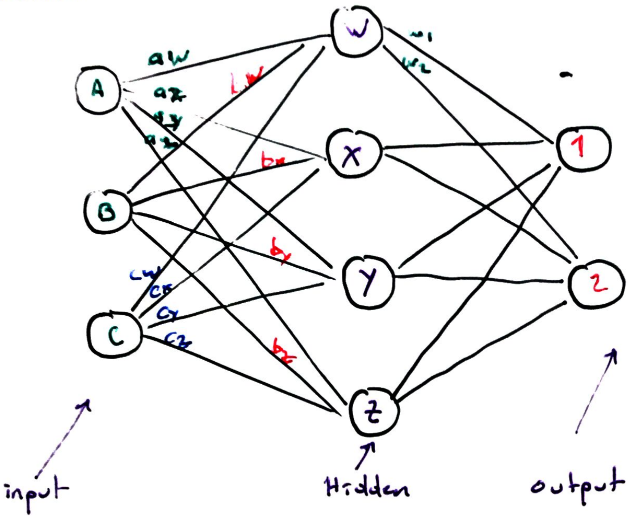
BASIC_NEURAL_NETWORKS

Feed Forward Back Propagation

Estructura



operaciones intermedia

$$A \cdot aw + B \cdot bw + C \cdot cw + W\_bias = W$$

$$[A, B, C] \cdot [aw, bw, cw] + W\_bias = W$$

funcion activacion

Sigmoid

$$= \phi = \frac{1}{1+e^{-x}}$$

ReLU

$$= \phi = \begin{cases} 0 \,; & x < 0 \\ x \,; & x > 0 \end{cases}$$

operaciones output

$$W \cdot w_1 + X \cdot x_1 + Y \cdot y_1 + Z \cdot z_1 + 1\_bias = 1$$

$$[W, X, Y, Z] \cdot [w_1, x_1, y_1, z_1] + 1\_bias = 1$$

Intermedia

$$[A, B, C] \cdot \begin{bmatrix} aw & ax & ay & az \\ bw & bx & by & bz \\ cw & cx & cy & cz \end{bmatrix} + \begin{bmatrix} w\_bias \\ x\_bias \\ y\_bias \\ z\_bias \end{bmatrix} \overset{f(x)}{\Rightarrow} \begin{bmatrix} W \\ X \\ Y \\ Z \end{bmatrix} \Rightarrow$$

Salida

funcion activación

$$[W, X, Y, Z] \cdot \underbrace{\begin{bmatrix} w_1 & w_2 \\ x_1 & x_2 \\ y_1 & y_2 \\ z_1 & z_2 \end{bmatrix}}_{\text{weights}} + \underbrace{\begin{bmatrix} 1\_bias \\ 2\_bias \end{bmatrix}}_{\text{bias}} \overset{f(x)}{\Rightarrow} \begin{bmatrix} 1 \\ 2 \end{bmatrix}$$

parametros de ajuste

Metodo de entrenamiento, back propagation

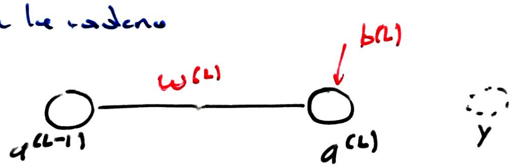
Funcion de costo

$C = cost[S, y]$ S: esperado y = salida

MSE cross entropy

Backpropagation y gradiente descendente

Regla de la cadena

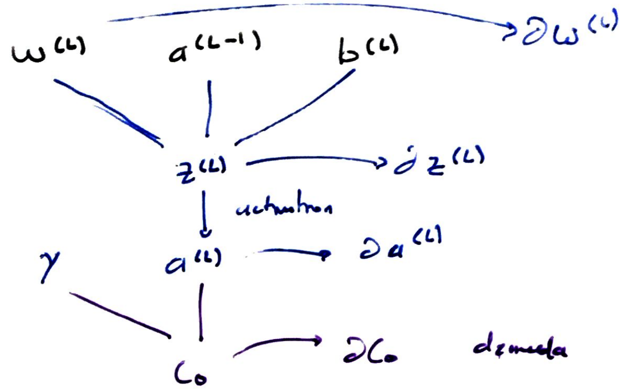


cost

$$\underbrace{C_0(\dots) = (a^{(L)} - y)^2}_{\text{costo por MSE}}$$

$$z^{(L)} = w^{(L)} \cdot a^{(L-1)} + b^{(L)}$$

obtencion de resultado



dependencia de C_0 con el $w^{(L)}$ (regla de la cadena

$$\frac{\partial C_0}{\partial w^{(L)}} = \frac{\partial z^{(L)}}{\partial w^{(L)}} \cdot \frac{\partial a^{(L)}}{\partial z^{(L)}} \cdot \frac{\partial C_0}{\partial a^{(L)}}$$

; $C_0 = (a^{(L)} - y)$ funcion de costo

$$\frac{\partial C_0}{\partial a^{(L)}} = 2\,(a^{(L)} - y) \Rightarrow$$ derivada funcion de costo

$a^{(L)} = \sigma(z^{(L)})$

σ: funcion de activacion (sigmoid)

$$\frac{\partial a^{(L)}}{\partial z^{(L)}} = \sigma'(z^{(L)})$$

σ': derivada del sigmoide

$$\frac{\partial z^{(L)}}{\partial w^{(L)}} = a^{(L-1)}$$

$$\sigma = \frac{1}{1+e^{-x}} \qquad \sigma' = \frac{e^{-x}}{(1+e^{-x})^2}$$

$$\frac{\partial C_0}{\partial w^{(L)}} = a^{(L-1)}\,\sigma'(z^{(L)})\,2\,(a^{(L)} - y)$$

$\sigma' = out(1-out)$

$$\frac{\partial C}{\partial w^{(L)}} = \frac{1}{n}\sum_{k=0}^{n-1} \frac{\partial C_k}{\partial w^{(L)}}$$

; el costo total

Se debe buscar esto respecto a todos los w y todos los bias

$$\nabla C = \left[\frac{\partial C}{\partial w^1} \quad \frac{\partial C}{\partial b^1} \quad \frac{\partial C}{\partial w^2} \quad \frac{\partial C}{\partial b^2} \cdots \frac{\partial C}{\partial w^L} \quad \frac{\partial C}{\partial b^L} \right]$$

en el bias $\frac{\partial z^{(L)}}{\partial b^{(L)}} = 1 \Rightarrow \frac{\partial C_0}{\partial b^{(L)}} = \sigma'(z^{(L)})\,2\,(a^{(L)} - y)$

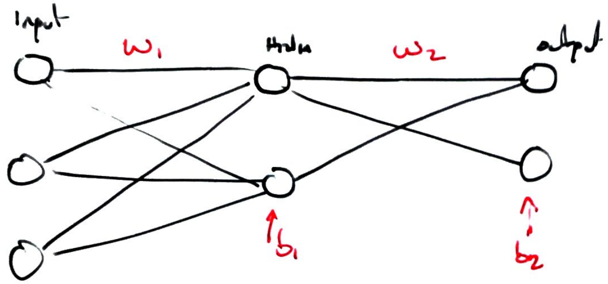
con respecto a la capa anterior

$$\frac{\partial C_0}{\partial a^{(L-1)}} = w^{(L)} \sigma'(z^{(L)})\, 2(a^{(L)}-y)$$

porque $\dfrac{\partial z^{(L)}}{\partial a^{(L-1)}} = w^{(L)}$

Para los datos de entrenamiento se hace la sumatoria de la funcion de costo

$$C_0 = \sum_{j=0}^{n-1} (a_j^{(L)} - y_j)^2$$



$$\frac{\partial C}{\partial w} = \sum$$

$$\frac{\partial C_0}{\partial a_k^{(L-1)}} = \sum_{j=0}^{n_{L-1}} \frac{\partial z_j^{(L)}}{\partial a_k^{(L-1)}} \frac{\partial a_j^{(L)}}{\partial z_j^{(L)}} \cdot \frac{\partial C_0}{\partial a_j^{(L)}}$$

$$\frac{\partial C}{\partial w^{(L)}} = \frac{\partial z^{(L)}}{\partial w^{L}} \cdot \frac{\partial a^{(L)}}{\partial z^{(L)}} \cdot \frac{\partial C}{\partial a^{L}}$$

para w^2

$$\frac{\partial C}{\partial w^2} = \underline{\frac{\partial z^2}{\partial w^2}} \cdot \underline{\frac{\partial a^2}{\partial z^2}} \cdot \underline{\frac{\partial C}{\partial a^2}} = a^1 \cdot \sigma'(z^2) \cdot 2(a^2 - y) \checkmark$$

$$\frac{\partial C}{\partial a^1} = \frac{\partial z^2}{\partial a^1} \cdot \frac{\partial a^2}{\partial z^2} \cdot \frac{\partial C}{\partial a^2} = w^2 \cdot \sigma'(z^2) \cdot 2(a^2 - y)$$

$$\frac{\partial C}{\partial b^2} = \frac{\partial z^2}{\partial b^2} \cdot \frac{\partial a^2}{\partial z^2} \cdot \frac{\partial C}{\partial a^2} = 1 \cdot \sigma'(z^2) \cdot 2(a^2 - y)$$

Para w^1

$$\frac{\partial C}{\partial w^1} = \frac{\partial z^1}{\partial w^1} \cdot \frac{\partial a^1}{\partial z^1} \cdot \frac{\partial C}{\partial a^1} = a^0 \cdot \sigma'(z^1) \frac{\partial C}{\partial a^1} =$$

$$\frac{\partial C}{\partial a^0} = \frac{\partial z^1}{\partial a^0} \cdot \frac{\partial a^1}{\partial z^1} \cdot \frac{\partial C}{\partial a^1} = w^1 \cdot \sigma'(z^1) \frac{\partial C}{\partial a^1}$$

$$\frac{\partial C}{\partial b^1} = \frac{\partial z^1}{\partial b^1} \cdot \frac{\partial a^1}{\partial z^1} \cdot \frac{\partial C}{\partial a^1} = \sigma'(z^1) \frac{\partial C}{\partial a^1}$$

actualiz. el gradiente.

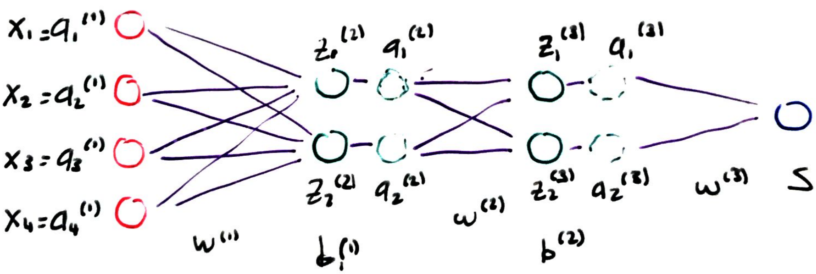
$$w_1 = w_0 - r\frac{\partial C}{\partial w}$$ r: learning rate $= 0{,}1$

Matrix propagation

$$\frac{\partial C}{\partial a^1} = w^2 \cdot \sigma'(z^2) \cdot 2(a^2 - y)$$

layer out $[4,3] \cdot \qquad \cdot [3,100]$ $\qquad 4 \to 3$

Ejemplo 2:



$x_i = a_i^{(1)}$ capa de entrada

$$z_1^{(2)} = W^{(1)} x + b^{(1)} \Rightarrow a^{(2)} = f(z^{(2)})$$

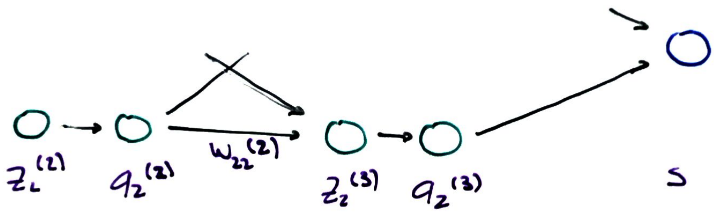
valor neurona — valor activacion

$$z^{(3)} = W^{(2)} a^{(2)} + b^{(2)} \Rightarrow a^{(3)} = f(z^{(3)})$$

$s = W^{(3)} a^{(3)}$ capa de salida

$C = \mathrm{cost}(s, y)$: funcion de costo

Backpropagation



$$\frac{\partial C}{\partial w^{(L)}} = \frac{\partial z^{(L)}}{\partial w^{(L)}} \cdot \frac{\partial a^{(L)}}{\partial z^{(L)}} \cdot \frac{\partial C}{\partial a^{(L)}} = a^{(L-1)} \cdot \sigma'(z^{(L)}) \cdot 2(a^{(L)} - y) \checkmark$$

$$\frac{\partial C}{\partial a^{(L-1)}} = \frac{\partial z^{(L)}}{\partial a^{(L-1)}} \cdot \frac{\partial a^{(L)}}{\partial z^{(L)}} \cdot \frac{\partial C}{\partial a^{(L)}} = w^{L} \cdot \sigma'(z^{(L)}) \cdot 2(a^{(L)} - y) \checkmark$$

$$\frac{\partial C}{\partial b^{(L)}} = \frac{\partial z^{(L)}}{\partial b^{(L)}} \cdot \frac{\partial a^{(L)}}{\partial z^{(L)}} \cdot \frac{\partial C}{\partial a^{(L)}} = 1 \cdot \sigma'(z^{(L)}) \cdot 2(a^{(L)} - y) \checkmark$$

Notas: para la ultima capa el $a^{(L-1)}$ es la activacion de la capa anterior, para la primera capa el $a^{(L-1)}$ es la entrada de datos.