# 1 Forward Phase

$$o_j^{\ell+1} = \sigma\left(\sum_{i=0}^{n^\ell} \sigma_i^\ell w_{ij}^\ell\right) \qquad (1 \le i < n^{L-1})$$

$$o_j^0 = I_j^d \qquad \text{(base case)}$$

where $o_0^\ell = 1$ always. Equivalently, $o_0^0 = 1$ and $w_{i0}^\ell = \delta_{i0}^{(\text{kronecker})}$

# 2 Backwards Phase

Base case:

$$\delta_j^{L-1} = o_j^{L-1}(1 - o_j^{L-1})(o_j^{L-1} - D_j^d) \qquad (0 \le j < n^{L-1})$$

$$\delta_j^\ell = o_j^\ell(1 - o_j^\ell)\sum_k w_{jk}^\ell \delta_k^{\ell+1} \qquad 0 \le j < n^\ell$$

Note that this is kind of funky for the term which would affect the biases, $\delta_0^\ell$. because $o_0^\ell = 1$, $\delta_0^\ell = 0$ always.

# 3 Stepping

$$\Delta w_{ij}^\ell = -\alpha \delta_j^{\ell+1} o_i^\ell$$

Note that $\Delta w_{i0}^\ell$ will always evaluate to zero. So this column is never changed. But $\Delta_{0j}^\ell$ can evaluate to nonzero values, so the biases are indeed updated.