1.1

a. $\tanh(x) = \dfrac{e^{x} - e^{x}}{e^{x} + e^{-x}}$

$$\tanh(x) = \frac{\sinh(x)}{\cosh(x)}$$

$$= \frac{\cosh(x)\cosh(x) - \sinh(x)\sinh(x)}{\cosh(x)^2} = \frac{\cosh(x)^2 - \sinh(x)^2}{\cosh(x)^2}$$

$$= 1 - \frac{\sinh(x)^2}{\cosh(x)^2} = 1 - \tanh^2(x)$$

Case I: j is an output layer

$$\delta_j = (t_j - o_j)(1 - o_j^2)$$

Case II: j is hidden layer node

$$\delta_j = (1 - o_j^2) \sum_{k \in Downstream(j)} \delta_k w_k$$

b. $ReLU(x) = \max(0, x)$

$$R = \begin{cases} x & x > 0 \\ 0 & x \leq 0 \end{cases} \qquad R' = \begin{cases} 1 & z > 0 \\ 0 & z < 0 \end{cases}$$

Case I: j is an output layer

$$\delta_j = (t_j - o_j)(R')$$

Case II: j is a hidden layer node

$$\delta_j = (R') \sum_{k \in Downstream(j)} \delta_k w_k$$

1.2

$$O = \sum w_i x_i$$

$$O = w_0 + w_1(x_1 + x_1^2) + \dots + w_n(x_n + x_n^2)$$

$$E = \frac{1}{2}(t - O)^2$$

$$\frac{\partial o}{\partial w_i} = \frac{1}{2}\left(t - \left(w_0 + w_1(x_1 + x_1^2) + \dots + w_n(x_n + x_n^2)\right)\right)^2$$

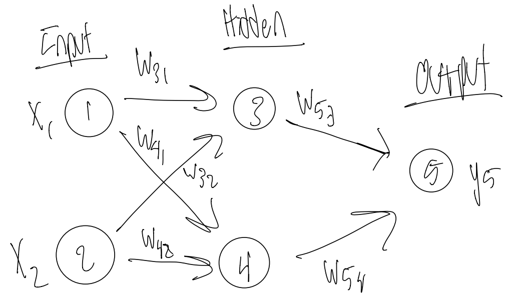
$$= -x_n - x_n^2$$

$$\frac{\partial E}{\partial w_i} = \frac{1}{2} \cdot 2(t - o)\left(\frac{\partial o}{\partial w_i}\right)$$

$$w_i^n = w_i^0 - n\left(\frac{\partial E}{\partial w_i}\right)$$

$$= w_i^0 - n(t - o)(-x_n - x_n^2)$$

1.3



a. in $1 = x_1$ out $1 = x_1$

in $2 = x_2$ out $2 = x_2$

in $3 = w_{31}x_1 + w_{32}x_2$ out $3 = h(w_{31}x_1 + w_{32}x_2)$

in $4 = w_{41}x_1 + w_{42}x_2$ out $4 = h(w_{41}x_1 + w_{42}x_2)$

in $5 = w_{53}(h(w_{31}x_1 + w_{32}x_2)) + w_{54}(h(w_{41}x_1 + w_{42}x_2))$

$y_5 = h((w_{53}(h(w_{31}x_1 + w_{32}x_2)) + w_{54}(h(w_{41}x_1 + w_{42}x_2)))$

$X = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$

b. $W^{(1)} = \begin{pmatrix} w_{31} & w_{32} \\ w_{41} & w_{42} \end{pmatrix}$

$W^2 = \begin{pmatrix} w_{53} & w_{54} \end{pmatrix}$

Input → hidden $= h(W' \cdot X)$

$= \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \cdot \begin{pmatrix} w_{31} & w_{32} \\ w_{41} & w_{42} \end{pmatrix} = \begin{pmatrix} x_1 w_{31} + x_2 w_{41} \\ x_1 w_{32} + x_2 w_{42} \end{pmatrix}$

Hidden → output $= h(W^2 \cdot h(W' \cdot X))$

$= \begin{pmatrix} w_{53} & w_{54} \end{pmatrix} \cdot \begin{pmatrix} x_1 w_{31} + x_2 w_{41} \\ x_1 w_{32} + x_2 w_{42} \end{pmatrix} = \begin{pmatrix} x_1 w_{31} w_{53} + x_2 w_{41} w_{53} + x_1 w_{32} w_{54} + x_2 w_{42} w_{54} \end{pmatrix}$

$y_5 = h(W^2 \cdot h(W' \cdot X))$

c. $h_s(x) = \frac{1}{1+e^{-x}}$

Sigmoid rearranged

$1 - \sigma(x) = \sigma(-x) \rightarrow 1 - \frac{1}{1+e^{-x}} = \frac{1}{1+e^{x}}$

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

↳ 2 * sigmoid

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

$= 1 - 2(1 - \sigma(2x))$

$= 1 - 2 + 2\sigma(2x)$

$= 2\sigma(2x) - 1$

$h_t(x) = 2h_s(2x) - 1$ => tanh is a rescaled sigmoid function so they both generate the same output function

1.4

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

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

$$= \sum_{d \in D} \sum_{k \in outputs} (t_{kd} - O_{kd}) + \frac{\partial E}{\partial w_{ji}} y \sum_{ij} w_{ji}^2$$

$$= \sum_{d \in D} \sum_{k \in outputs} (t_{kd} - O_{kd}) + 2y \sum_{ij} w_{ji}$$

$$= \sum_{d \in D} \sum_{k \in outputs} (t_{kd} - O_{kd}) + 2y\eta w_{ji}$$

$$= (t_{kd} - O_{kd})\, 2y\eta\delta w_{ji}$$

$$\delta_k \leftarrow (t_k - O_k)(2(t_k - O_k))$$

$$\delta_h \leftarrow (t_h - O_h) \sum w_{kh} \delta_k$$

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

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