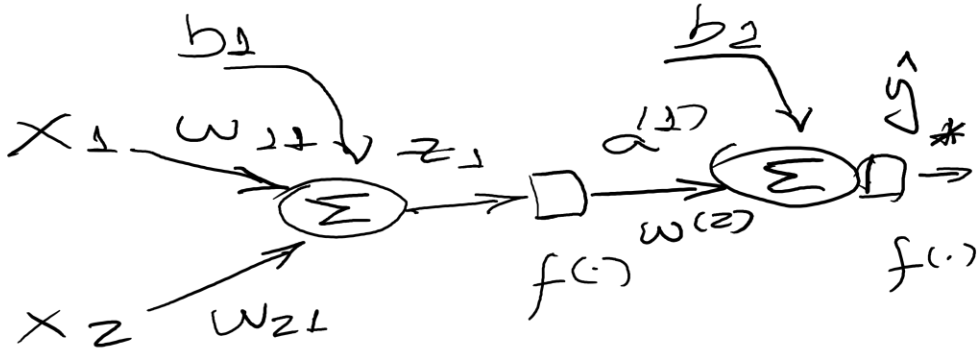
Suposição:

ENTRADA: $x = [x_1, x_2]$

oculta: 2 neuron $\rightarrow W^{(1)}$

SAÍDA: 1 neuron $\rightarrow W^{(2)}$

função: sigmoide $\rightarrow \sigma(z) = \frac{1}{1+e^{-z}}$



CAMADA OCULTA:

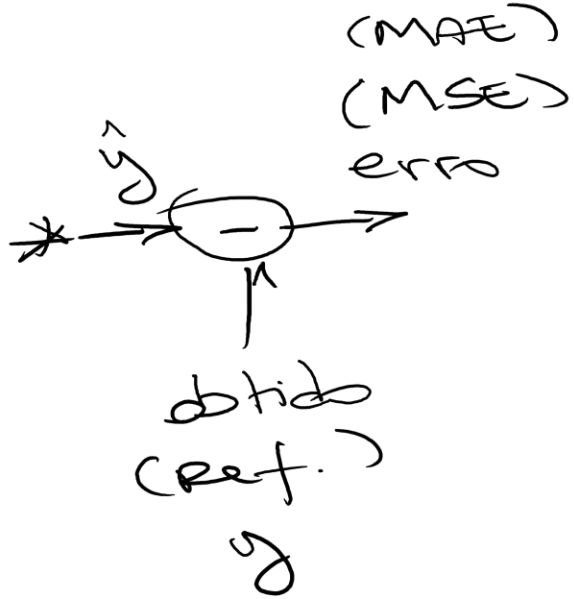
$$z^{(1)} = x_1 w_{11}^{(1)} + x_2 w_{21}^{(1)} + b_1^{(1)}$$

$$a^{(1)} = \sigma(z^{(1)})$$

CAMADA SAÍDA:

$$z^{(2)} = a^{(1)} w^{(2)} + b^{(2)}$$

$$\hat{y} = \sigma(z^{(2)})$$



Erro quadrático médio (MSE)

$$E = \frac{1}{2}(y_{desejado} - \hat{y}_{obtido})^2 \rightarrow \hat{y}_{obtido} = \sigma(z^{(2)})$$

Gradiente da saída:

$$\frac{\partial E}{\partial \hat{y}} = \hat{y} - y$$

$$\frac{\partial \hat{y}}{\partial z^{(2)}} = \hat{y}(1-\hat{y})$$

erro da saída $(\delta^{(2)}) = (\hat{y} - y) \cdot \hat{y}(1-\hat{y})$

$$\frac{\partial E}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial z^{(2)}} = \frac{\partial E}{\partial z^{(2)}} = \delta^{(2)}$$

Gradientes nos pesos:

$$\frac{\partial E}{\partial W^{(2)}} = \frac{\partial E}{\partial z^{(2)}} \cdot \frac{\partial z^{(2)}}{\partial W^{(2)}} = \delta^{(2)} \cdot a^{(1)}$$

$$\Rightarrow \frac{\partial z^{(2)}}{\partial a^{(1)}} = W^{(2)}$$

$$\boxed{\delta^{(1)} = \delta^{(2)} \cdot W^{(2)} \cdot a^{(1)}(1-a^{(1)})}$$

Gradiente nas entradas:

$$\frac{\partial E}{\partial w^{(1)}} = \delta^{(1)} \cdot x$$

Atualização dos Pesos:

$$\begin{cases} w^{(2)} \simeq w^{(2)} - \eta \cdot \dfrac{\partial E}{\partial w^{(2)}} \leftarrow \\ w^{(1)} \simeq w^{(1)} - \eta \cdot \dfrac{\partial E}{\partial w^{(1)}} \leftarrow \\ \eta \leftarrow \text{taxa de aprendizagem } (0.01 \leq \eta \leq 0.1) \end{cases}$$