I.

we will use $\tanh(x) = \frac{e^{x} - e^{-x}}{e^{x} + e^{-x}}$ activation

case I: $\frac{\partial E_d}{\partial net_k}$ for output unit

$$= \frac{\partial E_d}{\partial o_k} \cdot \frac{\partial o_k}{\partial net_k}$$

$$\frac{\partial E_d}{\partial o_k} = \frac{1}{2} \sum_{k \in outputs} \frac{\partial}{\partial o_j} (t_k - o_k)^2$$

$$= \frac{1}{2} \sum_{k \in outputs} 2(t_k - o_k) \frac{\partial}{\partial o_k} (t_k - o_k)$$

$$= \sum_{k \in outputs} -(t_k - o_k)$$

$$\frac{\partial o_k}{\partial net_k} = I$$

$$\frac{\partial E_d}{\partial net_{jk}} = -(t_k - o_k)$$

lets say

$$-(t_k - o_k) = -\delta_k$$

we will call the partial of Error wrt net for any unit k as follows.

$$\frac{\partial E_d}{\partial net_j} = -\delta_k$$

$$\Delta w_{ji} = -\eta \frac{\partial E_d}{\partial w_{ki}}$$

$$\Delta w_{ji} = \eta \delta_k x_{ki}$$ where $\delta_k = (t_k - o_k)$

$$\boxed{\delta_o \leftarrow (t_k - o_k)}$$

colour

Case II : K is a hidden unit

$$\frac{\partial E_d}{\partial net_h} = \sum_{K \in DownStream(h)} \frac{\partial E_d}{\partial net_K} \frac{\partial net_K}{\partial net_h}$$

$$= \sum_{K \in DownStream(h)} -\delta_K \frac{\partial net_K}{\partial net_h} \frac{\partial O_h}{\partial net_h}$$

$$= \sum_{K \in Downstream_h} -\delta_K \, \omega_{Kh} \frac{\partial O_h}{\partial net_h}$$

Now, $\frac{\partial O_h}{\partial net_h} = (1 - O_h)^2 \rightarrow$ tanh instead of sigmoid

$$\therefore \frac{\partial E_d}{\partial net_h} = -(1 - O_h^2) \sum_{K \in Down\,stream_h} \omega_{Kh} \, \delta_K$$

$$\boxed{\delta_h \leftarrow (1 - O_h^2) \sum_{K \in Down\,stream_h} \omega_{Kh} \, \delta_K}$$

2. $O = \omega_0 + \omega_1 (x_1 + x_2)^2 +$

Ans $O = \omega_0 + \omega_1 x_1 + \omega_1 x_1^2 + \ldots$

$\Rightarrow$ First, Error function is defined as

$$E(\vec{\omega}) = \frac{1}{2} \sum_{d \in D} (t_d - O_d)^2$$

$\Rightarrow$ update rule is the same $\omega_i = \omega_i + \Delta\omega_i$

where $\Delta\omega_i = -\eta \frac{\partial E}{\partial \omega_i}$

$\Rightarrow$ For ω_0, $\frac{\partial E}{\partial \omega_0} = \frac{\partial}{\partial \omega_0} \frac{1}{2} \sum_{d \in D} (t_d - O_d)^2$

$$= \frac{1}{2} \sum_{d \in D} \frac{\partial}{\partial \omega_0} (t_d - O_d)^2$$

$$= \frac{1}{2} \sum_{d \in D} 2 (t_d - O_d) \frac{\partial}{\partial \omega_0} (t_d - O_d)$$

$$= \sum_{d \in D} (t_d - O_d)(-1)$$

$$= -\sum_{d \in D} (t_d - O_d)$$

$$\boxed{\text{So, } \Delta\omega_0 = \eta \sum_{d \in D} (t_d - O_d)}$$

$\Rightarrow$ For $\omega_1, \omega_2, \ldots \omega_n$

$$\frac{\partial E}{\partial \omega_i} = \frac{\partial}{\partial \omega_i} \frac{1}{2} \sum_{d \in D} (t_d - O_d)^2$$

$$= \frac{1}{2} \sum_{d \in D} \frac{\partial}{\partial \omega_i} (t_d - O_d)^2$$

$$= \frac{1}{2} \sum_{d \in D} 2 (t_d - O_d) \frac{\partial}{\partial \omega_i} (t_d - O_d)$$

$$= \sum_{d \in D} (t_d - O_d)(-(x_{id} + x_{id}^2))$$

$$\boxed{\text{So, } \Delta\omega_i = \eta \sum_{d \in D} (t_d - O_d)(x_{id} + x_{id}^2)}$$

3.

Node	Net	output
1	X_1	X_1
2	X_2	X_2
3	$net3 = \omega_{31}x_1 + \omega_{32}x_2$	$X_3 = F(net3)$
4	$net4 = \omega_{41}x_1 + \omega_{42}x_2$	$X_4 = F(net4)$
5	$net5 = \omega_{53}x_3 + \omega_{54}x_4$	$X_5 = F(net5)$

(a) O/P $Y5 = F(net5)$

$$= F(\omega_{53}x_3 + \omega_{54}x_4)$$

$$= F(\omega_{53}[F(\omega_{31}x_1 + \omega_{32}x_2)] + \omega_{54}[F(\omega_{41}x_1 + \omega_{42}x_2)])$$

(b) $H[\omega^{(2)}, H(\omega^{(1)} \cdot X)]$

(c) $h_1(x) = \dfrac{1}{1+e^{-x}}$

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

$h_2(x) = \dfrac{e^x - e^{-x}}{e^x + e^{-x}}$

$$= \frac{e^{2x}-1}{e^{2x}+1}$$

$$h_1(2x) = \frac{e^{2x}}{1+e^{2x}}$$

$$2h_1(2x) = \frac{2e^{2x}}{e^{2x}+1}$$

$$2h_1(2x) - 1 = \frac{2e^{2x}}{e^{2x}+1} - 1$$

$$= \frac{e^{2x}-1}{e^{2x}+1} = h_2(x)$$

$\therefore h_2(x) = 2h_1(2x) - 1$

$\therefore h_2(x)$ is a rescaled $h_1(x)$ function

→ here $h_1(x)$ & $h_2(x)$ is differing by linear transformations & constants.

→ $h_1(2x)$ multiply by 2 is linear transformation of $h_2(x)$, where substracting 1 from $2h_1(2x)$ is constant scaling of $h_2(x)$.

4.

$$E(\vec{w}) = \frac{1}{2} \sum_{d \in D} \sum_{k \in outputs} (t_{kd} - O_{kd})^2 + \gamma_{ij} \sum \omega_{ji}^2$$

$$\omega_{ji} \leftarrow \omega_{ji} + \Delta\omega_{ji}$$

$$\Delta\omega_{ji} = -\eta \frac{\partial E(\vec{\omega})}{\partial \omega_{ji}}$$

$$\frac{\partial E(\vec{\omega})}{\partial \omega_{ji}} = \underbrace{\frac{\partial}{\partial \omega_{ji}} \frac{1}{2} \sum_{d \in D} \sum_{k \in O} (t_{kd} - O_{kd})^2}_{1} + \underbrace{\gamma \frac{\partial}{\partial \omega_{ji}} \omega_{ji}^2}_{2}$$

The first term already derived in Back propagation Algo.

$$\therefore \frac{\partial E(\vec{\omega})}{\partial \omega_{ji}} = -(t_j - O_j) O_j (1 - O_j) x_{ji} + 2\gamma \omega_{ji}$$

$\Rightarrow$ For output nodes,

$$\omega_{ji} \leftarrow \omega_{ji} + \eta (t_j - O_j) O_j (1 - O_j) x_{ji} - 2\eta\gamma \omega_{ji}$$

$$\omega_{ji} \leftarrow (1 - 2\eta) \omega_{ji} + \eta (t_j - O_j) O_j (1 - O_j) x_{ji}$$

$$\omega_{ji} \leftarrow \beta \omega_{ji} + \eta \delta_j x_{ji}$$

where $\beta = 1 - 2n\gamma$ & $\delta_j = (t_j - O_j) O_j (1 - O_j)$

$\Rightarrow$ $\boxed{\text{For hidden layers, } \omega_{ji} \leftarrow \beta \omega_{ji} + \eta \delta_j x_{ji}}$

where $\beta = 1 - 2n\gamma$ & $\delta_j = O_j (1 - O_j) \sum_{k \in Downstream(j)} \delta_k \omega_{kj}$

$\Rightarrow$ So we have multiplier β in both derivatives.