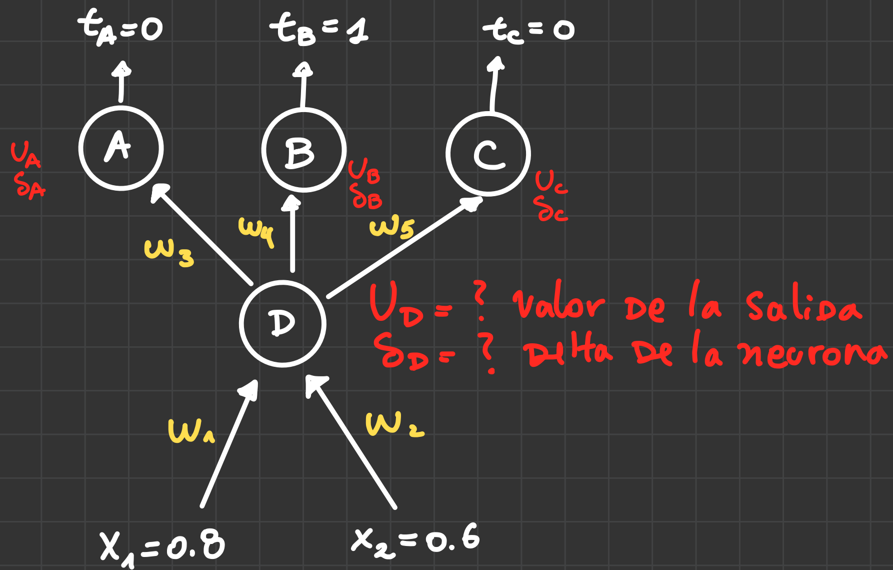
Joaquín Zepeda
20.127.150-9

→ para cada unidad se usa la función de activación logística:

$$u = sgm(x) = \frac{1}{1+e^{-x}}$$

y función de costos el error cuadrático:

$$E_K = (t_K - u_K)^2$$



Pesos inicialmente nulos, neuronas sin bias.

$$\boxed{\begin{aligned} W_3 &= W_A \\ W_4 &= W_B \\ W_5 &= W_C \end{aligned}}$$

Forward pass:

D] $\underbrace{S_D}_{\text{salida lineal neurona D}} = w_1 \cdot x_1 + w_2 \cdot x_2 + \cancel{b_D} = 0$ pesos inicialmente nulos y sin bias

Salida no lineal } $U_D = \text{sgm}(0) = \frac{1}{1+e^0} = 0.5$

Sigamos con la capa de salida:

A] $S_A = w_3 \cdot U_D = 0$

B] $S_B = w_4 \cdot U_D = 0$

C] $S_C = w_5 \cdot U_D = 0$

pues los pesos son inicialmente nulos y no hay bias

Luego, aplicando la función de activación logística:

A] $U_A = \text{sgm}(0) = 0.5 \rightarrow \varepsilon_A^2 = (0-0.5)^2 = 0.25$

B] $U_B = \text{sgm}(0) = 0.5 \rightarrow \varepsilon_B^2 = (1-0.5)^2 = 0.25$

C] $U_C = \text{sgm}(0) = 0.5 \rightarrow \varepsilon_C^2 = (0-0.5)^2 = 0.25$

$$\varepsilon_{T_K} = \varepsilon_A^2 + \varepsilon_B^2 + \varepsilon_C^2 = 3 \cdot 0.25 = 0.75$$

Backwards pass:

El w_3 se relaciona con la unidad A:

$$\Delta w_{ji}^K = \mu \cdot \delta_j^K \cdot z_i^K = \mu \cdot \frac{\partial(\varepsilon_{T_K})}{\partial S_{jK}} \cdot z_i^K$$

notación: $\tilde{\varepsilon}_K^2 = (t_K - \text{sgm}(s_K))^2$

$\varepsilon_{T_K} = \sum_j \tilde{\varepsilon}_{jK}^2$ } error total

Determinemos los Deltas:

$$\delta_A = -\frac{\partial \mathcal{E}_T}{\partial s_A} = -\frac{\partial}{\partial s_A}\Big(\sum_k \varepsilon_k^2\Big)$$

$$= -\frac{\partial}{\partial s_A}\Big(\varepsilon_A^2(s_A) + \varepsilon_B^2(s_B) + \varepsilon_c^2(s_c)\Big)$$

$$= -2\cdot\varepsilon_A\cdot\frac{\partial \varepsilon_A}{\partial s_A}$$

$$= -2\cdot\varepsilon_A\cdot\frac{\partial}{\partial s_A}\big(t_A - \mathrm{sgm}(s_A)\big)$$

$$= -2\cdot\varepsilon_A\cdot\big(-\mathrm{sgm}'(s_A)\big)$$

$$= 2\cdot\varepsilon_A\cdot\mathrm{sgm}(s_A)\cdot(1-\mathrm{sgm}(s_A))$$

$$\therefore\ \delta_A = 2\cdot\varepsilon_A\cdot\mathrm{sgm}(s_A)\cdot(1-\mathrm{sgm}(s_A))$$

$$= 2\cdot(-0{,}5)\cdot 0.5\cdot(1-0{,}5)$$

$$= -2\cdot 0.5^3$$

$$\frac{\partial\,\mathrm{sgm}(s)}{\partial s} = \frac{\partial}{\partial s}\left(\frac{1}{1+e^{-s}}\right)$$

$$= +(1+e^{-s})^{-2}\cdot +1\cdot e^{-s}$$

$$= \frac{1}{1+e^{-s}}\cdot\frac{1}{1+e^{-s}}\cdot\frac{1}{e^{s}}$$

POR ☆

$$= \underbrace{\frac{1}{1+e^{-s}}}\cdot\frac{1}{1+e^{s}}$$

$$= \mathrm{sgm}(s)\cdot(1-\mathrm{sgm}(s))$$

$$\bigstar\quad \frac{1}{1+e^{s}} = \frac{1}{1+e^{-s}}\cdot\frac{1}{e^{s}}$$

$$= \frac{e^{-s}}{1+e^{-s}}$$

$$\overset{\star}{=} 1 - \frac{1}{1+e^{-s}}$$

Ahora procediendo de igual manera para B y C:

$$\delta_B = -\frac{\partial \mathcal{E}_T}{\partial s_B} = -\frac{\partial}{\partial s_B}\Big(\sum_k \varepsilon_k^2\Big) = -2\cdot\varepsilon_B\cdot\frac{\partial \varepsilon_B}{\partial s_B}$$

$$= 2\cdot\varepsilon_B\cdot\mathrm{sgm}(s_B)\cdot(1-\mathrm{sgm}(s_B))$$

$$= 2\cdot 0.5\cdot 0.5\cdot(1-0.5)$$

$$= 2\cdot 0.5^3$$

$$\delta_c = -\frac{\partial \mathcal{E}_T}{\partial s_c} = -\frac{\partial}{\partial s_c}\Big(\sum_k \varepsilon_k^2\Big) = -2\cdot\varepsilon_c\cdot\frac{\partial \varepsilon_c}{\partial s_c}$$

$$= 2\cdot\varepsilon_c\cdot\mathrm{sgm}(s_c)\cdot(1-\mathrm{sgm}(s_c))$$

$$= 2\cdot -0.5\cdot 0.5\cdot(1-0.5)$$

$$= -2\cdot 0.5^3$$

En Resumen:

$$\boxed{\delta_A = -\delta_B = \delta_c = -2\cdot 0{,}5^3 = -0{,}25}$$

Veamos ahora el delta de la unidad D:

$$\delta_D = -\frac{\partial(\mathcal{E}_{T_k})}{\partial s_D} = -\sum_j \frac{\partial \mathcal{E}_T}{\partial s_j} \frac{\partial s_j}{\partial s_D}$$

$\frac{\partial \mathcal{E}_T}{\partial s_j} \rightarrow \delta_j$ } los deltas de la capa de salida

$$\left(\begin{array}{l} s_j = w_j \cdot sgm(s_i) \\ \frac{\partial s_j}{\partial s_i} = w_j \cdot sgm'(s_i) = w_j \cdot sgm(s_i)(1 - sgm(s_i)) \end{array} \right)$$

$$\Rightarrow \delta_D = -\left(\delta_A \cdot \frac{\partial s_A}{\partial s_D} + \delta_B \cdot \frac{\partial s_B}{\partial s_D} + \delta_C \cdot \frac{\partial s_C}{\partial s_D} \right)$$

$$= -\Big(\delta_A \cdot W_A \cdot sgm(s_D)(1 - sgm(s_D))$$
$$+ \delta_B \cdot W_B \cdot sgm(s_B)(1 - sgm(s_D))$$
$$+ \delta_C \cdot W_C \cdot sgm(s_C)(1 - sgm(s_D)) \Big)$$

$$= -\underbrace{sgm(s_D)(1 - sgm(s_D))}_{sgm'(s_D)} \cdot \left[\delta_A \cdot W_A + \delta_B \cdot W_B + \delta_C \cdot W_C \right]$$

Reemplazando $W_A = W_B = W_C = 0$

$$\Rightarrow \boxed{\delta_D = 0}$$

<u>Nota</u>: subíndice j se refiere a la capa de salida
subíndice i se refiere a la capa oculta.

② Considere $\mu = 0.1$

en donde

$$W_3 = W_A \quad \delta_3 = \delta_A \quad Z_3 = U_D$$

$$\therefore \quad \Delta W_3 = \mu \cdot \delta_3 \cdot Z_3$$

$$\Delta W_A = \mu \cdot \delta_A \cdot U_D = 0.1 \cdot (-0.25) \cdot 0.5 = -0.0125$$

$$\Delta W_B = \mu \cdot \delta_B \cdot U_D = 0.1 \cdot 0.25 \cdot 0.5 = 0.0125$$

$$\Delta W_C = \mu \cdot \delta_C \cdot U_D = 0.1 \cdot (-0.25) \cdot 0.5 = -0.0125$$

$$\Delta W_1 = \mu \cdot \delta_D \cdot X_1 = 0$$

$$\Delta W_2 = \mu \cdot \delta_D \cdot X_2 = 0$$

Por lo tanto, los nuevos pesos son:

Hidden layer
$$\begin{cases} W_1^+ = W_1 + \Delta W_1 = 0 \\ W_2^+ = W_2 + \Delta W_2 = 0 \end{cases}$$

output layer
$$\begin{cases} W_3^+ = W_A^+ = W_A + \Delta W_A = -0.0125 \\ W_4^+ = W_B^+ = W_B + \Delta W_B = +0.0125 \\ W_5^+ = W_C^+ = W_C + \Delta W_C = -0.0125 \end{cases}$$