Red neuronal 1

Jorge Butragueño Nieto

August 28, 2024

Abstract

Estudio de una red neuronal sencilla con dos nodos de entrada, dos nodos ocultos y dos nodos de salida. (ver Figura 1)

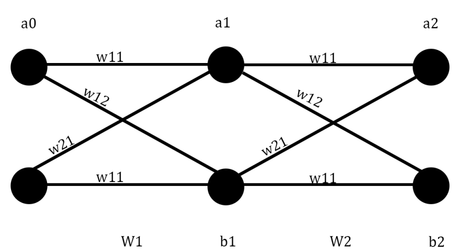

Figure 1: Red neuronal

1 Objetivos

El objetivo de este estudio será ajustar la red neuronal para que dados unos **input**, se ajusten todos los pesos y las bias para que devuelva unos **outputs** muy lo más parecidos a lo deseado y con el menor error posible.

2 Inputs y outputs

Lo primero será saber qué recibe la red y qué queremos que devuelva. En nuestro caso como recibe dos inputs, y devuelve 2 outputs. Construiremos una red que al recibir [1 0] devuelva el contrario es decir, [0 1].

$$a^0 = \begin{bmatrix} 1 & 0 \end{bmatrix}$$

$$y = \begin{bmatrix} 0 & 1 \end{bmatrix}$$

(donde a^0 es el input dado, y es el output esperado)

3 Forward propagation

Ahora inicializamos todos los datos que queremos calcular de forma aleatoria y calculamos cuál sería el resultado de la red.

3.1 Matrices

En el caso de este ejemplo tenemos $2 \times 2 + 2 \times 2 + 2 + 2 = 12$ valores que ajustar correctamente, y la mejor forma de organizar todos los cálculos es a través de las matrices.

$$W^1 = \begin{bmatrix} w_{11}^1 & w_{12}^1 \\ w_{21}^1 & w_{22}^1 \end{bmatrix} W^2 = \begin{bmatrix} w_{11}^2 & w_{12}^2 \\ w_{21}^2 & w_{22}^2 \end{bmatrix} b^1 = \begin{bmatrix} b_1 & b_2 \end{bmatrix} b^2 = \begin{bmatrix} b_1 & b_2 \end{bmatrix}$$

(ver la figura 1)

3.2 Calcular el valor de cada nodo

A partir de estos valores podemos calcular el valor de cada neurona hasta llegar al final.
Primero con los nodos ocultos:

$$z_1^1 = a_1^0 w_{11}^1 + a_2^0 w_{21}^1 + b_1^1 \qquad a_1^1 = \sigma(z_1^2)$$

$$z_2^1 = a_1^0 w_{12}^1 + a_2^0 w_{22}^1 + b_1^1 \qquad a_2^1 = \sigma(z_2^2)$$

Finalmente con el output:

$$z_1^2 = a_1^1 w_{11}^2 + a_2^1 w_{21}^2 + b_1^2 \qquad a_1^2 = \sigma(z_1^2)$$
$$z_2^2 = a_1^1 w_{12}^2 + a_2^1 w_{22}^2 + b_1^2 \qquad a_2^2 = \sigma(z_2^2)$$

Nótese que podemos reescribir las anteriores en términos matriciales:

$$Z^1 = a^0 W^1 + b^1$$

$$z^1 = \begin{bmatrix} z_1^1 & z_2^1 \end{bmatrix} = \begin{bmatrix} a_1^0 & a_2^0 \end{bmatrix} \begin{bmatrix} w_{11}^1 & w_{12}^1 \\ w_{21}^1 & w_{22}^1 \end{bmatrix} + \begin{bmatrix} b_1^1 & b_2^1 \end{bmatrix} = \begin{bmatrix} a_1^0 w_{11}^1 + a_2^0 w_{21}^1 + b_1^1 & a_1^0 w_{12}^1 + a_2^0 w_{22}^1 + b_1^1 \end{bmatrix}$$

$$Z^2 = a^1 W^2 + b^2$$

$$z^2 = \begin{bmatrix} z_1^2 & z_2^2 \end{bmatrix} = \begin{bmatrix} a_1^1 & a_2^1 \end{bmatrix} \begin{bmatrix} w_{11}^2 & w_{12}^2 \\ w_{21}^2 & w_{22}^2 \end{bmatrix} + \begin{bmatrix} b_1^2 & b_2^2 \end{bmatrix} = \begin{bmatrix} a_1^1 w_{11}^2 + a_2^1 w_{21}^2 + b_1^2 & a_1^1 w_{12}^2 + a_2^1 w_{22}^2 + b_1^2 \end{bmatrix}$$

y finalmente:

$$a^1 = \sigma(z^1) = \begin{bmatrix} \sigma(z_1^1) & \sigma(z_2^1) \end{bmatrix}$$

$$a^2 = \sigma(z^2) = \begin{bmatrix} \sigma(z_1^2) & \sigma(z_2^2) \end{bmatrix}$$

3.3 Error

Finalmente una medida para saber cómo de bien está el resultado dado es utilizando el error cuadrático medio MSE

$$MSE = \frac{1}{n}\sum_{0}^{n}(y_i - \hat{y}_i)$$

En nuestro caso $\hat{y} = a^2$ y $n = 2$ por lo que la función de Coste (C) quedaría:

$$C = \frac{1}{2}(y_1 - a_1^2)^2 + (y_2 - a_2^2)^2$$

4 Backpropagation

4.1 Cálculo del gradiente

Para poder reducir al máximo el error usaremos el descenso por gradiente, donde será necesario calcular el gradiente de nuestra función. El gradiente se obtiene como la derivada de cada una de nuestras variables, en nuestro caso son 12:

$$\frac{\partial C}{\partial W^2} = \begin{bmatrix} \frac{\partial C}{\partial w_{11}^2} & \frac{\partial C}{\partial w_{12}^2} \\ \frac{\partial C}{\partial w_{21}^2} & \frac{\partial C}{\partial w_{22}^2} \end{bmatrix} \frac{\partial C}{\partial W^1} = \begin{bmatrix} \frac{\partial C}{\partial w_{11}^1} & \frac{\partial C}{\partial w_{12}^1} \\ \frac{\partial C}{\partial w_{21}^1} & \frac{\partial C}{\partial w_{22}^1} \end{bmatrix}$$

$$\frac{\partial C}{\partial b^2} = \begin{bmatrix} \frac{\partial C}{\partial b_1^2} & \frac{\partial C}{\partial b_2^2} \end{bmatrix} \frac{\partial C}{\partial b^1} = \begin{bmatrix} \frac{\partial C}{\partial b_1^1} & \frac{\partial C}{\partial b_2^1} \end{bmatrix}$$

Para calcular las derivadas haremos uso de la regla de la cadena

$$\frac{\partial C}{\partial W^2} = \begin{bmatrix} \frac{\partial C}{\partial a_1^2}\frac{\partial a_1^2}{\partial z_1^2}\frac{\partial z_1^2}{\partial w_{11}^2} & \frac{\partial C}{\partial a_2^2}\frac{\partial a_2^2}{\partial z_2^2}\frac{\partial z_2^2}{\partial w_{12}^2} \\ \frac{\partial C}{\partial a_1^2}\frac{\partial a_1^2}{\partial z_1^2}\frac{\partial z_1^2}{\partial w_{21}^2} & \frac{\partial C}{\partial a_2^2}\frac{\partial a_2^2}{\partial z_2^2}\frac{\partial z_2^2}{\partial w_{22}^2} \end{bmatrix}$$

$$\frac{\partial C}{\partial b^2} = \begin{bmatrix} \frac{\partial C}{\partial a_1^2}\frac{\partial a_1^2}{\partial z_1^2}\frac{\partial z_1^2}{\partial b_1^2} & \frac{\partial C}{\partial a_2^2}\frac{\partial a_2^2}{\partial z_2^2}\frac{\partial z_2^2}{\partial b_2^2} \end{bmatrix}$$

$$\frac{\partial C}{\partial W^1} = \begin{bmatrix} \frac{\partial C}{\partial a_1^2}\frac{\partial a_1^2}{\partial z_1^2}\frac{\partial z_1^2}{\partial a_1^1}\frac{\partial a_1^1}{\partial z_1^1}\frac{\partial z_1^1}{\partial w_{11}^1} + \frac{\partial C}{\partial a_2^2}\frac{\partial a_2^2}{\partial z_2^2}\frac{\partial z_2^2}{\partial a_1^1}\frac{\partial a_1^1}{\partial z_1^1}\frac{\partial z_1^1}{\partial w_{11}^1} & \frac{\partial C}{\partial a_1^2}\frac{\partial a_1^2}{\partial z_1^2}\frac{\partial z_1^2}{\partial a_2^1}\frac{\partial a_2^1}{\partial z_2^1}\frac{\partial z_2^1}{\partial w_{12}^1} + \frac{\partial C}{\partial a_2^2}\frac{\partial a_2^2}{\partial z_2^2}\frac{\partial z_2^2}{\partial a_2^1}\frac{\partial a_2^1}{\partial z_2^1}\frac{\partial z_2^1}{\partial w_{12}^1} \\ \frac{\partial C}{\partial a_1^2}\frac{\partial a_1^2}{\partial z_1^2}\frac{\partial z_1^2}{\partial a_1^1}\frac{\partial a_1^1}{\partial z_1^1}\frac{\partial z_1^1}{\partial w_{21}^1} + \frac{\partial C}{\partial a_2^2}\frac{\partial a_2^2}{\partial z_2^2}\frac{\partial z_2^2}{\partial a_1^1}\frac{\partial a_1^1}{\partial z_1^1}\frac{\partial z_1^1}{\partial w_{21}^1} & \frac{\partial C}{\partial a_1^2}\frac{\partial a_1^2}{\partial z_1^2}\frac{\partial z_1^2}{\partial a_2^1}\frac{\partial a_2^1}{\partial z_2^1}\frac{\partial z_2^1}{\partial w_{22}^1} + \frac{\partial C}{\partial a_2^2}\frac{\partial a_2^2}{\partial z_2^2}\frac{\partial z_2^2}{\partial a_2^1}\frac{\partial a_2^1}{\partial z_2^1}\frac{\partial z_2^1}{\partial w_{22}^1} \end{bmatrix}$$

$$\frac{\partial C}{\partial b^1} = \begin{bmatrix} \frac{\partial C}{\partial a_1^2}\frac{\partial a_1^2}{\partial z_1^2}\frac{\partial z_1^2}{\partial a_1^1}\frac{\partial a_1^1}{\partial z_1^1}\frac{\partial z_1^1}{\partial b_1^1} + \frac{\partial C}{\partial a_2^2}\frac{\partial a_2^2}{\partial z_2^2}\frac{\partial z_2^2}{\partial a_1^1}\frac{\partial a_1^1}{\partial z_1^1}\frac{\partial z_1^1}{\partial b_1^1} & \frac{\partial C}{\partial a_1^2}\frac{\partial a_1^2}{\partial z_1^2}\frac{\partial z_1^2}{\partial a_2^1}\frac{\partial a_2^1}{\partial z_2^1}\frac{\partial z_2^1}{\partial b_2^1} + \frac{\partial C}{\partial a_2^2}\frac{\partial a_2^2}{\partial z_2^2}\frac{\partial z_2^2}{\partial a_2^1}\frac{\partial a_2^1}{\partial z_2^1}\frac{\partial z_2^1}{\partial b_2^1} \end{bmatrix}$$

5 Descenso por Gradiente

La actualización de los pesos w_{ij} y los biases b_j en una red neuronal utilizando descenso por gradiente se realiza de la siguiente manera:

Actualización de los Pesos

$$w_{ij}^{(t+1)} = w_{ij}^{(t)} - \eta \frac{\partial C}{\partial w_{ij}} \tag{1}$$

Actualización de los Biases

$$b_j^{(t+1)} = b_j^{(t)} - \eta \frac{\partial C}{\partial b_j} \tag{2}$$

Versión Combinada

Si consideramos todos los pesos $\mathbf{W}$ y biases $\mathbf{b}$:

$$\mathbf{W}^{(t+1)} = \mathbf{W}^{(t)} - \eta \nabla_{\mathbf{W}} C \tag{3}$$

$$\mathbf{b}^{(t+1)} = \mathbf{b}^{(t)} - \eta \nabla_{\mathbf{b}} C \tag{4}$$