Weights Correction

$$w_{ji}(n+1) = w_{ji}(n) + \Delta w_{ji}(n)$$

$$\Delta w_{ji}(n) = \eta\, \delta_j x_i$$

$$\delta_j = e_j\, \phi'(v_j(n)) = \frac{\partial \mathcal{L}(n)}{\partial e_j(n)} \times \frac{\partial e(n)}{\partial y_j(n)} \times \frac{\partial y_j(n)}{\partial v_j(n)}$$

$$\phi'(v_j(n)) = \frac{d}{dv_j} \text{ (Activation Function)}$$

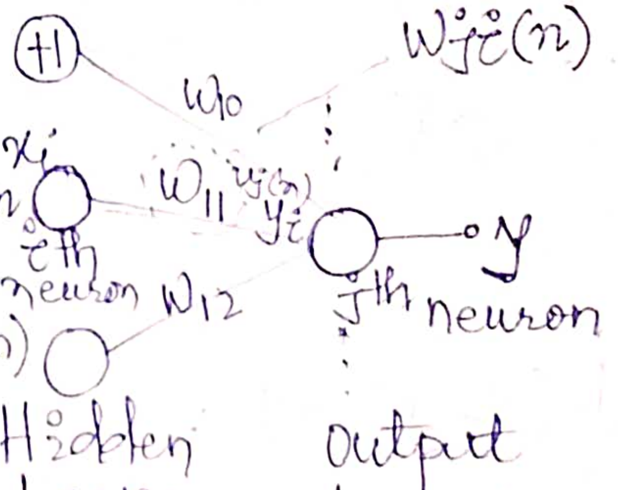
$$\frac{\partial y_j(n)}{\partial v_j(n)} = \phi'(v_j(n))$$

Update weights in output layer

$\Delta w_{ji}(n)$

nth training example or observation

$$\Delta w_{ji}(n) = \eta\, \delta_j(n)\, y_i(n)$$



$$\frac{\partial \mathcal{L}(n)}{\partial w_{ji}(n)} = \underbrace{\frac{\partial \mathcal{L}(n)}{\partial e_j(n)} \times \frac{\partial e_j(n)}{\partial y_j(n)} \times \frac{\partial y_j(n)}{\partial v_j(n)}}_{\delta_j(n)} \times \frac{\partial v_j(n)}{\partial w_{ji}(n)}$$

$$v_j(n) = \sum_{i=0}^{m} w_{ji}(n)\, x_i(n)$$

loss function / cost function $\mathcal{L}(n)$

$\Downarrow$

$$\mathcal{L}(n) = \frac{1}{N} \sum_{n=1}^{N} \sum_{j=1}^{K} e_j^2(n)$$

$N \rightarrow$ # of training examples

$$y_j(n) = \phi(v_j(n))$$

$$e_j(n) = \cancel{\tfrac{1}{2}}\left[d_j(n) - y_j(n)\right]^2$$

$$e_j(n) = \cancel{\tfrac{1}{2}}\left[y_j(n) - \hat{y}_j(n)\right]^2$$

$$\mathcal{E}(n) = \frac{1}{2N}\sum_{n=1}^{N}\sum_{j=1}^{K} e_j(n)^2$$

$$\frac{\partial \mathcal{E}_j(n)}{\partial e_j(n)} = e_j(n)$$

$$\frac{\partial e_j(n)}{\partial y_j(n)} = -1 \qquad \frac{\partial y_j(n)}{\partial v_j(n)} = \phi'(v_j(n))$$

$$y_j(n) = \phi(v_j(n)) = \frac{1}{1+e^{-v_j(n)}} = \frac{u}{v}$$

$$\frac{\partial \phi(v_j(n))}{\partial v_j(n)} = \phi'(v_j(n)) = \frac{-e^{-v_j(n)} - 0}{\left[1+e^{-v_j(n)}\right]^2}$$

$$= \frac{-e^{-v_j(n)}}{\left(1+e^{-v_j(n)}\right)^2}$$

$$\phi(v_j(n)) - 1 = \frac{1}{1+e^{-v_j(n)}} - 1$$

$$= \frac{\cancel{1} - \cancel{1} - e^{-v_j(n)}}{1+e^{-v_j(n)}} = \frac{-e^{-v_j(n)}}{1+e^{-v_j(n)}}$$

$$\phi'(v_j(n)) = \frac{1}{1+e^{-v_j(n)}} \times \frac{-e^{-v_j(n)}}{1+e^{-v_j(n)}}$$

$$\phi'(v_j(n)) = \phi(v_j(n))\left[\phi(v_j(n)) - 1\right] \quad \text{—— } (*)$$

$$\frac{\partial y_j(n)}{\partial v_j(n)} = \phi(v_j(n))\left(\phi(v_j(n)) - 1\right)$$

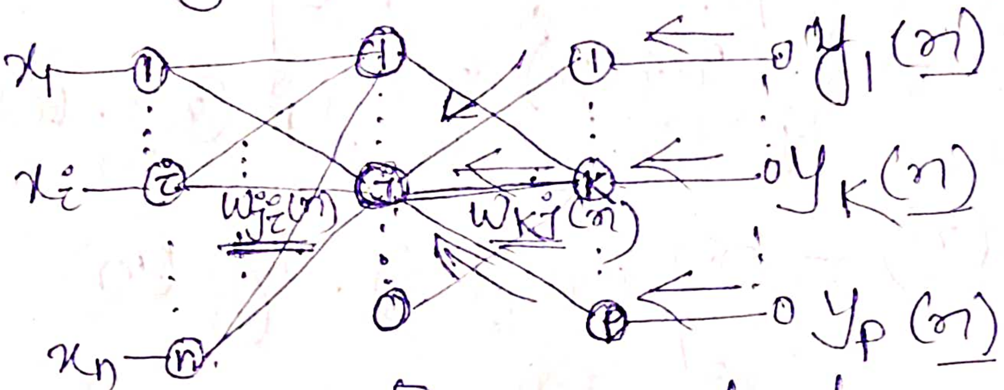
$$\Delta W_{ji}(n) = -e_j(n)\,\phi(v_j(n))\,[1-\phi(v_j(n))]\,y_i(n)$$

$$\underset{\text{new}}{W_{ji}(n+1)} = \underset{\text{old}}{W_{ji}(n)} + e_j(n)\,\phi(v_j(n))\,[1-\phi(v_j(n))]\,y_i(n)$$

$$\Downarrow \qquad \Downarrow$$

$$W_{ji}(n+1) = W_{ji}(n) + e_j(n)\,y_j(n)\,[1-y_j(n)]\,y_i(n)$$

Update weights in hidden layer



$$\frac{\partial \mathcal{L}_j(n)}{\partial W_{ji}(n)} = \Delta W_{ji}(n) = -\eta\,\delta_j(n)\,x_i(n)$$

$$\delta_j(n) = -\frac{\partial \mathcal{L}_j(n)}{\partial v_j(n)} \times \cancel{\frac{\partial v_j(n)}{\partial W_{ji}(n)}}$$

$$\delta_j(n) = -\underbrace{\frac{\partial \mathcal{L}_j(n)}{\partial v_j(n)}}_{\text{chain rule}} \times x_i(n)$$

$$\frac{\partial \mathcal{L}_j(n)}{\partial v_j(n)} = \underbrace{\frac{\partial \mathcal{L}_j(n)}{\partial e_K(n)} \times \frac{\partial e_K(n)}{\partial y_K(n)} \times \frac{\partial y_K(n)}{\partial v_K(n)}}_{\delta_K(n)} \times \frac{\partial v_K(n)}{\partial y_j(n)} \times \frac{\partial y_j(n)}{\partial v_j(n)}$$

$$\delta_K = \frac{\partial \mathcal{E}(n)}{\partial v_K(n)} = -e_K(n) \cdot \phi'(v_K(n))\, y_j(n)$$

$$\frac{\partial \mathcal{E}(n)}{\partial y_j(n)} = -\sum_{K=1}^{P} \delta_K(n)\, w_{Kj}(n)$$

$$\delta_j(n) = \phi_j'(v_j(n)) \cdot \sum_{K=1}^{P} -\delta_K(n)\, w_{Kj}(n)$$

$$\delta_j(n) = \phi_j'(v_j(n)) \sum_{K=1}^{P} \delta_K(n) \cdot w_{Kj}(n)$$

$$\Delta w_{ji}(n) = \eta \cdot \delta_j(n) \cdot y_i(n)$$

$$\boxed{\delta_K(n) = \sum_{K=1}^{P} e_K(n)\, \phi_K'(v_K(n))}$$

$e_1(n)\, y_1(n)\,[1 - y_1(n)]$

$\Rightarrow e_1(n)\, y_1(n)\,[1 - y_1(n)]\, w_{1j}(n) + e_2(n)\, y_2(n)\,[1 - y_2(n)]\, w_{2j}(n) + e_3(n)\, y_3(n)\,[1 - y_3(n)]\, w_{3j}(n)$

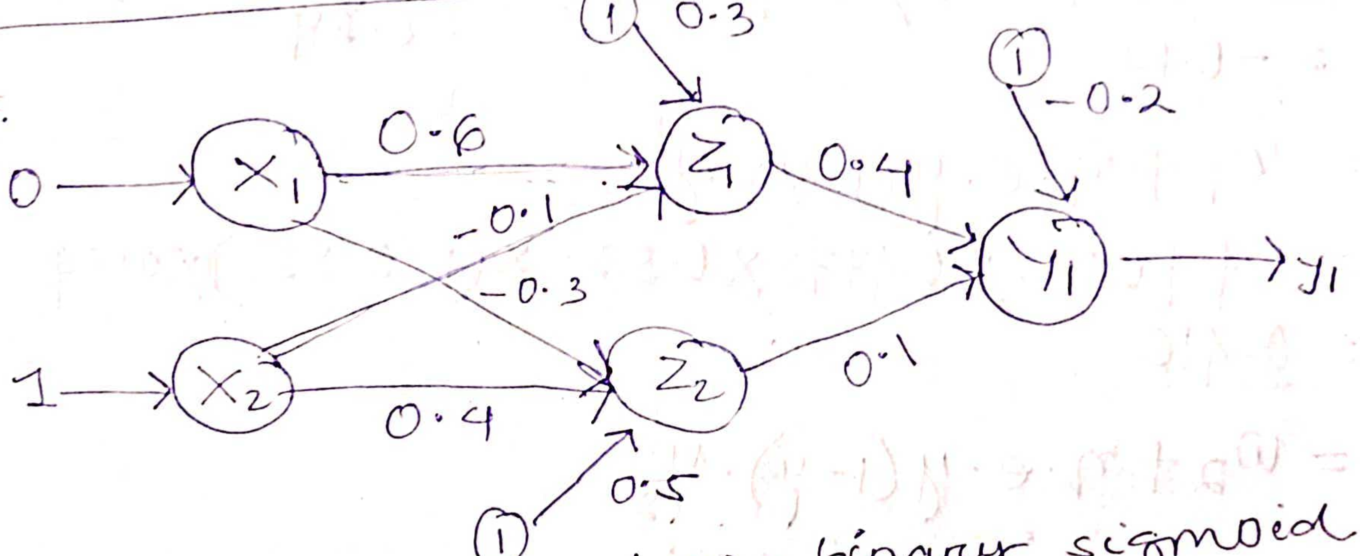
$\Rightarrow \sum \delta_K(n)\, w_{Kj}(n)$

$$\delta_j(n) = y_j(n)\,[1-y_j(n)]\sum_{k=1}^{p}\delta_k(n).\,w_{kj}(n)$$

$$\Delta w_{ji}(n) = \eta\,.\,\delta_j(n)\,.\,x_i(n)$$

$$\boxed{w_{ji}^{\text{new}}(n+1) = w_{ji}^{\text{old}}(n) + \eta\,\delta_j(n)\,.\,x_i(n)}$$

Q.



$\eta = 0.25$, $f(x) =$ binary sigmoid

<u>Forward computation</u>

$$z_1^h = 1\times0.3 + 0\times0.6 + 1\times(-0.1) = 0.2$$

$$y_1^h = \frac{1}{1+e^{-0.2}} = 0.54$$

$$z_2^h = 1\times0.5 - 0\times0.3 + 1\times0.4 = 0.9$$

$$y_2^h = \frac{1}{1+e^{-0.9}} = 0.71$$

$$z^o = -1\times0.2 + 0.4\times0.54 + 0.71\times0.1 = 0.087 = \underline{0.09}$$

$$y^o = \frac{1}{1+e^{-0.09}} = \underline{\underline{0.522}}$$

Backward computation

(i) Update output layer weights

$$e = d - y = 1 - 0.5227 = 0.4773$$

$$y(1-y) = 0.24$$

$$w_0 = w_0 + \eta \cdot e \cdot y(1-y)$$

$$= -0.2 + 0.25 \times 0.4773 \times \underbrace{0.522 \times (1-0.522)}_{0.24}$$

$$= -0.17$$

$$w_1 = w_1 + \eta \cdot e \cdot y(1-y) \cdot y_1^h$$

$$= 0.4 + 0.25 \times 0.4773 \times 0.522 \times (1-0.522) \times 0.54$$

$$= 0.416$$

$$w_2 = w_2 + \eta \cdot \underbrace{e \cdot y(1-y)}_{\delta_K = 0.114} \cdot y_2^h$$

$$= 0.1 + 0.25 \times 0.114 \times 0.71 = 0.12$$

(ii) Update hidden layer weights

$$w_0^h = w_0^h + \eta \cdot \delta_j^h \cdot x_i$$

$$\delta_j^h = y_j^h (1 - y_j^h) \cdot \sum_{K=1} \delta_K w_{Kj}$$

$\delta_1^h \Rightarrow$ Local gradient of first hidden neuron

$$\delta_1^h = y_1^h (1 - y_1^h) \cdot \delta^o \, w_{11}^o = 0.54 \times (1 - 0.54) \times 0.114 \times 0.4$$

$$= \cancel{0.602}\ 0.0117$$

$$w_{11}^h = w_{11}^h + \eta \cdot \delta_1^h \cdot x_1$$

$$w_{11}^h = 0.6 + 0.25 \times \overset{0.0117}{\cancel{0.602}} \times 0 = 0.6$$

$$w_{12}^h = -0.1 + 0.25 \times \overset{0.0117}{\cancel{0.602}} \times 1 = -0.09$$

$$w_{10}^{h} = w_{10}^{h} + \eta \cdot \delta_1^{h} \cdot 1$$
$$= 0.3 + 0.25 \times 0.0117 = \underline{\underline{0.3029}}$$

$$\delta_2^{h} = y_2^{h}(1 - y_2^{h}) \cdot \delta^{o} \cdot w_{12}$$
$$= 0.71(1 - 0.71) \times 0.114 \times 0.1 = 0.023$$

$$w_{20}^{h} = w_{20}^{h} + \eta \cdot \delta_2^{h} \cdot 1$$
$$= 0.5 + 0.25 \times 0.023 = \underline{\underline{0.50575}}$$

$$w_{21}^{h} = w_{21}^{h} + \eta \cdot \delta_2^{h} \cdot x_1 = \underline{\underline{-0.3}}$$

$$w_{22}^{h} = w_{22}^{h} + \eta \cdot \delta_2^{h} \cdot x_2 = 0.4 + 0.25 \times 0.023 \times 1$$
$$= \underline{\underline{0.40575}}$$