Ejercicio teórico

Sea una red neuronal de dos capas, la primera de 3 neuronas y la segunda de 1 con los parámetros inicializados con los siguientes valores:

$$w^{(1)} = \begin{pmatrix} 0.1 & -0.5 \\ -0.3 & -0.9 \\ 0.8 & 0.02 \end{pmatrix}, b^{(1)} = \begin{pmatrix} 0.1 \\ 0.5 \\ 0.8 \end{pmatrix}, w^{(2)} = (-0.4 \quad 0.2 \quad -0.5), b^{(2)} = 0.7$$

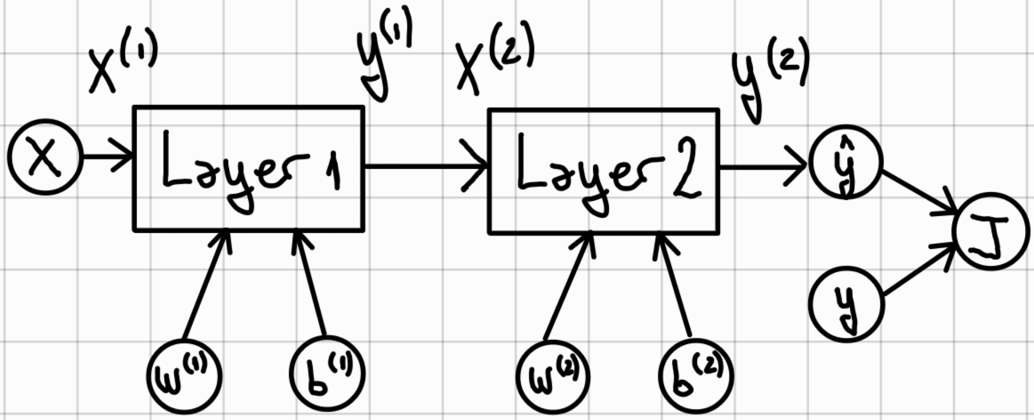
y donde cada capa calcula su salida vía

$$y^{(i)} = \sigma(w^{(i)} \cdot x^{(i)} + b^{(i)})$$

donde $\sigma(z) = \frac{1}{1+e^{-z}}$ es la función sigmoidea .

\ Dada la observación $x = \begin{pmatrix} 1.8 \\ -3.4 \end{pmatrix}$, $y = 5$ y la función de costo $J(\theta) = \frac{1}{2}(\hat{y}_\theta - y)^2$, calcular las derivadas de J respecto de cada parámetro $w^{(1)}$, $w^{(2)}$, $b^{(1)}$, $b^{(2)}$.

Nota: Con una sigmoidea a la salida jamás va a poder estimar el 5 "pedido", pero eso no afecta al mecanismo de backpropagation!



Redes neuronales: Resumen

Sea $X \in \mathbb{R}^n$, $W \in \mathbb{R}^{k\times n}$, $b \in \mathbb{R}^k$ y $g : \mathbb{R} \to \mathbb{R}$ no lineal, si se recibe $\frac{dJ}{dy} \in \mathbb{R}^k$ entonces:

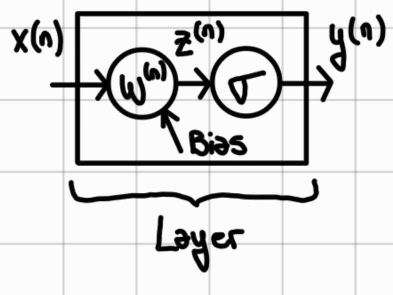
$$\frac{dJ}{db} = dY \odot g'(z)$$

$$\frac{dJ}{dW} = dY \odot g'(z) \cdot X^T$$

$$\frac{dJ}{dX} = W^T \cdot dY \odot g'(z)$$

donde $z = W \cdot X + b \in \mathbb{R}^k$ e $y = g(z)$ con g aplicada elemento a elemento.
Luego como $y^{(m)} = x^{(m+1)}$, entonces $\frac{dJ}{dX^{(m+1)}} = \frac{dJ}{dy^{(m)}}$ que es lo que le pasamos a la capa anterior.

- $\hat{y} = y^{(2)}$
- $x^{(2)} = y^{(1)}$



Para calcular la salida de la red neuronal, necesitamos propagar la entrada a través de las capas. La entrada x pasa por la primer capa oculta con pesos w(1) y bias b(1):

$$z^{(1)} = \underbrace{\begin{pmatrix} 0{,}1 & -0{,}5 \\ -0{,}3 & -0{,}9 \\ 0{,}8 & 0{,}02 \end{pmatrix}}_{3X2} \underbrace{\begin{pmatrix} 1{,}8 \\ -3{,}4 \end{pmatrix}}_{2X1} + \underbrace{\begin{pmatrix} 0{,}1 \\ 0{,}5 \\ 0{,}8 \end{pmatrix}}_{3X1} \rightarrow 3X1$$

$$z^{(1)} = \begin{pmatrix} 1{,}88 \\ 2{,}52 \\ 1{,}372 \end{pmatrix} + \begin{pmatrix} 0{,}1 \\ 0{,}5 \\ 0{,}8 \end{pmatrix} = \begin{pmatrix} 1{,}98 \\ 3{,}02 \\ 2{,}172 \end{pmatrix}$$

Ahora se aplica la función sigmoide a z(1):

$$y^{(1)} = \sigma\left(z^{(1)}\right) = \begin{pmatrix} \sigma(1{,}98) \\ \sigma(3{,}02) \\ \sigma(2{,}172) \end{pmatrix} = \begin{pmatrix} 1/1+e^{-1{,}98} \\ 1/1+e^{-3{,}02} \\ 1/1+e^{-2{,}172} \end{pmatrix} = \begin{pmatrix} 0{,}879 \\ 0{,}953 \\ 0{,}898 \end{pmatrix}$$

La salida de la primer capa se pasa como entrada a la segunda capa:

$$z^{(2)} = \underbrace{\begin{pmatrix} -0{,}4 & 0{,}2 & -0{,}5 \end{pmatrix}}_{1X3} \underbrace{\begin{pmatrix} 0{,}879 \\ 0{,}953 \\ 0{,}898 \end{pmatrix}}_{3X1} + \underbrace{0{,}7}_{1X1} \rightarrow 1X1$$

$$z^{(2)} = 0{,}09$$

Finalmente, se aplica la función sigmoide para obtener y_hat:

$$\hat{y} = \sigma\left(z^{(2)}\right) = \frac{1}{1+e^{-0{,}09}} = 0{,}522$$

Con estos valores, es posible ahora proceder al cálculo de la función de costo:

$$J(\theta) = \frac{1}{2}\left(\hat{y}-y\right)^2 = \frac{1}{2}\left(0{,}522-5\right)^2$$

$$J(\theta) = 10{,}02$$

Ahora se procede a calcular las derivadas de la función de costo respecto a cada uno de los parametros w(1), w(2), b(1) y b(2).

$$\frac{\partial J}{\partial w^{(2)}} = \frac{\partial J}{\partial \hat{y}} \odot \sigma'\left(z^{(2)}\right) \cdot \underbrace{X^{(2)T}}_{= y^{(1)T}}$$

$$\frac{\partial J}{\partial \hat{y}} = \hat{y} - y = -4{,}478$$

$$\sigma'\left(z^{(2)}\right) = \sigma\left(z^{(2)}\right) \cdot \left[1-\sigma\left(z^{(2)}\right)\right] = 0{,}249$$

$$y^{(1)T} = \begin{pmatrix} 0{,}879 \\ 0{,}953 \\ 0{,}898 \end{pmatrix}^T = \begin{pmatrix} 0{,}879 & 0{,}953 & 0{,}898 \end{pmatrix}$$

$$\frac{\partial J}{\partial W^{(2)}} = -4{,}478 \cdot 0{,}249 \cdot \begin{pmatrix} 0{,}879 & 0{,}953 & 0{,}898 \end{pmatrix}$$

$$\boxed{\frac{\partial J}{\partial W^{(2)}} = \begin{pmatrix} -0{,}980 & -1{,}062 & -1{,}001 \end{pmatrix}} \in \mathbb{R}^{1\times 3}$$

$$\frac{\partial J}{\partial b^{(2)}} = \frac{\partial J}{\partial \hat{y}} \odot \sigma'\left(z^{(2)}\right) = -4{,}478 \cdot 0{,}249$$

$$\boxed{\frac{\partial J}{\partial b^{(2)}} = -1{,}115} \in \mathbb{R}^{1}$$

$$\frac{\partial J}{\partial x^{(2)}} = W^{(2)T} \cdot \frac{\partial J}{\partial \hat{y}} \odot \sigma'\left(z^{(2)}\right)$$

$$= \begin{pmatrix} -0{,}4 \\ 0{,}2 \\ -0{,}5 \end{pmatrix} \cdot \begin{pmatrix} -4{,}478 \end{pmatrix} \cdot 0{,}249$$

Para analizar la primer capa, se considera la siguiente expresión:

$$\frac{dJ}{dX^{(m+1)}} = \frac{dJ}{dy^{(m)}}$$

$$\Rightarrow \boxed{\frac{\partial J}{\partial y^{(1)}} = \begin{pmatrix} 0,446 \\ -0,223 \\ 0,558 \end{pmatrix}} \in \mathbb{R}^{3\times 1}$$

$$\frac{\partial J}{\partial w^{(1)}} = \frac{\partial J}{\partial y^{(1)}} \odot \sigma'\left(Z^{(1)}\right) \cdot X^{T}$$

$$Z^{(1)} = \begin{pmatrix} 1,98 \\ 3,02 \\ 2,172 \end{pmatrix}$$

$$\sigma'\left(Z^{(1)}\right) = \sigma\left(Z^{(1)}\right)\left[1 - \sigma\left(Z^{(1)}\right)\right]$$

$$= \begin{pmatrix} 0,879\,(1-0,878) \\ 0,953\,(1-0,953) \\ 0,898\,(1-0,898) \end{pmatrix} = \begin{pmatrix} 0,106 \\ 0,045 \\ 0,092 \end{pmatrix}$$

$$\frac{\partial J}{\partial W^{(1)}} = \begin{pmatrix} 0,446 \\ -0,223 \\ 0,558 \end{pmatrix} \odot \begin{pmatrix} 0,106 \\ 0,045 \\ 0,092 \end{pmatrix} \begin{pmatrix} 1,8 & -3,4 \end{pmatrix} = \begin{pmatrix} 0,05 \\ -0,01 \\ 0,05 \end{pmatrix} \begin{pmatrix} 1,8 & -3,4 \end{pmatrix}$$

$$\boxed{\frac{\partial J}{\partial w^{(1)}} = \begin{pmatrix} 0{,}09 & -0{,}17 \\ -0{,}02 & 0{,}03 \\ 0{,}09 & -0{,}17 \end{pmatrix}} \in \mathbb{R}^{3\times 2}$$

$$\frac{\partial J}{\partial b^{(1)}} = \frac{\partial J}{\partial y^{(1)}} \odot \sigma'(z^{(1)}) = \begin{pmatrix} 0{,}446 \\ -0{,}223 \\ 0{,}558 \end{pmatrix} \odot \begin{pmatrix} 0{,}106 \\ 0{,}045 \\ 0{,}092 \end{pmatrix}$$

$$\boxed{\frac{\partial J}{\partial b^{(1)}} = \begin{pmatrix} 0{,}05 \\ -0{,}01 \\ 0{,}05 \end{pmatrix}} \in \mathbb{R}^{3\times 1}$$