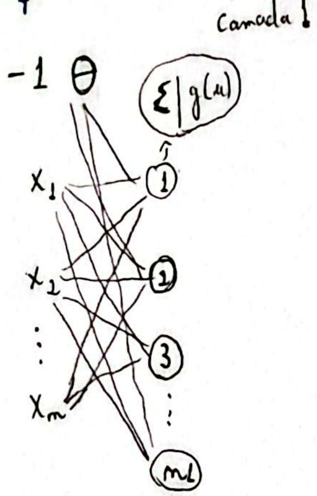
Aula de Multilayer Perceptron (MLP)

Esquema



$\Sigma = I_j^{camada}$

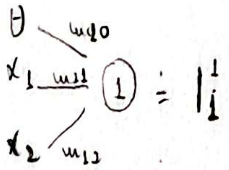
↓
nesse caso

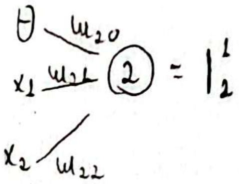
I_j^1

2 neurônios e 1 camada

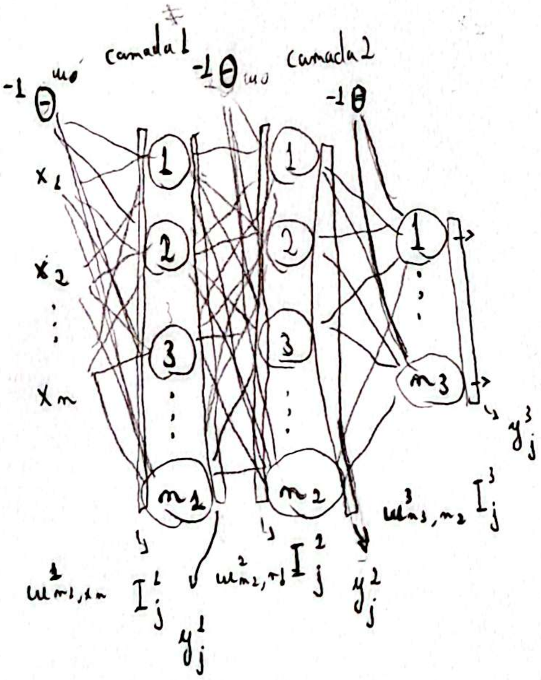
matriz de pesos

$$w_{ji} = \begin{bmatrix} w_{11} & w_{12} \\ w_{21} & w_{22} \end{bmatrix}$$





$w_{j,i}^{camada}$ -> j é os neurônios
i é os pesos



j = neurônio

$I_j^{camada} = \sum w_i \cdot x_i$

$w^{m}_{m+1,m-1}$ = matriz de pesos

y_j^{camada} = saída dos neurônios

Fase Forward

Camada 1 (c=1)

$$I_j^1 = \sum_{i=0}^{m1} w_{ji} \cdot x_i \rightarrow I_j^1 = w_{10}^1 \cdot x_0 + w_{11}^1 \cdot x_1 + w_{12}^1 \cdot x_2 \cdots w_{1n}^1 \cdot x_n \rightarrow y_j^1 = g(I_j^1)$$

Camada 2 (c=2)

$$I_j^2 = \sum_{i=0}^{m2} w_{ji} \cdot y_i^1 \rightarrow I_j^2 \Rightarrow y_j^2 = g(I_j^2)$$

Camada 3 (c=3)

$$I_j^3 = \sum_{i=0}^{m3} w_{ji} \cdot y_j^2 \rightarrow I_j^3 \Rightarrow y_j^3 = g(I_j^3)$$

Generalizando

$$\boxed{I_j^c = \sum_{i=0}^{m} w_{ji}^c \cdot x_i}$$

Regras de derivação

$$\frac{\partial a}{\partial b}(f(g)) = f'(g) \cdot g'$$

$$(a-x) = -1$$

Cálculo do erro

Considerando que K é o indice da amostra:

$c=3$

$$e(K) = \frac{1}{2} \sum_{j=1}^{m^c} \left[d(K) - Y_j^c(K)\right]^2 \quad d = \text{desejado}$$

c

$$EQM = \frac{1}{p} \sum_{K=0}^{p} e(K)$$

Fase Backward (Backpropagation)

Nosso interesse é calcular o gradiente dos erros

$$\nabla e^3 = \frac{\partial e^3}{\partial w_{ij}^3} \quad \rightsquigarrow \frac{\partial e^3}{\partial w_{ij}^3} = \overset{1}{\frac{\partial e^3}{\partial y_j^3}} \cdot \overset{2}{\frac{\partial y_j^3}{\partial I_j^3}} \cdot \overset{3}{\frac{\partial I_j^3}{\partial w_{ji}}}$$

Passo 1: Derivar o e em relação a y_j^3

$$\frac{\partial e^3}{\partial y_j^3} = -(d_j - y_j^3) \rightarrow \text{Destrinchando...} \quad \frac{\partial e}{\partial y} f(u) = f'(u) \cdot u'$$

$$e(K) \frac{1}{2}\left(\sum_{j=1}^{m} d_j - \sum_{j=1}^{m} y_j\right)^2$$

↓ ↓ (vai somar e virar um escalar)

$$\frac{1}{2}(d - y)^2$$

$$f(\cdot) = (\cdot)^2$$

$$u = d - y$$

$$f'(u) = (d-y)$$

$$u' = -1$$

$$\frac{\partial e^3}{\partial y_j^3}(f(u)) = -(d-y)$$

$$\frac{\partial e^3}{\partial y_j^3} = -(d-y) \quad - \quad \frac{\partial y_j^3}{\partial l_j^3} = \sec h^2(y_j^3) \rightarrow g'(l)$$

$$\frac{\partial y_j^3}{\partial l_j^3} = \tanh(l_j^3) \rightarrow \tanh'(l_j^3) \rightarrow \boxed{\sec h^2(y_j^3)} \Big\} \text{ Passo 2}$$

$$\frac{\partial l_j^3}{\partial w_{j,i}^3} = y_j^2 \rightarrow \left(l_j^3 = \sum_{i=0}^{m2} w_{ij}^3 \cdot y_j^2 \right)' \rightarrow y_j^2 \Big\} \text{ Passo 3}$$

$$\frac{\partial e^3}{\partial w_{ij}^3} = -(d-y) \cdot g'(l) \cdot y_j^2 = \nabla e^3$$

Passo 4: Aplicando a regra de atualização da camada 3

$$\nabla e^3 = \frac{\partial e^3}{\partial w_{ij}^3} \qquad \boxed{\Delta w_{ji}^3 = \eta \cdot \frac{\partial e^3}{\partial w_{ij}^3}} \rightarrow \Delta w_{ji}^3 = \eta \left[\underbrace{\boxed{-(d-y) \cdot g'(l)}}_{\delta_j^3} \cdot y_j^2 \right]$$

$$w_{ji}^3(t+1) = w_{ji}^3(t) + \underbrace{\left[\eta \cdot \delta_j^3 \cdot y_j^2 \right]}_{\Delta w_{ji}^3} \rightarrow \boxed{w_{ji}^3(t+1) = w_{ji}^3(t) + \Delta w_{ji}^3}$$ ajustes dos pesos da camada 3

Camada 2 (intermediária)

$$\nabla e^2 = \frac{\partial e^2}{w_{ji}^2} \rightarrow \frac{\partial e^2}{\partial w_{ji}^2} = \overset{①}{\boxed{\frac{\partial e^2}{\partial y_j^2}}} \cdot \overset{②}{\frac{\partial y_j^2}{\partial l_j^2}} \cdot \overset{③}{\frac{\partial l_j^2}{\partial w_{ji}^2}}$$

O erro de ajuste dos pesos calculado com base em l_k^3 e δ_k^3

$$\frac{\partial e^2}{\partial y_j^2} = \sum_{k=1}^{m3} \underbrace{\boxed{\frac{\partial e^3}{\partial l_k^3}}}_{\delta_k^3} \cdot \underbrace{\boxed{\frac{\partial l_k^3}{\partial y_j^2}}}_{w_{kj}^3}$$ → Pesos já ajustados

$$\frac{\partial l_k^3}{\partial y_j^2} = \frac{\partial \left(\sum_{k=1}^{m3} w_{kj}^3 \cdot \cancel{y_j^2} \right)}{\cancel{\partial (y_j^2)}} \rightarrow \frac{\partial l_k^3}{\partial y_j^2} = w_{kj}^3$$

$$\boxed{\frac{\partial e^2}{\partial y_j^2} = -\left(\sum_{k=1}^{m3} \delta_k^3 \cdot w_{kj}^3 \right)}$$

$$\frac{\partial e^2}{\partial w_{ji}^2} = -\left(\sum_{k=1}^{m3} \delta_k^3 \cdot w_{kj}^3 \right) \cdot g'(l_j^2) \cdot y_j^1 = \nabla e^2$$

$$\boxed{\frac{\partial y_j^2}{\partial l_j^2} = g'(l_j^2)}$$

$$\boxed{\frac{\partial l_j^2}{\partial w_{ji}^2} = y_j^1}$$

Aplicação da regra de atualização da camada 2

$$\nabla e^2 = \frac{\partial e^2}{\partial w_{ji}^2} = \eta \cdot \left[\underbrace{\left(-\left(\sum_{k=1}^{n_3} \delta^3 \, w_{kj}^3\right) \cdot g'(I_j^2)\right)}_{\delta_j^2} \cdot y_j^1 \right]$$

$$\Delta w_{ji}^2 = \eta \cdot \delta_j^2 \cdot y_j^1 \rightarrow w_{ji}^2(t+1)$$

Camada 1 (entradas)

$$\nabla e^1 = \frac{\partial e^1}{\partial w_{ji}^1}$$

$$\frac{\partial e^1}{\partial w_{ji}^1} = \boxed{\frac{\partial e^1}{\partial y_j^1}} \cdot \frac{\partial y_j^1}{\partial I_j^1} \cdot \frac{\partial I_j^1}{\partial w_{ji}^1}$$

$$\frac{\partial e^1}{\partial y_j^1} = \sum_{k=1}^{n_2} \underbrace{\boxed{\frac{\partial e^1}{\partial I_k^2}}}_{\delta_k^2} \cdot \underbrace{\boxed{\frac{\partial I_k^2}{\partial y_k^1}}}_{w_{kj}^2} \rightarrow \frac{\partial\left(\sum_{k=1}^{n_2} w_{kj}^2 \cdot \cancel{y_j^1}\right)}{\partial(\cancel{y_j})} \rightarrow w_{kj}^2$$

$$\boxed{\frac{\partial e^1}{\partial y_j^1} = -\sum_{k=1}^{n_2} \delta_k^2 \cdot w_{kj}^2}$$

$$\boxed{\frac{\partial y_j^1}{\partial I_j^1} = g'(I_j^1) \rightarrow \mathrm{sech}^2(I_j^1)}$$

$$\boxed{\frac{\partial I_j^1}{\partial w_{ji}^1} = x_i}$$ essa derivada é sempre a entrada da camada situada

$$\frac{\partial e^1}{\partial w_{ij}^1} = -\left(\sum_{k=1}^{n_2} \delta_k^2 \cdot w_{kj}^2\right) \cdot g'(I_j^1) \cdot x_i$$

Aplicação da regra de atualização da camada 1

$$\nabla e^1 = \frac{\partial e^1}{\partial w_{ji}^1} \rightarrow -\left(\sum_{k=1}^{n_2} \delta_k^2 \cdot w_{kj}^2\right) \cdot g'(I_j^1) \cdot x_i$$

$$\Delta w_{ji}^1 = \eta \cdot \left[\underbrace{-\left(\sum_{k=1}^{n_2} \delta_k^2 \cdot w_{kj}^2\right) \cdot g'(I_j^1)}_{\delta_k^1} \cdot x_i\right]$$

$$\Delta w_{ji}^1 = \eta \cdot \delta_k^1 \cdot x_i$$

$$w_{ji}^1(t+1) = w_{ji}^1(t) + \underbrace{\boxed{\eta \cdot \delta_k^1 \cdot x_i}}_{\Delta w_{ji}^1 =}$$

$$\boxed{w_{ji}^1(t+1) = w_{ji}^1(t) + \Delta w_{ji}^1}$$