John Kenney
jfk180030

1.

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

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

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

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

$$1 - (\tanh(x))^2$$

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

$$\omega_{ji}^{n} = \omega_{ji}^{old} + \boxed{\Delta\omega_{ji}} \qquad \Delta\omega_{ji} = -\eta\frac{\partial E_d}{\partial \omega_{ji}}$$

$$\frac{\partial E_d}{\partial \omega_{ji}} = \frac{\partial E_d}{\partial net_j}\frac{\partial net_j}{\partial \omega_{ji}} \qquad \frac{\partial net_j}{\partial \omega_{ji}} = \left(\sum x_{ji}\omega_{ji}\right)' = x_{ji}$$

$$\frac{\partial E_d}{\partial net_j}$$

Case 1: j is an output unit $\quad o_j = \tanh(net_j)$

$$\frac{\partial E_d}{\partial net_j} = \frac{\partial E_d}{\partial o_j}\frac{\partial o_j}{\partial net_j} \qquad \frac{\partial o_j}{\partial net_j} = 1 - (o_j)^2$$

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

$$= \frac{\partial}{\partial o_j}\left[\frac{1}{2}(t_j - o_j)^2\right]$$

$$= -(t_j - o_j)$$

$$\frac{\partial E_d}{\partial net_j} = -(t_j - o_j)(1 - o_j^2) = -\delta_j$$

$$\Delta\omega_{ji} = -\eta\frac{\partial E_d}{\omega_{ji}} = \eta\,\delta_j\,x_{ji}$$

Case 2: j is a hidden unit:

$$\frac{\partial E_d}{\partial net_j} = \sum_{k \in downstream(j)} \frac{\partial E_d}{\partial net_k} \frac{\partial net_k}{\partial net_j}$$

$$= \sum_{k \in downstream(j)} -\delta_k \frac{\partial net_k}{\partial net_j} \qquad \left( O_j = \tanh(net_j) \right.$$

$$= \sum_{k \in downstream(j)} -\delta_k \frac{\partial net_k}{\partial o_j} \frac{\partial o_j}{\partial net_j}$$

$$= \sum_{k \in downstream(j)} -\delta_k w_{kj} (1 - o_j^2)$$

$$-\delta_j = (1 - o_j^2) \sum_{k \in downstream(j)} -\delta_k w_{kj}$$

$$\Delta w_{ji} = -n \frac{\partial E_d}{\partial w_{ji}} = n \delta_j x_{ji}$$

where $o_j = \tanh(net_j)$

Summary:

Case 1: j is output layer node:

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

$$w_{ji}^{new} = w_{ji}^{old} + \Delta w_{ji} \quad \text{where } \Delta w_{ji} = n \delta_j x_{ji}$$

Case 2: j is hidden layer node:

$$-\delta_j = (1 - o_j^2) \sum_{k \in downstream(j)} -\delta_k w_{kj}$$

$$w_{ji}^{new} = w_{ji}^{old} + \Delta w_{ji}$$

where $\Delta w_{ji} = n \delta_j x_{ji}$

1\. b.

$$ReLu(x) = \begin{cases} x & ; x \in (0, \infty) \\ 0 & ; x \in (-\infty, 0] \end{cases}$$

$$\frac{d\,ReLu(x)}{dx} = \begin{cases} 1 & ; x \in (0, \infty) \\ 0 & ; x \in (-\infty, 0] \end{cases}$$

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

$$w_{ji}^{new} = w_{ji}^{old} + \Delta w_{ji} \qquad \Delta w_{ji} = -n \frac{\partial E_d}{\partial w_{ji}}$$

$$\frac{\partial E_d}{\partial w_{ji}} = \frac{\partial E_d}{\partial net_j} \frac{\partial net_j}{\partial w_{ji}} \qquad \frac{\partial net_j}{\partial w_{ji}} = (\Sigma x_{ji} w_{ji})' = x_{ji}$$

$$\frac{\partial E_d}{\partial net_j}$$

Case 1: j is an output unit $\qquad o_j = ReLu(net_j)$

$$\frac{\partial E_d}{\partial net_j} = \frac{\partial E_d}{\partial o_j} \frac{\partial o_j}{\partial net_j} \qquad \frac{\partial o_j}{\partial net_j} = \begin{cases} 1 & ; net_j \in (0, \infty) \\ 0 & ; net_j \in (-\infty, 0] \end{cases}$$

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

$$= \frac{\partial}{\partial o_j}\left[\frac{1}{2}(t_j - o_j)^2\right]$$

$$= -(t_j - o_j)$$

$$\frac{\partial E_d}{\partial net_j} = \begin{cases} -(t_j - o_j) & ; net_j \in (0, \infty) \\ 0 & , net_j \in (-\infty, 0] \end{cases}$$

$$= -\delta_j$$

$$\Delta w_{ji} = -n \frac{\partial E_d}{\partial w_{ji}} = n \delta_j x_{ji}$$

Case 2: j is a hidden unit:

$O_j = ReLu(net_j)$

$$\frac{\partial E_d}{\partial net_j} = \sum_{k \in downstream(j)} \frac{\partial E_d}{\partial net_k} \frac{\partial net_k}{\partial net_j}$$

$$= \sum_{k \in downstream(j)} -\delta_k \frac{\partial net_k}{\partial net_j}$$

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

$$-\delta_j = \begin{cases} \sum_{k \in downstream(j)} -\delta_k w_{kj} & ; net_j \in (0, \infty) \\ 0 & ; net_j \in (-\infty, 0] \end{cases}$$

$$\Delta w_{ji} = -n \frac{\partial E_d}{\partial w_{ji}} = n \delta_j x_{ji}$$

Summary: where $O_j = ReLu(net_j)$

Case 1: j is output layer node:

$$\delta_j = \begin{cases} (t_j - O_j) & , net_j \in (0, \infty) \\ 0 & , net_j \in (-\infty, 0] \end{cases}$$

$$w_{ji}^{new} = w_{ji}^{old} + \Delta w_{ji} \quad \text{where } \Delta w_{ji} = n \delta_j x_{ji}$$

Case 2: j is hidden layer node:

$$-\delta_j = \begin{cases} \sum_{k \in downstream(j)} -\delta_k w_{kj} & ; net_j \in (0, \infty) \\ 0 & ; net_j \in (-\infty, 0] \end{cases}$$

$$w_{ji}^{new} = w_{ji}^{old} + \Delta w_{ji} \quad \text{where } \Delta w_{ji} = n \delta_j x_{ji}$$

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

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

$w_i^{new} = w_i^{old} + \Delta w_i \qquad \Delta w_j = -n\left(\frac{\partial E}{\partial w_i}\right)$

$$\frac{\partial E}{\partial w_i} = \frac{\partial E}{\partial(t-o)} \frac{\partial(t-o)}{\partial w_i}$$

$$\frac{\partial E}{\partial(t-o)} = (t-o)$$

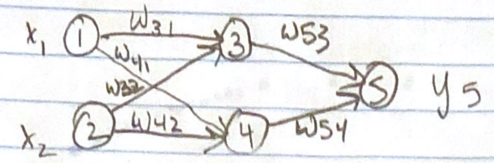
$$\frac{\partial(t-o)}{\partial w_i} = \frac{-\partial o}{\partial w_i} = \frac{-d(w_i(x_i + x_i^2))}{\partial w_i} = -(x_i + x_i^2)$$

$$\Delta w_j = -n\frac{\partial E}{\partial w_i}$$

$$\frac{\partial E}{\partial w_i} = -(t-o)(x_i + x_i^2) \qquad \Delta w_j = n(t-o)(x_i + x_i^2)$$

$$w_i^{new} = w_i^{old} + \Delta w_j$$

3.




a)
$$x_3 = w_{31}x_1 + w_{32}x_2$$
$$x_4 = w_{41}x_1 + w_{42}x_2$$
$$y_3 = h(x_3)$$
$$y_4 = h(x_4)$$
$$x_5 = w_{53}y_3 + w_{54}y_4$$
$$y_5 = h(x_5)$$

b)
$$\begin{pmatrix} x_3 \\ x_4 \end{pmatrix} = W^{(1)} X$$
$$\begin{pmatrix} y_3 \\ y_4 \end{pmatrix} = h(W^{(1)}X)$$
$$x_5 = W^{(2)} h(W^{(1)}X)$$
$$y_5 = h(W^{(2)} h(W^{(1)}X))$$

c) sigmoid $h = \sigma$
tanh $h = t$

$$y_{5\sigma} = A\, y_{5t} + B$$
$$\frac{1}{1+e^{-x_5}} = A\,\frac{e^{x_5} - e^{-x_5}}{e^{x_5} + e^{-x_5}} + B$$
$$\frac{1}{1+e^{-x_5}} = \frac{1}{1+e^{-x_5}} \cdot \frac{e^{x_5} - e^{-x_5}}{e^{x_5} + e^{-x_5}} + \frac{2e^{-x_5}}{(1+e^{-x_5})(e^{x_5} + e^{-x_5})}$$
$$\frac{1}{1+e^{-x_5}} = \frac{(e^{x_5} + e^{-x_5})}{(1+e^{-x_5})(e^{x_5} + e^{-x_5})}$$
$$\frac{1}{1+e^{-x_5}} = \frac{1}{1+e^{-x_5}}$$
$$A = \frac{1}{1+e^{-x_5}} \qquad B = \frac{2e^{-x_5}}{(1+e^{-x_5})(e^{x_5} + e^{-x_5})}$$

$$\frac{\partial o_j}{\partial w_{ji}} = \frac{\partial O(net_j)}{\partial w_{ji}}$$

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

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

$$\frac{\partial E_d}{\partial w_{ji}} = \frac{\partial}{\partial w_{ji}}\left[\frac{1}{2}\sum_{d\in D}\sum_{k\in outputs}(t_{kd} - o_{kd})^2 + \gamma\sum_{i,j} w_{ji}^2\right]$$

$$\frac{\partial E_d}{\partial w_{ji}} = \frac{\partial}{\partial w_{ji}}\left[\frac{1}{2}\sum_{d\in D}\sum_{k\in outputs}(t_{kd} - o_{kd})^2\right] + \frac{\partial}{\partial w_{ji}}\left[\gamma\sum_{i,j} w_{ji}^2\right]$$

$$= \frac{\partial}{\partial net_j}\left[\frac{1}{2}\sum_{d\in D}\sum_{k\in outputs}(t_{kd} - o_{kd})^2\right]\cdot\frac{\partial net_j}{\partial w_{ji}} + \gamma\sum_{ij}\frac{\partial}{\partial w_{ji}}(w_{ji}^2)$$

$$\frac{\partial E_d}{\partial net_j} = \frac{\partial E_d}{\partial o_j}\frac{\partial o_j}{\partial net_j}$$

$$= \frac{\partial}{\partial o_j}\left[\frac{1}{2}\sum_{d\in D}\sum_{k\in outputs}(t_{kd} - o_{kd})^2\right]\frac{do_j}{dnet_j}x_{ji} + 2\gamma\sum_{ij} w_{ji}$$

$$= \frac{\partial}{\partial o_j}\left[\frac{1}{2}\sum_{d\in D}(t_{kd} - o_{kd})^2\right]\frac{do_j}{dnet_j}x_{ji} + '' \qquad ''$$

$$= \left[\sum_{d\in D}(t_{kd} - o_{kd})\right](-1)\,o_j(1-o_j)x_{ji} + 2\gamma\sum_{ji} w_{ji}$$

$$= \left(-x_{ji}o_j + x_{ji}o_j^2\right)\sum_{d\in D}(t_{kd} - o_{kd}) + 2\gamma\sum_{ji} w_{ji}$$