1.1

a. tanh

$$\tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$$

$$\frac{\partial \tanh(x)}{\partial x} = \frac{(e^x - e^{-x})'(e^x + e^{-x}) - (e^x - e^{-x})(e^x + e^{-x})'}{(e^x + e^{-x})^2}$$

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

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

$$\Delta w = -\eta \frac{\partial E_d}{\partial w_{ji}} = -\eta \left(\frac{\partial E_d}{\partial net_j} \cdot \frac{\partial net_j}{\partial w_{ji}} \right) = -\eta \frac{\partial E_d}{\partial net_j} \cdot x_{ji}$$

For output: $\frac{\partial E_d}{\partial net_j} = \frac{\partial E_d}{\partial O_j} \cdot \frac{\partial O_j}{\partial net_j} = \frac{\partial}{\partial O_j} [\frac{1}{2} (t_j - O_j)^2] \cdot (1 - O_j^2)$

$$= -(t_j - O_j)(1 - O_j^2)$$

$$\Delta w_{ji} = \eta (t_j - O_j)(1 - O_j^2) \cdot x_{ji}$$

For hidden layer:

$$\frac{\partial E_d}{\partial net_j} = \Sigma -\delta_k \frac{\partial net_k}{\partial net_j} = \Sigma -\delta_k \frac{\partial net_k}{\partial O_j} \cdot \frac{\partial O_j}{\partial net_j}$$

$$\Delta w_{ji} = \eta x_{ji} (1 - O_j^2) \Sigma \delta_k w_{kj}$$

b. Relu

$$\mathrm{Relu}(x) = \max(0, x)$$

$$\frac{\partial \mathrm{Relu}(x)}{\partial x} = \begin{cases} 0 & x \le 0 \\ 1 & x > 0 \end{cases}$$

$$\Delta W_{ji} = -\eta \frac{\partial E_d}{\partial W_{ji}} = -\eta \frac{\partial E_d}{\partial net_j} \cdot x_{ji}$$

For output unit:

$$\frac{\partial E_d}{\partial net_j} = \begin{cases} 0 & x \le 0 \\ -(t_j - O_j) & x > 0 \end{cases}$$

$$\Delta W_{ji} = \begin{cases} 0 & x \le 0 \\ \eta (t_j - O_j) x_{ji} & x > 0 \end{cases}$$

For hidden layer:

$$\frac{\partial E_d}{\partial net_j} = \begin{cases} 0 & x \le 0 \\ \sum \delta_k W_{kj} & x > 0 \end{cases}$$

$$\Delta W_{ji} = \begin{cases} 0 & x \le 0 \\ \eta x_{ji} \sum \delta_k W_{kj} & x > 0 \end{cases}$$

1.2

$$E = \sum_{x \in X} \frac{1}{2} (t_x - O_x)^2$$

$$\frac{\partial E}{\partial W_i} = \frac{\partial}{\partial W_i} \sum_{x \in X} \frac{1}{2} (t_x - O_x)^2$$

$$= \frac{1}{2} \sum_{x \in X} 2 (t_x - O_x) \cdot \frac{\partial (t_x - O_x)}{\partial W_i}$$

$$(O_x = W_0 + W_1 X_{1x} + W_1 X_{1x}^2 + \dots + W_n X_{nx} + W_n X_{nx}^2)$$

$$= \sum_{x \in X} (t_x - O_x) \cdot (-1) \cdot (X_{ix} + X_{ix}^2)$$

$$= \sum_{x \in X} (t_x - O_x) \cdot (-X_{ix} - X_{ix}^2)$$

1.3
a.

$net_3 = w_{31}x_1 + w_{32}x_2$ $\qquad x_3 = h(net_3)$

$net_4 = w_{41}x_1 + w_{42}x_2$ $\qquad x_4 = h(net_4)$

$net_5 = w_{53}x_3 + w_{54}x_4$ $\qquad y_5 = h(net_5)$

$$= h(w_{53}h(net_3) + w_{54}h(net_4))$$
$$= h(w_{53}h(w_{31}x_1 + w_{32}x_2) + w_{54}h(w_{41}x_1 + w_{42}x_2))$$

b. The vector format is:

$y_5 = h(w^{(2)} \cdot h(w^{(1)} \cdot \mathbf{x}))$

c. $h_t(x) = \dfrac{e^x - e^{-x}}{e^x + e^{-x}}$

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

if we set $u = -2x$, then $2 \cdot \dfrac{1}{e^{2x}+1} = 2 \cdot \dfrac{1}{e^{-u}+1}$

so, $1 - h_t(x) = 2 \cdot h_s(u)$ $(u = -2x)$

$1 - h_t(x)$ and $2 \cdot h_s(u)$ output the same function.

1.4

$w_{ji} = w_{ji} + \Delta w_{ji}$

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

$\Rightarrow \dfrac{\partial E_d}{\partial w_{ji}} = \dfrac{\partial}{\partial w_{ji}} \left[\frac{1}{2}\sum\sum(t_{kd} - O_{kd})^2\right] + \dfrac{\partial}{\partial w_{ji}}(\gamma \sum w_{ji}^2)$

Case I (output unit):

$\dfrac{\partial E_d}{\partial w_{ji}} = -(t_j - O_j)O_j(1 - O_j)x_{ji} + 2\gamma w_{ji}$

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

$$= \underbrace{(1 - 2\eta\gamma)}_{\alpha} w_{ji} + \eta\underbrace{[(t_j - O_j)O_j(1 - O_j]}_{\delta_j} x_{ji}$$

case II (hidden layer):

$$\frac{\partial E_d}{\partial w_{ji}} = -O_j(1-O_j)\sum \delta_k w_{kj}\, x_{ji} + 2\gamma w_{ji}$$

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

$$= w_{ji} + \eta O_j(1-O_j)\sum \delta_k w_{kj}\, x_{ji} - 2\gamma\eta w_{ji}$$

$$= \underbrace{(1-2\eta\gamma)}_{\alpha} w_{ji} + \eta\,[\underbrace{O_j(1-O_j)\sum \delta_k w_{kj}}_{\delta_j}]\, x_{ji}$$