones de costo

Entropía cruzada: **problemas de clasificación binaria**

$$H(\mathbf{x}^{(\mathbf{L})}, \mathbf{y}) = -\frac{1}{N}\sum_{n=1}^{N}\left(y_n \log x_n^{(L)} + (1 - y_n)\log\left(1 - x_n^{(L)}\right)\right)$$

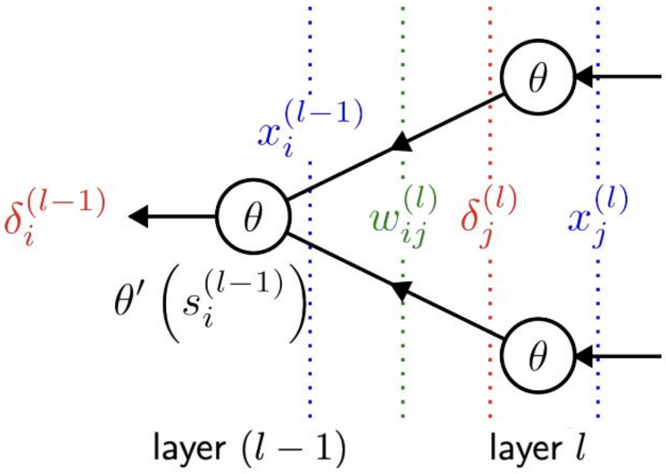
MSE: **problemas de**

$$\mathrm{MSE}(\mathbf{x}^{(L)}, \mathbf{y}) = \frac{1}{2N}\sum_{n=1}^{N}\left(y_n - x_n^{(L)}\right)^2$$

Backpropagation

Objetivo: minimizar el error respecto a los parámetros.

¿Minimizar? **DERIVADA.**



$$\frac{\partial \mathrm{e}(\mathbf{w})}{\partial w_{ij}^{(l)}} = \frac{\partial \mathrm{e}(\mathbf{w})}{\partial s_j^{(l)}} \underbrace{\frac{\partial s_j^{(l)}}{\partial w_{ij}^{(l)}}}_{x_i^{(l-1)}}$$

?

$$\frac{\partial \mathrm{e}(\mathbf{w})}{\partial s_j^{(l)}} = \delta_j^{(l)}$$

En el ejercicio: **cache** del forward

Cálculo del delta

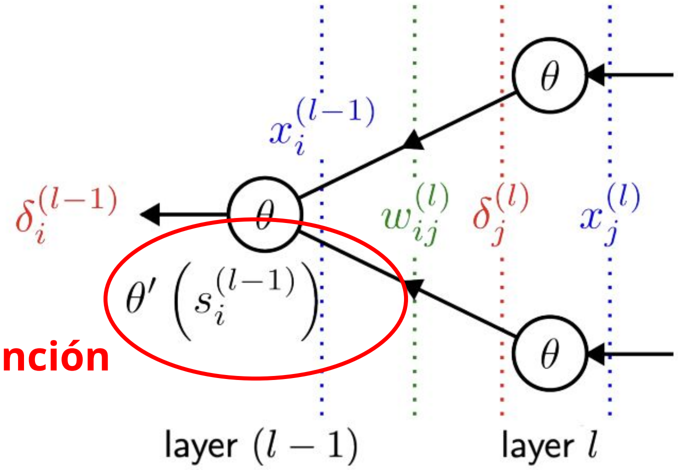
Uno de ellos es fácilmente calculable:

$$
\begin{aligned}
\boxed{\delta_1{}^{(L)} = \frac{\partial \mathbf{e}_n(\mathbf{w})}{\partial s_1{}^{(L)}}} \\
&= \frac{\partial}{\partial s_1{}^{(L)}} \left(x_1{}^{(L)} - y_n \right)^2 \\
&= 2 \left(x_1{}^{(L)} - y_n \right) \frac{\partial x_1{}^{(L)}}{\partial s_1{}^{(L)}} \\
&= 2 \left(x_1{}^{(L)} - y_n \right) \theta' \left(s_1{}^{(L)} \right)
\end{aligned}
$$

Donde:

$$x_1^{(L)} = \theta\left(s_1^{(L)}\right)$$

y suponiendo MSE.

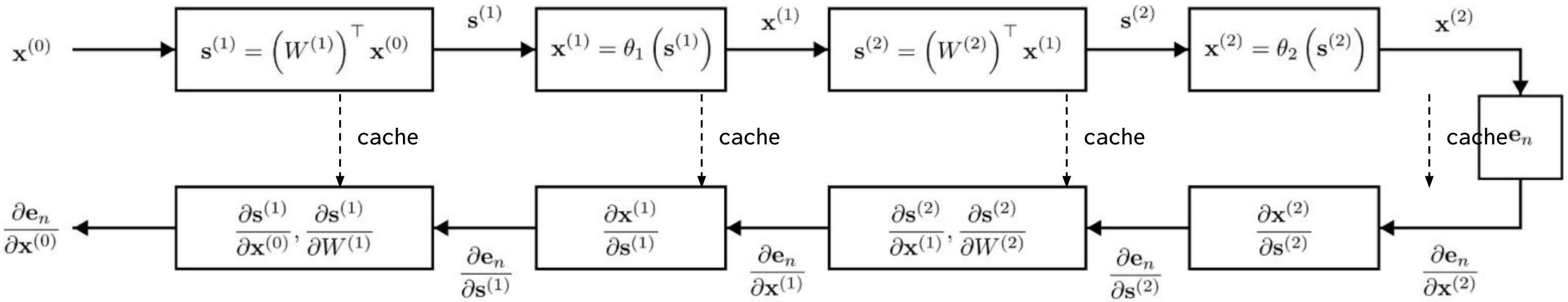


Derivada de la función de activación
de activación

Ejercicio 1: implementación de una red neuronal de dos

Forward $\longrightarrow$

$$\mathbf{x} = \mathbf{x}^{(0)} \xrightarrow{\mathrm{W}^{(1)}} \mathbf{s}^{(1)} \xrightarrow{\theta} \mathbf{x}^{(1)} \xrightarrow{\mathrm{W}^{(2)}} \mathbf{s}^{(2)} \xrightarrow{\theta} \mathbf{x}^{(2)} \cdots \longrightarrow \mathbf{s}^{(L)} \xrightarrow{\theta} \mathbf{x}^{(L)} = h(\mathbf{x})$$



$$\boldsymbol{\delta}^{(1)} \longleftarrow \boldsymbol{\delta}^{(2)} \cdots \longleftarrow \boldsymbol{\delta}^{(L-1)} \longleftarrow \boldsymbol{\delta}^{(L)}$$

$\longleftarrow$ Backward

Implementación:

Forward: Afín → Activación

Backward: Activación → Afín

$$\boxed{dS^{(l)} = dX^{(l)}\theta'(S^{(l)})}$$

$$\frac{\partial E}{\partial W^{(l)}} = dW^{(l)} = \left(X^{(l-1)}\right)^T dS^{(l)}$$

$$\frac{\partial E}{\partial \mathbf{b}^{(l)}} = d\mathbf{b}^{(l)} = 1^T dS^{(l)}$$

$$\frac{\partial E}{\partial X^{(l-1)}} = dX^{(l-1)} = dS^{(l)}\left(W^{(l)}\right)^T$$

Actualización de los parámetros

Utilizando **Descenso por gradiente:**

$$W^{(l)} = W^{(l)} - \eta\, dW^{(l)}$$

$$\mathbf{b}^{(l)} = \mathbf{b}^{(l)} - \eta\, \mathbf{db}^{(l)}$$

Ejercicio 1: repaso final completo