# DERIVATION OF BACKPROPAGATION WEIGHT UPDATES

$$E(\vec{w}) = \frac{1}{2} \sum_{d \in D} \underbrace{\sum_{k \in outputs} (t_{kd} - o_{kd})^2}_{\text{SUM OVER ALL NETWORK OUTPUT UNITS}}$$

$d \in D$: Training Examples.

$$\underbrace{E_d(\vec{w}) = \frac{1}{2} \sum_{k \in outputs} (t_k - o_k)^2}_{\text{STOCHASTIC VERSION OF GRADIENT DESCENT}}$$

(SUMMING OVER ALL NETWORK OUTPUT UNITS FOR EACH TRAINING EXAMPLE)

$x_{ji}$ = ith INPUT TO UNIT $j$

$w_{ji}$ = weight associated with ith input to unit $j$

$net_j$ = $\sum_i w_{ji} x_{ji}$ [WEIGHTED SUM OF INPUTS FOR UNIT $j$]

$o_j$ = the output computed by unit $j$

$t_j$ = target output for unit $j$

$\sigma$ = sigmoid function

Outputs = the set of units in the final layer of the network.

Downstream($j$) = the set of units whose immediate inputs include the output of unit $j$

$$\frac{\partial E_d}{\partial w_{ji}} = \frac{\partial E_d}{\partial net_j} \frac{\partial net_j}{\partial w_{ji}} \quad \text{[USING CHAIN RULE]}$$

$$= \underbrace{\frac{\partial E_d}{\partial net_j}}_{} x_{ji} \quad \dots\dots \quad (1)$$

DERIVE AN EXPRESSION FOR THIS.

TWO CASES :

CASE (1) $j$ is an OUTPUT UNIT

CASE (2) $j$ is an INTERNAL UNIT

CASE 1 :

AGAIN, APPLYING CHAIN RULE,

$$\frac{\partial E_d}{\partial net_j} = \overbrace{\frac{\partial E_d}{\partial o_j}}^{\text{TERM1}} \overbrace{\frac{\partial o_j}{\partial net_j}}^{\text{TERM2}} \quad \dots \quad (4)$$

HINT : $w_{ji}$ INFLUENCES NETWORK THROUGH $net_j$

$net_j$ INFLUENES NETWORK THROUGH $o_j$

CONSIDER TERM 1,

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

$$\frac{\partial}{\partial o_j} (t_k - o_k)^2 = 0 \quad \text{except when } j = k.$$

$$\frac{\partial E_d}{\partial o_j} = \frac{\partial}{\partial o_j} \frac{1}{2}(t_j - o_j)^2$$ [NOTE THAT WE DON'T HAVE A SUMMATION ANYMORE]

$$= \frac{-1}{2} \cdot 2 \cdot (t_j - o_j)$$

$$= -(t_j - o_j) \quad \ldots \quad (2)$$

CONSIDER TERM 2,

$$\frac{\partial o_j}{\partial net_j} = \frac{\partial \sigma(net_j)}{\partial net_j}$$ [DERIVATIVE OF SIGMOID FUNCTION]

$$= o_j(1 - o_j) \quad \ldots \quad (3)$$

SUBSTITUTING (2) & (3) IN (4)

$$\frac{\partial E_d}{\partial net_j} = -(t_j - o_j)\, o_j (1 - o_j)$$

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

$$= -n\left[-(t_j - o_j)\, o_j(1 - o_j)\right] x_{ji}$$

$$= \eta\left[(t_j - o_j) \cdot o_j (1 - o_j)\right] x_{ji}$$

CASE 2 :

DOWNSTREAM $(j)$ : All units immediately downstream of unit $j$.

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

Let $\delta_k = -\dfrac{\partial E_d}{\partial net_k}$.

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

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

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

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

Let $\delta_j = -\dfrac{\partial E_d}{\partial net_j}$

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