1.1. BackPropagation Algorithm :-

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

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

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

$$\frac{\partial E_d}{\partial w_{ji}} = \frac{\partial E_d}{\partial net_j} \times \frac{\partial net_j}{\partial w_{ji}} = \frac{\partial E_d}{\partial net_j} \times x_{ji}$$

Case 1. for o/p units, $\dfrac{\partial E_d}{\partial net_j} = \dfrac{\partial E_d}{\partial o_j} \dfrac{\partial o_j}{\partial net_j}$

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

$$\frac{\partial o_j}{\partial net_j} = 1 - \tanh^2(x)$$

$$\Delta w_{ji} = -\eta \left(1 - \tanh^2(x)\right) \left[-(t_j - o_j)\right] x_{ji}$$

$$= \eta (t_j - o_j) \left(1 - \tanh^2(x)\right) x_{ji}$$

Case 2. for hidden units, $\dfrac{\partial E_d}{\partial net_j} = \sum_{k \in downstream(j)} \dfrac{\partial E_d}{\partial net_k} \dfrac{\partial net_k}{\partial net_j}$

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

δ_k is error term associated with k.

$$\delta_k = (t_j - o_j)\left(1 - \tanh^2(x)\right)$$

$$= \sum -\delta_k \frac{\partial net_k}{\partial o_j} \times \frac{\partial o_j}{\partial net_j} = \sum -\delta_k w_{kj} \frac{\partial o_j}{\partial net_j}$$

$$= \sum -\delta_k w_{kj} \left[1 - \tanh^2(x)\right]$$

$$\Delta w_{ji} = -\eta \sum -\delta_k w_{kj} \left[1 - \tanh^2(x)\right] x_{ji}$$

$$= \eta \left[1 - \tanh^2(x)\right] x_{ji} \sum \delta_k w_{kj}$$

(b) ReLu Activation function : $ReLu = max(0, \cancel{z}\, x)$

$$R'(z) = \begin{cases} 0 & z < 0 \\ 1 & z > 0. \end{cases}$$

$$\Delta w_{ji} = -\eta \frac{\partial E_d}{\partial w_{ji}} \qquad E_d(w) = \frac{1}{2}\sum (t_k - o_k)^2$$

$$\frac{\partial E_d}{\partial w_{ji}} = \frac{\partial E_d}{\partial net_j} \frac{\partial net_j}{\partial w_{ji}} = \frac{\partial E_d}{\partial net_j} x_{ji}$$

Case 1. for o/p units, $\frac{\partial E_d}{\partial net_j} = \frac{\partial E_d}{\partial o_j} \times \frac{\partial o_j}{\partial net_j}$

$$\frac{\partial E_d}{\partial o_j} = \frac{\partial}{\partial o_j}\left(\frac{1}{2}\sum (t_k - o_k)^2\right) = -(t_k - o_k)$$

$$\frac{\partial o_j}{\partial net_j} = max(0, 1)$$

$$\Delta w_{ji} = -\eta \frac{\partial E_d}{\partial w_{ji}} = +\eta (t_k - o_k)\, max(0,1) \times x_{ji}$$

i.e. $\Delta w_{ji} = 0$ if $net_j < 0$

$\underline{\underline{\Delta w_{ji} = \eta (t_k - o_k) x_{ji}}}$ if $net_j > 0$.

Case 2. for hidden layers, $\frac{\partial E_d}{\partial net_j} = \sum_{k \in downstream(j)} \frac{\partial E_d}{\partial net_k} \times \frac{\partial net_k}{\partial net_j}$

$$= \sum -\delta_k \frac{\partial net_k}{\partial net_j}$$

$$= \sum -\delta_k \frac{\partial net_k}{\partial o_j} \frac{\partial o_j}{\partial net_j}$$

$$= \sum -\delta_k w_{kj}\, max(0, 1)$$

$$\Delta w_{ji} = \eta\, max(0,1) \sum \delta_k w_{kj}\, x_{ji}$$

$\Delta w_{ji} = 0$ if $net_j < 0$

$= \underline{\underline{\eta x_{ji} \sum \delta_k w_{kj}}}$ if $net_j > 0$

1.2 output $o = w_0 + w_1(x_1 + x_1^2) + \dots w_n(x_n + x_n^2)$

$$o = w_0 + \sum w_i (x_i + x_i^2)$$

Activation function $f(x) = x$

Learning rate : η

$$\Delta w_{ji} = -\eta \frac{\partial E_d}{\partial w_{ji}} \qquad E_d(w) = \frac{1}{2} \sum_{k \in outputs} (t_k - o_k)^2$$

$$\frac{\partial E_d}{\partial w_{ji}} = \frac{\partial E_d}{\partial net_j} \frac{\partial net_j}{\partial w_{ji}} = \frac{\partial E_d}{\partial net_j} [x_{ji} + x_{ji}^2]$$

Case 1. O/p unit weights $\frac{\partial E_d}{\partial net_j} = \frac{\partial E_d}{\partial o_j} \times \frac{\partial o_j}{\partial net_j}$

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

$$\frac{\partial o_j}{\partial net_j} = 1$$ (derivative of identity function w.r.t net_j)

$$\Delta w_{ji} = -\eta \frac{\partial E_d}{\partial w_{ji}} = \underline{\eta (t_j - o_j) [x_{ji} + x_{ji}^2]}.$$

Case 2. hidden unit weights $\frac{\partial E_d}{\partial net_j} = \sum_{k \in downstream(j)} \frac{\partial E_d}{\partial net_k} \frac{\partial net_k}{\partial net_j}$

$$= \sum -\delta_k \frac{\partial net_k}{\partial net_j}$$

δ_k is error term

$\delta_k = (t_j - o_j)$

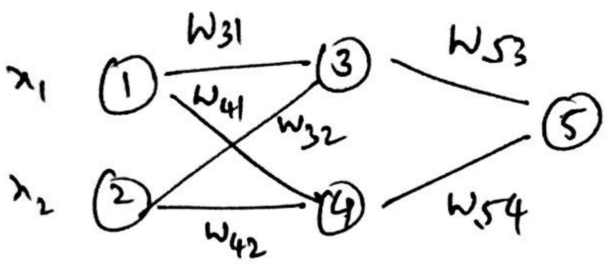
$$= \sum -\delta_k \frac{\partial net_k}{\partial o_j} \times \frac{\partial o_j}{\partial net_j}$$

$$= \sum -\delta_k \; w_{kj} \times 1$$

$$\Delta w_{ji} = -\eta \frac{\partial E_d}{\partial w_{ji}} = \underline{\eta [x_{ji} + x_{ji}^2] \sum \delta_k w_{kj}}$$

1.3 Activation function $f(x) = x$ for input layer

$h(x)$ for hidden layer & O/p layer



ⓐ O/p of 3 : $h(w_{31} x_1 + w_{32} x_2)$

O/p of 4 : $h(w_{41} x_1 + w_{42} x_2)$

O/p of 5 : $y_5 = h\left[w_{53}\, h(w_{31} x_1 + w_{32} x_2) + w_{54}\, h(w_{41} x_1 + w_{42} x_2)\right]$

ⓑ Vector notation $X = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$ $W^{(1)} : \begin{pmatrix} w_{31} & w_{32} \\ w_{41} & w_{42} \end{pmatrix}$

$W^{(2)} : \begin{pmatrix} w_{53} & w_{54} \end{pmatrix}$

O/p of hidden layer = $h(X \cdot W^{(1)})$

O/p of 5, $y_5 = h(W^{(2)} \cdot h(X \cdot W^{(1)}))$

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

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

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

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

$$\boxed{h_t(x) = 2\,h_s(2x) - 1}$$

This proves outputs of activation functions $h_s(x)$ and $h_t(x)$ differ by linear and constants.

4.10. $E(w) = \frac{1}{2} \sum \sum (t_{kd} - o_{kd})^2 + \gamma \sum w_{ji}^2$

$w = w + \Delta w$; $\Delta w = -\eta \frac{\partial E}{\partial w_i}$

$$\frac{\partial E}{\partial w_i} = \frac{\partial E}{\partial net_j} \times \frac{\partial net_j}{\partial w_i}$$

$$\frac{\partial net_i}{\partial w_i} = \frac{\partial}{\partial w_i}(w_i x_i) = x_{ji}$$

This

and

$$\frac{\partial E}{\partial net_j} = \frac{\partial}{\partial net_j}\left[\frac{1}{2}\sum\sum (t_{kd} - o_{kd})^2 \cancel{+ \gamma \sum w_{ji}^2}\right].$$

$$= -(t_{kd} - o_{kd}) \cancel{+ 2\gamma w}$$

$$\frac{\partial}{\partial w}\left[\frac{1}{2}\sum\sum (t_{kd} - o_{kd})^2\right] + \frac{\partial}{\partial w}\left(\gamma \sum w_{ji}^2\right).$$

$$= -(t_{kd} - o_{kd}) x_{ji} + 2\gamma w.$$

$$\Delta w = \eta \left[t_{kd} - o_{kd}\right] x_{ji} + \underbrace{2\gamma w}.$$

constant multiplied to weight before weight update