1. for the output layer neuron uses the identity function $f(x)=x$

$E_d(W) = \frac{1}{2}\sum_{k\in outputs}(t_k - O_k)^2$ $\quad \Delta W_{ji} = -\eta \frac{\partial E_d}{\partial W_{ji}}$

$\frac{\partial E_d}{\partial W_{ji}} = \frac{\partial E_d}{\partial net_j} \cdot \frac{\partial net_j}{\partial W_{ji}}$ $\quad \frac{\partial net_j}{\partial W_{ji}} = x_{ji}$

$\frac{\partial E_d}{\partial net} = \frac{\partial E_d}{\partial O_j} \cdot \frac{\partial O_j}{\partial net_j}$ $\quad \frac{\partial E_d}{\partial O_j} = \frac{\partial}{\partial O_j}\left[\frac{1}{2}\sum_{k\in outputs}(t_k - O_k)^2\right] = \frac{\partial}{\partial O_j}\left[\frac{1}{2}(t_j - O_j)^2\right]$

$= -(t_j - O_j)$ $\quad \frac{\partial O_j}{\partial net_j} = 1$ $\quad \frac{\partial E_d}{\partial O_j} \cdot \frac{\partial O_j}{\partial net_j} = -(t_j - O_j)$ $\quad \Delta W_{ji} = \eta (t_j - O_j) x_{ji}$

for hidden layer uses tanh function $f(x) = \tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$

$\frac{\partial E_d}{\partial W_{ji}} = \frac{\partial E_d}{\partial net_j} \cdot \frac{\partial net_j}{\partial W_{ji}}$ $\quad \frac{\partial net_j}{\partial W_{ji}} = x_{ji}$

$\frac{\partial E_d}{\partial net_j} = \sum_{k\in Downstream(j)} \frac{\partial E_d}{\partial net_k} \cdot \frac{\partial net_k}{\partial net_j} = \sum_{k\in Downstream(j)} -\delta_k \cdot \frac{\partial net_k}{\partial net_j} = \sum_{k\in Downstream(j)} \frac{\partial net_k}{\partial O_j} \cdot \frac{\partial O_j}{\partial net_j}$

$= \sum_{k\in Downstream(j)} -\delta_k W_{kj} \cdot \frac{\partial O_j}{\partial net_j} = \sum_{k\in Downstream(j)} -\delta_k W_{kj}(1 - O_j^2)$ $\quad \delta_j = (1 - O_j^2)\sum_{k\in Downstream(j)} \delta_k W_{kj}$

$\Delta W_{ji} = \eta \delta_j x_{ji}$

2. $O = W_0 + W_1(x_1 + x_1^2) + \cdots + W_n(x_n + x_n^2)$

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

$= \frac{1}{2}\sum_{d\in D}(t_d - O_d)\frac{\partial}{\partial W_i}(t_d - O_d) = \sum_{d\in D}(t_d - O_d)(-(x_{id} + x_{id}^2)) = \sum_{d\in D}(t_d - O_d)(-x_{id} - x_{id}^2)$

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

3. a. $net_3 = x_1 \cdot W_{31} + x_2 \cdot W_{32}$ $\quad net_4 = x_1 \cdot W_{41} + x_2 \cdot W_{42}$

$net(y_5) = h(net_3)\cdot W_{53} + h(net_4)\cdot W_{54}$ $\quad O = h(net(y_5))$

b. $output = h(h(X \cdot W^{(1)}) \cdot W^{(2)})$

c. $h_1(x) = \frac{1}{1+e^{-x}} = \frac{e^x}{1+e^x}$ $\quad h_2(x) = \tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$

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

$output = h_1(h_1(net_3)\cdot W_{53} + h_1(net_4)\cdot W_{54})$

$output = h_2(h_2(net_3) \cdot W_{53} + h_2(net_4) \cdot W_{54}) = 2h_1(2h_2(net_3) W_{53} + 2h_2(net_4) W_{54})$

$\dashv$

4. $W_{ji} = W_{ji} + \Delta W_{ji}$ $\quad \Delta W_{ji} = -\eta \frac{\partial E_d}{\partial W_{ji}}$

for output layer $\frac{\partial E_d}{\partial W_{ji}} = \frac{\partial}{\partial W_{ji}} \frac{1}{2} \sum_{d \in D} \sum_{k \in outputs} (t_{kd} - O_{kd})^2 + \frac{\partial}{\partial W_{ji}} r W_{ji}^2$

$= -(t_j - O_j) O_j (1 - O_j) x_{ji} + 2r W_{ji}$ $\quad \Delta W_{ji} = \eta (t_j - O_j) O_j (1 - O_j) x_{ji} - 2r\eta W_{ji}$

$W_{ji} = W_{ji} + \Delta W_{ji} = (1 - 2r\eta) W_{ji} + \eta (t_j - O_j) O_j (1 - O_j) x_{ji}$

for hidden layer. $\delta_j = O_j (1 - O_j) \sum_{k \in Downstream(j)} \delta_k W_{kj}$ $\quad \Delta W_{ji} = \eta \delta_j x_{ji} - 2r\eta W_{ji}$

$W_{ji} = W_{ji} + \Delta W_{ji} = (1 - 2r\eta) W_{ji} + \eta \delta_j x_{ji}$