Låt $w_{ij}$ vara alla vikter mellan två noder i och j.
Låt $y$ vara utdatan.
Låt $t$ vara önskad utdata.
Låt $o_i$ vara utdatan från noden $i$.
Låt $\sigma$ vara en sigmoid funktion.
Låt errorfunktionen vara:

$$E(w_{ij}) = (\frac{1}{2})(y-t)^2$$

Enligt gradientsökning så kommer:

$$E(w_{ij}) \geq E(w_{ij} - \alpha \nabla E(w_{ij}))$$

Så för att minska felet så behöver vi få ut $\nabla E(w_{ij})$.
Vilket bara är de partiella derivatorna $\frac{\partial E}{\partial w_{ij}}$.
Vilket är, enligt kedjeregeln:

$$\frac{\partial E}{\partial o_j} \frac{\partial o_j}{\partial net_j} \frac{\partial net_j}{\partial w_{ij}}$$

då $o_j = \sigma(net_j)$,
och $net_j = \sum_{k=1}^{n}(w_{kj} o_k)$ utom för indata där $net_j$ är givet.
Vi beräknar de tre derivatorna var för sig.
Vi börjar med den längst bak:

$$\frac{\partial net_j}{\partial w_{ij}} = \frac{\partial \sum_{k=1}^{n} w_{kj} o_k}{\partial w_{ij}} = o_i$$

då allt utom $w_{ij} o_i$ (d.v.s. när k=i) är konstanter och försvinner.

$$\frac{\partial o_j}{\partial net_j} = \frac{\partial \sigma(net_j)}{\partial net_j} = \sigma'(net_j)$$

$\frac{\partial E}{\partial o_j}$ behöver delas upp i två fall, en då $j$ är utdata och en då det är i vårat gömda lager.
Låt oss börja med utdatanoden.
Notera att utdatan för ett utdatalager är $y$.

$$\frac{\partial E}{\partial o_j} = \frac{\partial E}{\partial y} = \frac{\partial (\frac{1}{2})(y-t)^2}{\partial y} = y - t$$

Nu, om den är i det gömda lagret istället.
Notera att $E$ är en funktion av indatan till alla neuroner som får indata av $j$.
Låt $L$ vara mängden av alla dessa tal

$$\frac{\partial E}{\partial o_j} = \frac{\partial E(o_j)}{\partial o_j} = \frac{\partial E(net)}{\partial o_j}$$

Notera att vi nu kan derivera totalt och med hjälp av kedjeregeln få:

$$\frac{\partial E}{\partial o_j} = \sum_{l \in L} \left( \frac{\partial E}{\partial net_l} \frac{\partial net_l}{\partial o_j} \right) = \sum_{l \in L} \left( \frac{\partial E}{\partial net_l} \frac{\partial \sum_{k=1}^{n} w_{kl} o_k}{\partial o_j} \right) = \sum_{l \in L} \left( \frac{\partial E}{\partial o_l} \frac{\partial o_l}{\partial net_l} w_{jl} \right)$$

notera att $\frac{\partial E}{\partial o_l} \frac{\partial o_l}{\partial net_l}$ redan är uträknad.
då har vi

$$\frac{\partial E}{\partial w_{ij}}$$

till

$$(o_j - t_j)\sigma'(net_j)o_i$$

om det är utdata och

$$\frac{\partial E}{\partial w_{ij}}$$

till

$$\left(\sum_{l \in L} \frac{\partial E}{\partial o_l} \frac{\partial o_l}{\partial net_l} w_{jl}\right)\sigma'(net_j)o_i$$

om det är i det gömda lagret.