=) In this project we are told to give both the weight update rule for output layer weight and hidden layer weights

=) Given that output of a single unit to be

$$O = \tanh(\vec{\omega} \cdot \vec{x})$$

Derivation

Note: we will be using ~~Soto~~ stochastic gradient descent rule in order to derive the weights updates in backpropagation

=) Recall $\omega_{ji} \leftarrow \omega_{ji} + \Delta \omega_{ji}$ where

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

(*) here we use the negative sign b relate

=) In this project we are told to give both the weight update rule for output layer weight and hidden layer weight

=) Given that output of a single unit to be

$$O = \tanh(\vec{w} \cdot \vec{x})$$

Derivation

Note: we will be using ~~Sotio~~ stochastic gradient descent rule inorder to derive the weights updates in backpropagation

=) Recall $w_{ji} \leftarrow w_{ji} + \Delta w_{ji}$ where

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

(*) here we use the negative sign to reduce the error

=) Thus, the error equation E_d is given by

$$E_d(\vec{w}) = \frac{1}{2} \sum \sum (t_{k,d} - O_{k,d})^2 \quad —①$$

rule for output layer weight and hidden layer weight

=) Given that output of a single unit to be

$$O = \tanh(\vec{w}, \vec{x})$$

Derivation

Note: we will be using ~~Sato~~ stochastic gradient descent rule inorder to derive the weights updates in backpropagation

=) Recall $w_{ji} \longleftarrow w_{ji} + \Delta w_{ji}$ where

$$\Delta w_{ji} = -\eta \frac{\partial E_d}{\partial w_{ji}}$$ here we use the negative sign to reduce the error (*)

=) Thus, the error equation E_d is given by

$$E_d(\vec{w}) = \frac{1}{2} \sum_{d \in D} \sum_{k \in outputs} (t_{k,d} - O_{k,d})^2 \quad \text{——} (1)$$

=) Next we need to find the derivative of E_d with respect to w_{ji}, in that case we use the chain rule as follows descent rule inorder to derive the weights updates in backpropagation

=) Recall $w_{ji} \leftarrow w_{ji} + \Delta w_{ji}$ where

$$\Delta w_{ji} = -\eta \frac{\partial E_d}{\partial w_{ji}} \quad (*)$$

here we use the negative sign to reduce the error

=) Thus the error equation E_d is given by

$$E_d(\vec{w}) = \frac{1}{2} \sum_{d \in D} \sum_{k \in outputs} (t_{k,d} - O_{k,d})^2 \quad \text{--- (1)}$$

=) Next we need to find the derivative of E_d with respect to w_{ji}, in that case we use the chain rule as follows

$$\frac{\partial E_d}{\partial w_{ji}} = \frac{\partial E_d}{\partial net_j} \frac{\partial net_j}{\partial w_{ji}} \quad \text{--- (2)}$$

where net_j is given by net_j ...

∂w_{ji} negative sign to reduce the error

⇒ Thus the error equation E_d is given by

$$E_d(\vec{w}) = \frac{1}{2} \sum_{d \in D} \sum_{k \in outputs} (t_{k,d} - O_{k,d})^2 \quad \text{—— (1)}$$

⇒ Next we need to find the derivative of E_d with respect to w_{ji}, in that case we use the chain rule as follows

$$\frac{\partial E_d}{\partial w_{ji}} = \frac{\partial E_d}{\partial net_j} \frac{\partial net_j}{\partial w_{ji}} \quad \text{—— (2)}$$

where net_j is given by $net_j = \sum_i w_{ji} x_{ji}$ —— (2.1)

⇒ Differentiating net_j w.r.t w_{ji} we have

$$\frac{\partial net_j}{\partial w_{ji}} = x_{ji} \quad \text{—— (3)}$$

Notice that, since we

$$E_d(\vec{w}) = \frac{1}{2} \sum_{d \in D} \sum_{k \in outputs} (t_{k,d} - O_{k,d})^2 \quad \text{—— (1)}$$

=) Next we need to find the derivative of E_d with respect to w_{ji}, in that case we use the chain rule as follows

$$\frac{\partial E_d}{\partial w_{ji}} = \frac{\partial E_d}{\partial net_j} \frac{\partial net_j}{\partial w_{ji}} \quad \text{—— (2)}$$

where net_j is given by $net_j = \sum_i w_{ji} x_{ji}$ —— (2.1)

=) Differentiating net_j w.r.t w_{ji} we have

$$\frac{\partial net_j}{\partial w_{ji}} = x_{ji} \quad \text{—— (3)}$$

Notice that, since we are using stochastic gradient descent only 1 training example is considered

$$E_d(\vec{w}) = \frac{1}{2} \sum_{d \in D} \sum_{k \in outputs} (t_{k,d} - O_{k,d})^2 \quad \text{—— (1)}$$

=) Next we need to find the derivative of E_d with respect to w_{ji}, in that case we use the chain rule as follows

$$\frac{\partial E_d}{\partial w_{ji}} = \frac{\partial E_d}{\partial net_j} \frac{\partial net_j}{\partial w_{ji}} \quad \text{—— (2)}$$

where net_j is given by $net_j = \sum_i w_{ji} x_{ji}$ —— (2.1)

=) Differentiating net_j w.r.t w_{ji} we have

$$\frac{\partial net_j}{\partial w_{ji}} = x_{ji} \quad \text{—— (3)}$$

Notice that, since we are using stochastic gradient descent only 1 training example is considered

update
weights

descent
backpropagation

=> Substituting (3) into eqn (2) we have

$$\frac{\partial E_d}{\partial w_{ji}} = \frac{\partial E_d}{\partial net_j} x_{ji} \quad \text{—— (4)}$$

=> From eqn (5), making ∂w_{ji} subject formula

$$\partial w_{ji} = \frac{\partial net_j}{x_{ji}} \quad \text{—— (5)}$$

=> Substituting eqn (5) into (4) we have

$$\partial E_d \; x_{ji} \quad \text{—— (6)}$$

update

weights

descent

backpropagation

to

reduce

=> Substituting (3) into eqn (2) we have

$$\frac{\partial E_d}{\partial w_{ji}} = \frac{\partial E_d}{\partial net_j} x_{ji} \quad \text{——} (4)$$

=> From eqn (3), making ∂w_{ji} subject yields

$$\partial w_{ji} = \frac{\partial net_j}{x_{ji}} \quad \text{——} (5)$$

=> Substituting eqn (5) into (*) we have

$$\Delta w_{ji} = -\eta \frac{\partial E_d}{\partial net_j} x_{ji} \quad \text{——} (6)$$

=> What we do not know so far is E_d/ derivative net$_j$

=> Note, in order to calculate the expression $\partial E_d / \partial net_j$

ights

$$\frac{\partial E_d}{\partial w_{ji}} = \frac{\partial E_d}{\partial net_j} x_{ji} \quad \text{——} (4)$$

=) From eqn (3), making ∂w_{ji} subject yields

descent
propagation

$$\partial w_{ji} = \frac{\partial net_j}{x_{ji}} \quad \text{——} (5)$$

=) Substituting eqn (5) into (*) we have

$$\Delta w_{ji} = -\eta \frac{\partial E_d}{\partial net_j} x_{ji} \quad \text{——} (6)$$

to
reduce

=) What we do not know so far is E_d / derivative wrt net_j

=) Note, in order to calculate the expression $\partial E_d / \partial net_j$, then we shall consider 2 cases

i.e (1) for unit j being an output unit to the network
(2) for unit j being an inner unit to the network

(1) ... rule derivation for output unit weight

$$\frac{\partial w_{ji}}{} = \frac{\partial net_j}{x_{ji}} \quad \text{— (5)}$$

=) Substituting eqn (5) into (*) we have

$$\Delta w_{ji} = -\eta \frac{\partial E_d}{\partial net_j} x_{ji} \quad \text{— (6)}$$

=) What we do not know so far is $E_d /$ derivative wrt net_j

=) Note, inorder to calculate the expression $\partial E_d / \partial net_j$, then we shall consider 2 cases

i.e (1) for unit j being an output unit to the network
(2) for unit j being an inner unit to the network

(1) Training rule derivation for output unit weight

=) We begin by taking the chain rule for $\partial E_d / \partial net_j$

i.e $$\partial E_d / = \frac{\partial E_d}{\partial O_j} \frac{\partial O_j}{\partial net_j} \quad \text{— (7)}$$

∂net_j

⇒ What we do not know so far is E_d/ derivative wrt net_j

⇒ Note, inorder to calculate the expression $\partial E_d / \partial net_j$, then we shall consider 2 cases

i.e (1) for unit j being an output unit to the network
(2) for unit j being an inner unit to the network

(1) Training rule derivation for output unit weight

⇒ We begin by taking the chain rule for $\partial E_d / \partial net_j$

i.e $$\frac{\partial E_d}{\partial net_j} = \frac{\partial E_d}{\partial O_j} \frac{\partial O_j}{\partial net_j} \quad \text{---} (7)$$

⇒ What is $\partial E_d / \partial O_j$ = ?

... know so far is E_d / derivative w.r.t. net_j

1 => Note, in order to calculate the expression $\partial E_d / \partial net_j$, then we shall consider 2 cases

i.e (1) for unit j being an output unit to the network
(2) for unit j being an inner unit to the network

(1) Training rule derivation for output unit weight

=> We begin by taking the chain rule for $\partial E_d / \partial net_j$

i.e $$\frac{\partial E_d}{\partial net_j} = \frac{\partial E_d}{\partial O_j} \frac{\partial O_j}{\partial net_j} \qquad (7)$$

=> What is $\partial E_d / \partial O_j$ = ?

$$\Rightarrow \frac{\partial E_d}{\partial O'_j} = \frac{\partial}{\partial O'_j}\left[\frac{1}{2}\sum_{d\in D}\sum_{k\in outputs}(t_{k,d} - O_{k,d})^2\right]$$

$\Rightarrow$ Thus in cases when $k=j$, we ~~have~~ differentiate the above equation to yield

$$\Rightarrow \frac{\partial E_d}{\partial O'_j} = \frac{\partial}{\partial O'_j}\left(\frac{1}{2}\left[t_{k,d} - O_{k,d}\right]^2\right) = 2\cdot\frac{1}{2}\left(t_{j,d} - O_{j,d}\right)\frac{\partial\left[t_{j,d} - O_{j,d}\right]}{\partial O'_j}$$

$$\Rightarrow \frac{\partial E_d}{\partial O'_j} = 2\cdot\frac{1}{2}\left(t_{j,d} - O_{j,d}\right)\left[0-1\right] \quad \text{for same } d\in D$$

$$= -\left(t_j - O_j\right) \quad \text{for same } d\in D$$

... we need to ask ourselves,

$\frac{\partial}{\partial O_j'}$ $\frac{}{\partial O_j}\Big[\sum_{d\in D}\sum_{k\in outputs}(t_{k,d}-O_{k,d})\Big]$

=> Thus in case, when $k=j$, we ~~have~~ differentiate the above equation to yield

$$\Rightarrow \frac{\partial E_d}{\partial O_j'} = \frac{\partial}{\partial O_j}\left(\frac{1}{2}\left[t_{k,d}-O_{k,d}\right]^2\right) = 2\cdot\frac{1}{2}\left(t_{j,d}-O_{j,d}\right)\frac{\partial\left[t_{j,d}-O_{j,d}\right]}{\partial O_j}$$

$$\Rightarrow \frac{\partial E_d}{\partial O_j'} = 2\cdot\frac{1}{2}\left(t_{j,d}-O_{j,d}\right)\left[0-1\right] \quad \text{for same } d\in D$$

$$= -\left(t_j - O_j'\right) \quad \text{for same } d\in D$$

Next, for $\frac{\partial O_j'}{\partial net_j'}$, we need to ask ourselves, what is O_j in our case?

=> Remember $O_j = \tanh(\vec{w}\cdot\vec{x})$

$$\Rightarrow \frac{\partial O_j'}{\partial net_j'} = \frac{\partial\left[\tanh(\vec{w},\vec{x})\right]}{\partial net_j'} \quad \text{—— (8)}$$

$$\frac{\partial O_j'}{\partial O_j'^2} \ldots = \ldots \frac{1}{2}(t_{j,d} - O_{j,d}) \frac{\partial(t_{j,d} - O_{j,d})}{\partial O_j}$$

$$\Rightarrow \frac{\partial E_d}{\partial O_j'} = 2 \cdot \frac{1}{2}(t_{j,d} - O_{j,d})[0-1] \quad \text{for same } d \in D$$

$$= -(t_j - O_j) \quad \text{for same } d \in D$$

Next, for $\frac{\partial O_j'}{\partial net_j'}$, we need to ask ourselves, what is O_j in our case?

$\Rightarrow$ Remember $O_j = \tanh(\vec{w} \cdot \vec{x})$

$$\Rightarrow \frac{\partial O_j'}{\partial net_j'} = \frac{\partial[\tanh(\vec{w}, \vec{x})]}{\partial net_j'} \quad \text{—— } (8)$$

$\Rightarrow$ Next we need to focus on differentiating

$$\frac{\partial[\tanh(\vec{w}, \vec{x})]}{\partial net_j'}$$

$\Rightarrow$ But since $\tanh'(x) = 1 - \tanh^2(x)$, then we have

$$\frac{\partial[\tanh(\vec{w}, \vec{x})]}{\ldots} = \ldots 1 - \tanh^2(\vec{w}, \vec{x}) \ldots$$

Next, for $\frac{\partial O_j}{\partial net_j}$, we need to ask ourselves, what is O_j in our case?

=) Remember $O_j = \tanh(\vec{\omega} . \vec{x})$

$$\Rightarrow \frac{\partial O_j}{\partial net_j} = \frac{\partial [\tanh(\vec{\omega}, \vec{x})]}{\partial net_j} \qquad \text{——} \quad 8$$

=) Next we need to focus on differentiating $\frac{\partial [\tanh(\vec{\omega}, \vec{x})]}{\partial net_j}$

=) But since $\tanh'(x) = 1 - \tanh^2(x)$, then we have that

$$\frac{\partial [\tanh(\vec{\omega}, \vec{x})]}{\partial net_j} = 1 - \tanh^2(\vec{\omega}, \vec{x}) = 1 - \tanh^2 [net_j$$

$$= [1 - (O_j)^2] \qquad \text{——} \quad 8.1$$

Therefore $\frac{\partial E_d}{\partial net_j} = -[t_j - O_j][1 - (O_j)^2] \qquad \text{——} \quad 9$

$\Rightarrow$ Remember $O_j = \tanh(\vec{\omega} . \vec{x})$

$$\Rightarrow \frac{\partial O_j}{\partial net_j} = \frac{\partial [\tanh(\vec{\omega}, \vec{x})]}{\partial net_j} \quad \text{———} \quad (8)$$

$\Rightarrow$ Next we need to focus on differentiating $\frac{\partial [\tanh(\vec{\omega}, \vec{x})]}{\partial net_j}$

$\Rightarrow$ But since $\tanh'(x) = 1 - \tanh^2(x)$, then we have that

$$\frac{\partial [\tanh(\vec{\omega}, \vec{x})]}{\partial net_j} = 1 - \tanh^2(\vec{\omega}, \vec{x}) = 1 - \tanh^2[net_j]$$

$$= [1 - (O_j)^2] \quad \text{———} \quad (8.1)$$

$\Rightarrow$ Therefore $$\frac{\partial E_d}{\partial net_j} = -[t_j - O_j][1 - (O_j)^2] \quad \text{———} \quad (9)$$

$\Rightarrow$ This is the error term function w.r.t output unit j

Now for the training rule

$$\Delta w_{ji} = -\eta \, \partial E_d / \partial w_{ji} \quad , \text{ from eqn } (6) \text{ we have}$$

a bove

$$\Rightarrow \Delta w_{ji} = -\eta \, \partial E_d / \partial net_j \; x_{ji}$$

$- O_{jd}]$

$\Rightarrow$ Subtituting eqn (7) into the above eqn we have

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

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

let

the above

$[t_{jd} - O_{jd}]$

O_j

$\Delta w_{ji} = -\eta \frac{\partial E_d}{\partial w_{ji}}$, from eqn (6) we have

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

$\Rightarrow$ Substituting eqn (9) into the above eqn we have

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

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

We can conclude that, let

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

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

Training rule derivation for hidden unit weights the above

$\frac{\partial E_d}{\partial w_{ji}}$, from eqn ⑤ we ...

$[t_{jd} - O_{jd}]$

∂O_j

Δ

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

$\Rightarrow$ Substituting eqn ① into the above eqn we have

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

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

We can conclude that, let

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

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

② Training rule derivation for hidden unit weights

$\Rightarrow$ note that

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

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

substituting eqn (1) into the above eqn we have

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

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

we can conclude that, let

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

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

② Training rule derivation for hidden unit weights

$\Rightarrow$ note that

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

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

where $k \in Downstream(j)$ implies that we need k to be inclusive of the j values

$\Rightarrow$ k is the output unit in this case, since it is ~~...~~ from j values which was ~~...~~ entirely output

...clude that, let

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

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

② Training rule derivation for hidden unit weights

⇒ note that

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

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

where $k \in Downstream(j)$ implies that we need k to be inclusive of the j values,

⇒ k is the output unit in this case, since it is downstreamed from j value which ~~was~~ was ~~our~~ earlier output u...

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

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

② Training rule derivation for hidden unit weights

⇒ Note that

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

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

where $k \in Downstream(j)$ implies that we need k to be inclusive of the j values

⇒ K is the output unit in this case, since it is downstreamed from j value which was earlier output unit

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

let $\frac{\partial net_k}{\partial O_j} = w_{kj}$ $\Rightarrow \frac{\partial E_d}{\partial net_j} = \sum_{k \in Downstream(j)} - \delta_k w_{kj} \frac{\partial O_j}{\partial net_j}$ — (10)

where $\frac{\partial net_k}{\partial O_j} = \frac{\partial x_{kj}}{\partial O_j} w_{kj} = \frac{\partial O_j w_{kj}}{\partial O_j} = w_{kj}$

$\Rightarrow$ But $O_j = \tanh(\vec{w}.\vec{x})$

$$\Rightarrow \frac{\partial E_d}{\partial net_j} = \sum_{k \in Downstream(j)} - \delta_k w_{kj} \frac{\partial\left[\tanh(\vec{w}.\vec{x})\right]}{\partial net_j} \quad \text{— (11)}$$

But from

let $\frac{\partial net_k}{\partial O_j} = w_{kj}$ $\Rightarrow \frac{\partial E_d}{\partial net_j} = \sum_{k \in Downstream(j)} -\delta_k w_{kj} \frac{\partial O_j}{\partial net_j}$ — (10)

where $\frac{\partial net_k}{\partial O_j} = \frac{\partial \cancel{O_{kj}} w_{kj}}{\partial O_j} = \frac{\partial O_j w_{kj}}{\partial O_j} = w_{kj}$

$\Rightarrow$ But $O_j = \tanh(\vec{w}.\vec{x})$

$$\Rightarrow \frac{\partial E_d}{\partial net_j} = \sum_{k \in Downstream(j)} -\delta_k w_{kj} \frac{\partial [\tanh(\vec{w}.\vec{x})]}{\partial net_j} \quad \text{— (11)}$$

$\Rightarrow$ But from eqn (8.1) we know that $\frac{\partial [\tanh(\vec{w}\vec{x})]}{\partial net_j} = [1-(O_j)^2]$

$\Rightarrow$ Substituting (8.1) into (11) we have

$$\frac{\partial E_d}{\partial net_j} = \sum_{k \in Downstream(j)} -\delta_k w_{kj} [1-(O_j)^2] \quad \text{— (12)}$$

whare $\frac{\partial net_k}{\partial O_j} = \frac{\partial O_{kj} w_{kj}}{\partial O_j} = \frac{\partial O_j w_{kj}}{\partial O_j} = w_{kj}$

=> But $O_j = \tanh(\vec{w}.\vec{x})$

=> $\frac{\partial E_d}{\partial net_j} = \sum_{k \in Downstream(j)} -\delta_k w_{kj} \frac{\partial [\tanh(\vec{w}.\vec{x})]}{\partial net_j}$ —— (11)

=> But from eqn (8.1) we know that $\frac{\partial [\tanh(\vec{w}\vec{x})]}{\partial net_j} = [1-(O_j)^2]$

=> Substituting (8.1) into (11) we have

$$\frac{\partial E_d}{\partial net_j} = \sum_{k \in Downstream(j)} -\delta_k w_{kj} [1-(O_j)^2] \quad \text{—— (12)}$$

=> From eqn (6) we substitute (12), thus

$$\Delta w_{ji} = -\eta [1-(O_j)^2] \sum_{k \in Downstream(j)} \delta_k w_{kj} x_{ji}$$

Further simplification

=> let $\delta_j = [1-(O_j)^2] \sum \delta$

$\Rightarrow$ But $O_j = \tanh(\ldots)$

$$\Rightarrow \frac{\partial E_d}{\partial net_j} = \sum_{K \in Downstream(j)} -\delta_K \omega_{Kj} \frac{\partial [\tanh(\vec{\omega}.\vec{x})]}{\partial net_j} \quad \text{—— (11)}$$

$\Rightarrow$ But from eqn (8.1) we know that $\frac{\partial [\tanh(\vec{\omega}\,\vec{x})]}{\partial net_j} = [1-(O_j)^2]$

$\Rightarrow$ Substituting (8.1) into (11) we have

$$\frac{\partial E_d}{\partial net_j} = \sum_{K \in Downstream(j)} -\delta_K \omega_{Kj} [1-(O_j)^2] \quad \text{—— (12)}$$

$\Rightarrow$ From eqn (6) we substitute (12), thus

$$\Delta \omega_{ji} = -2 [1-(O_j)^2] \sum_{K \in Downstream(j)} \delta_K \omega_{Kj} x_{ji}$$

Further simplification

$$\Rightarrow \text{let} \quad \delta_j = [1-(O_j)^2] \sum_{K \in Downstream(j)} \delta_K \omega_{Kj}$$

where K belongs to downstream of j and j is the internal node

$\Rightarrow$ $\Delta \omega_{ji}$ ...

=) But from eqn (8.1) we know that $\frac{\partial [\tanh(\ldots)]}{\partial net_j} = [1-(O_j)^2]$

=) Substituting (8.1) into (11) we have

$$\frac{\partial E_d}{\partial net_j} = \sum_{k \in Downstream(j)} - \delta_k \, w_{kj} \, [1-(O_j)^2] \quad \text{—— (12)}$$

=) From eqn (6) we substitute (12), thus

$$\Delta w_{ji} = -\eta \, [1-(O_j)^2] \sum_{k \in Downstream(j)} \delta_k \, w_{kj} \, x_{ji}$$

Further simplification

=) let $\delta_j = [1-(O_j)^2] \sum_{k \in Downstream(j)} \delta_k \, w_{kj}$

where k belongs to downstream of 'j' and 'j' is the internal node

therefore =) $\Delta w_{ji} = \eta \, \delta_j \, x_{ji}$

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

=) Substituting (8.1) into (11) we have

$$\frac{\partial E_d}{\partial net_j} = \sum_{k \in Downstream(j)} -\delta_k w_{kj} [1-(O_j)^2] \quad \text{—— (12)}$$

=) From eqn (6) we substitute (12), thus

$$\Delta w_{ji} = -\eta [1-(O_j)^2] \sum_{k \in Downstream(j)} \delta_k w_{kj} x_{ji}$$

Further simplification

=) let $\delta_j = [1-(O_j)^2] \sum_{k \in Downstream(j)} \delta_k w_{kj}$

where k belongs to downstream of 'j' and 'j' is the internal node

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

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

END OF PROOF / DERIVATION $\boxtimes$