6.1 Since we are evaluating just one output neuron o, we can drop all of the 2nd degree "j" indices. ~~i~~

general rule for updating a single weight:

$$w_i \leftarrow w_i + \Delta w_i$$

t is the "target" or y

assuming sigmoid activation

where $\Delta w_i = -\eta \dfrac{\partial E}{\partial w_i} = \eta (t - o) o (1 - o) x_i$

we say $\delta_o = (t - o) o (1-o)$ since this stays the same for all weights for any given training example,

and $\Delta w_i = \eta \delta_o x_i$

1.2

a. $\boxed{3} = h(x_1) \cdot w_{31} + h(x_2) \cdot w_{32}$

$\boxed{4} = h(x_1) \cdot w_{41} + h(x_2) \cdot w_{42}$

$y_5 = \boxed{5} = h\left(\boxed{3}\right) \cdot w_{53} + h\left(\boxed{4}\right) \cdot w_{54}$

$= h\left(h(x_1) \cdot w_{31} + h(x_2) \cdot w_{32}\right) \cdot w_{53} +$

$\left(h(x_1) \cdot w_{41} + h(x_2) \cdot w_{42}\right) \cdot w_{54}$

1.2 b

$$\begin{bmatrix} ③ \\ ④ \end{bmatrix} = \begin{bmatrix} w_{31} & w_{32} \\ w_{41} & w_{42} \end{bmatrix} h \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = W^{(1)} h(X)$$

$$⑤ = \begin{bmatrix} w_{53} & w_{54} \end{bmatrix} h \begin{bmatrix} ③ \\ ④ \end{bmatrix} = W^{(2)} h(W^{(1)} h(X))$$

c. —

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

$\sigma(x)$ is symmetric about the origin

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

$$\Rightarrow 1 - \frac{1}{1+e^{-x}} = \frac{1}{1+e^{x}} \qquad \text{, remember this}$$

$$\tanh(x) = \frac{e^x - e^{-x} + e^{-x} - e^{-x}}{e^x + e^{-x}} \qquad \text{add } e^{-x} - e^{-x} \text{ to numerator}$$

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

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

$\hookrightarrow$

1.2 c continued

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

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

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

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

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

thus tanh is just a scaled/shifted sigmoid function, and can therefore generate the same function.