

## 7: Retropropagación

This page is a draft and is under active development.

### Idea intuitiva de la retropropagación

El algoritmo de retropropagación es un método muy utilizado en redes neuronales para minimizar la conocida como función de pérdida o de error de la red. Esta función proporciona, durante el entrenamiento de la red neuronal, una medida de cuánto se desvían los valores proporcionados por la red de los valores reales o deseados. Los parámetros desconocidos de la red, que son las incógnitas del problema y suelen denominarse pesos, se modifican de tal forma que los valores de salida de la red se ajusten lo más posible a los deseados. Para ello, se utiliza el método del gradiente descendente, que permite minimizar el error en función de los pesos. Puesto que el error se calcula al final de la red (en la capa de salida) es más fácil calcular el efecto de las variaciones de los pesos (y, por tanto, ajustarlos) comenzando desde el final, esto es, cambiando los pesos que llegan a la capa de salida. Sucesivamente, se modifican los de cada capa intermedia en dirección hacia la capa inicial. En este proceso será de utilidad la regla de la cadena.

#### ✓ Ejemplo 7.15

Vamos a considerar las funciones escalares  $f(x) = x^2$ ,  $g(y) = \sin y$  y  $h(z) = \exp(-z)$ . La composición de las tres funciones  $(h \circ g \circ f)(x)$  da lugar a  $J(x) = \exp(-\sin x^2)$ . Vamos a calcular  $\frac{dJ}{dx}$  utilizando el algoritmo de retropropagación. Aunque la composición de funciones se realiza hacia delante:

$$x \rightarrow y = f(x) = x^2 \rightarrow z = g(y) = \sin y \rightarrow J(x) = h(z) = \exp(-z) \quad (7.1)$$

el cálculo de la derivada lo realizamos hacia atrás, utilizando la regla de la cadena:

1. Empezamos por  $h(z) = \exp(-z)$ :

$$\frac{dJ}{dz} \rightarrow -\exp(-z) \quad (7.2)$$

2. Después por  $z = g(y) = \sin(y)$ :

$$\frac{dJ}{dy} \rightarrow \frac{dJ}{dz} \frac{dz}{dy}, \quad \frac{dz}{dy} = \cos(y) \quad (7.3)$$

3. Por último por  $y = f(x) = x^2$ :

$$\frac{dJ}{dx} \rightarrow \frac{dJ}{dy} \frac{dy}{dx}, \quad \frac{dy}{dx} = 2x \quad (7.4)$$

4. Así, resulta

$$\frac{dJ}{dx}(x) = -2x \cos(x^2) \exp(-\sin(x^2)) \quad (7.5)$$

#### ✓ Ejemplo 7.16

Sean  $\mathbf{x} = (x_1, x_2)$  y las funciones:  $f(x_1, x_2) = a_1 x_1 + a_2 x_2$  con  $a_1$  y  $a_2$  constantes,  $g(y) = 1 + e^{-y}$  y  $h(z) = \frac{1}{z}$ . La composición de las tres funciones  $(h \circ g \circ f)(x)$  da lugar a  $J(x_1, x_2) = \frac{1}{1 + e^{-(a_1 x_1 + a_2 x_2)}}$ . Esto es:

$$(x_1, x_2) \rightarrow y = f(x_1, x_2) = a_1 x_1 + a_2 x_2 \rightarrow z = g(y) = 1 + e^{-y} \rightarrow J(\mathbf{x}) = h(z) = \frac{1}{z} \quad (7.6)$$

El objetivo en este caso es el cálculo del gradiente de la función  $J$ ,  $\nabla J(\mathbf{x})$ , que tiene por coordenadas  $\frac{\partial}{\partial x_i} J(\mathbf{x})$ , para  $i = 1, 2$ .

1. Empezamos por  $h(z) = \frac{1}{z}$  :

$$\frac{\partial J}{\partial z} \rightarrow -\frac{1}{z^2} \quad (7.7)$$

2. Después tomamos  $z = g(y) = 1 + e^{-y}$  :

$$\frac{\partial J}{\partial y} \rightarrow \frac{\partial J}{\partial z} \frac{\partial z}{\partial y}, \quad \frac{\partial z}{\partial y} = -e^{-y} \quad (7.8)$$

3. Por último, para  $y = f(x_1, x_2) = a_1 x_1 + a_2 x_2$  :

- $\frac{\partial J}{\partial x_1} \rightarrow \frac{\partial J}{\partial y} \frac{\partial y}{\partial x_1}, \quad \frac{\partial y}{\partial x_1} = a_1$
- $\frac{\partial J}{\partial x_2} \rightarrow \frac{\partial J}{\partial y} \frac{\partial y}{\partial x_2}, \quad \frac{\partial y}{\partial x_2} = a_2$

This page titled [7: Retropropagación](#) is shared under a [not declared](#) license and was authored, remixed, and/or curated by [Joaquín López Herraiz](#).