1 The Delta Rule

Let E_p be the error for a particular input pattern. We will just look at one pattern, and drop the index. Let T_j be the desired *Target* pattern for node j. The output of a simple linear net is computed as follows:

$$O_j = \sum_{\forall i} X_i W_{i,j}$$

$$E = 1/2(T_j - O_j)^2$$

This is a composite function: **(ERROR (OUT (**$W_{i,j}$**)))**
We can apply the Chain Rule:

$$\frac{\delta E}{\delta W_{i,j}} = \frac{\delta E}{\delta O_j}\frac{\delta O_j}{\delta W_{i,j}}$$

$$\frac{\delta E}{\delta O_j} = -(T_j - O_j) \qquad \frac{\delta O_j}{\delta W_{i,j}} = X_i$$

$$-\frac{\delta E}{\delta W_{i,j}} = (T_j - O_j)X_i$$

===

For networks with "sigmoid units" the raw output is squashed by the sigmoid. Denote the squashed output by S_j. Using the logistic function:

$$S_j = \frac{1}{1 + e^{-O_j/t}} \qquad O_j = \sum_{\forall i} X_i W_{i,j}$$

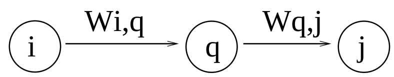
Again, a composite function: **(ERROR (SIG (OUT (**$W_{i,j}$**))))**

$$\frac{\delta E}{\delta W_{i,j}} = \frac{\delta E}{\delta S_j}\frac{\delta S_j}{\delta O_j}\frac{\delta O_j}{\delta W_{i,j}}$$

$$\frac{\delta E}{\delta S_j} = -(T_j - S_j) \qquad \frac{\delta S_j}{\delta O_j} = S_j(1 - S_j) \qquad \frac{\delta O_j}{\delta W_{i,j}} = X_i$$

$$-\frac{\delta E}{\delta W_{i,j}} = (T_j - S_j)S_j(1 - S_j)X_i$$

Now consider a 2-layer network:



(ERROR (SIG.j (OUT.j (SIG.q (OUT.q ($W_{i,q}$)**))))))**

$$\frac{\delta E}{\delta W_{i,q}} = \frac{\delta E}{\delta S_j}\frac{\delta S_j}{\delta O_j}\frac{\delta O_j}{\delta S_q}\frac{\delta S_q}{\delta O_q}\frac{\delta O_q}{\delta W_{i,q}}$$

$$\frac{\delta E}{\delta O_j} = \frac{\delta E}{\delta S_j}\frac{\delta S_j}{\delta O_j} = -(T_j - S_j)S_j(1 - S_j)$$

$$\frac{\delta O_j}{\delta S_q} = w_{q,j}$$

$$\frac{\delta S_q}{\delta O_q} = S_q(1 - S_q) \qquad \frac{\delta O_q}{\delta W_{i,q}} = X_i$$

$$-\frac{\delta E}{\delta W_{i,q}} = \{\sum_j (T_j - S_j)S_j(1 - S_j)w_{q,j}\}S_q(1 - S_q)X_i$$