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

$\Rightarrow$ Derivative of $\tanh(x) = \tanh'(x)$
$= 1 - \tanh^2(x)$

$\Rightarrow$ case 1: k is an output unit
we know that Error for data d is:

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

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

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

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

~~where $O_k = \tanh(net_k)$ and $net_j = \sum w_{ji} x_{ji}$~~

$\Rightarrow$ ~~Derivative of O_k = derivative of $\tanh(net_k)$~~

$\Rightarrow$ ~~$1 - \tanh^2 h(net_k)$~~

$\therefore$ ~~$1 - O_k^2$~~

So, putting this value in $\frac{\partial (t_k - O_k)}{\partial o_j}$

$$\therefore \; (0 - (1)\cancel{(1-O_k^2)}) = -\cancel{(1-O_k^2)} = -1$$

So, $$\frac{\partial E_d}{\partial net_k} = -(t_k - O_k)\cancel{(1-O_k^2)} = -\delta_k$$

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

$$= \eta (t_k - O_k)\cancel{(1-O_k^2)}\, x_{ki}$$

and $$\boxed{\delta_k \leftarrow (t_k - O_k)\cancel{(1-O_k^2)}}$$

$\Rightarrow$ case 2: k is a hidden unit

$$\frac{\partial E_d}{\partial net_h} = \sum_{k \in Downstream(h)} \frac{\partial E_d}{\partial net(k)} \cdot \frac{\partial net_k}{\partial net_h}$$

$$= \sum_{k \in Downstream(h)} -\delta_k \frac{\partial net_k}{\partial net_h}$$

$$= \sum_{k \in Downstream(h)} -\delta_k \frac{\partial net_k}{\partial O_h} \cdot \frac{\partial O_h}{\partial net_h}$$

where $\frac{\partial O_h}{\partial net_h} = (1 - O_h^2)$ as defined previously

$$\frac{\partial E_d}{\partial net_h} = \sum_k -\delta_k w_{kh} (1 - O_h^2)$$

and $$\boxed{\delta_h \leftarrow \sum_k w_{kh} \delta_k (1 - O_h^2)}$$

2. $o = w_0 + w_1(x_1 + x_1)^2 + \ldots w_n(x_n + x_n^2)$

Ans: $O = w_0 + w_1 x_1 + w_1 x_1^2 + \ldots w_n x_n + w_n x_n^2$

⇒ First, error function is defined as

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

⇒ update rule is the same $w_i = w_i + \Delta w_i$

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

⇒ for w_0, $\frac{\partial E}{\partial w_0} = \frac{\partial}{\partial w_0} \frac{1}{2} \sum_{d \in D} (t_d - o_d)^2$

$$= \frac{1}{2} \sum_{d \in D} \frac{\partial}{\partial w_0} (t_d - o_d)^2$$

$$= \frac{1}{2} \sum_{d \in D} 2 (t_d - o_d) \frac{\partial}{\partial w_0} (t_d - o_d)$$

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

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

so, $\Delta w_0 = \eta \sum_{d \in D} (t_d - o_d)$

⇒ for $w_1, w_2 \ldots w_n$

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

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

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

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

so, $\Delta w_i = \eta \sum_{d \in D} (t_d - o_d)(x_{id} + x_{id}^2)$

3.9

Node	Net	Output
1	x_1	x_1
2	x_2	x_2
3	$net_3 = w_{31}x_1 + w_{32}x_2$	$x_3 = f(net_3)$
4	$net_4 = w_{41}x_1 + w_{42}x_2$	$x_4 = f(net_4)$
5	$net_5 = w_{53}x_3 + w_{54}x_4$	$x_5 = f(Net_5)$

a. $y_5 = f(net_5)$

$= f(w_{53}x_3 + w_{54}x_4)$

$= f(w_{53}[f(w_{31}x_1 + w_{32}x_2)]$
$+ w_{54}[f(w_{41}x_1 + w_{42}x_2)]$

b. $H[w^{(2)}, H(w^{(1)} \cdot x)]$

c. $h_1(x) = \dfrac{1}{1+e^{-x}}$ $\qquad$ $h_2(x) = \dfrac{e^{x} - e^{-x}}{e^{x} + e^{-x}}$

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

($\because$ multiply both eqn by e^{x})

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

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

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

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

$= h_2(x)$

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

$h_2(x)$ is a rescalled function of $h_1(x)$

- $h_1(x)$ & $h_2(x)$ is different in terms of linear transformations & constants.

- $h_1(2x) * 2$ is a linear transformation of $h_2(x)$ where subtracting 1 from $2h_1(2x)$ is a constant scaling of $h_2(x)$

4. $$E(\vec{w}) = \frac{1}{2} \sum_{d \in D} \sum_{k \in outputs} (t_{kd} - O_{kd})^2 + \gamma \sum_{i,j} w_{ji}^2$$

$\Rightarrow$ $w_{ji} \leftarrow w_{ji} + \Delta w_{ji}$

$$\Delta w_{ji} = -\eta \frac{\partial E(\vec{w})}{\partial w_{ji}}$$

$$\frac{\partial E(\vec{w})}{\partial w_{ji}} = \underbrace{\frac{\partial}{\partial w_{ji}} \frac{1}{2} \sum_{d \in D} \sum_{k \in outputs} (t_{kd} - O_{kd})^2}_{1} + \underbrace{\frac{\partial}{\partial w_{ji}} \gamma \sum_{i,j} w_{ji}^2}_{2}$$

We defined derivative of 1^{st} Term previously

It would be $= -(t_j - O_j) O_j (1 - O_j) x_{ji}$

$\Rightarrow$ We will continue to work on 2^{nd} Term.

$$\frac{\partial}{\partial w_{ji}} \gamma \sum_{i,j} w_{ji}^2 = 2\gamma w_{ji}$$

This is for output nodes

So, $\Delta w_{ji} = \eta (t_j - O_j) O_j (1 - O_j) x_{ji} - 2\eta\gamma w_{ji}$

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

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

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

where $\beta = (1 - 2\eta)$

$\delta_j = (t_j - O_j) O_j (1 - O_j)$

$\Rightarrow$ For Hidden units,

$$\delta_j = O_j (1 - O_j) \sum_{k \in Downstream(j)} \delta_k w_{kj}$$

So, $w_{ji} \leftarrow \beta w_{ji} + \eta \delta_j x_{ji}$

$\Rightarrow$ These both derivation shows that we have multiplied β (constant) to the weight (w_{ji})