$$K = W_K X$$
$$k_j^i = w_r^i x_j^r$$
$$q_j^i = v_r^i x_j^r$$
$$\frac{\partial k_j^i}{\partial w_b^a} = \delta_{rb}^{ia} x_j^r$$
$$= \delta^{ia} x_j^b$$
$$\frac{\partial q_j^i}{\partial v_b^a} = \delta_{rb}^{ia} x_j^r$$
$$= \delta^{ia} x_j^b$$

$$S = K^T Q$$
$$s_j^i = k_i^r q_j^r$$
$$\frac{\partial s_j^i}{\partial q_b^a} = k_i^r \delta_{jb}^{ra}$$
$$= k_i^a \delta_{jb}$$

$$\frac{\partial s_j^i}{\partial v_b^a} = \frac{\partial s_j^i}{\partial q_d^c} \frac{\partial q_d^c}{\partial v_b^a}$$
$$= k_i^c \delta_{jd} \delta^{ca} x_d^b$$
$$= k_i^a x_j^b$$

$$K = WX$$
$$k_d^c = w_r^c x_d^r$$
$$\frac{\partial k_d^c}{\partial w_b^a} = \delta_{rb}^{ca} x_d^r = \delta^{ca} x_d^b$$
$$S = K^T Q$$
$$s_j^i = k_i^r q_j^r$$
$$\frac{\partial s_j^i}{\partial k_b^a} = q_j^r \delta_{ib}^{ra}$$
$$= q_j^a \delta_{ib}$$

$$\frac{\partial s_j^i}{\partial w_b^a} = \frac{\partial s_j^i}{\partial k_d^c} \frac{\partial k_d^c}{\partial w_b^a}$$
$$= q_j^c \delta_{id} \delta^{ca} x_d^b$$
$$= q_j^a x_i^b$$

$$Y = f(S)$$
$$y_j^i = s_{ij}/t_j$$
$$\frac{\partial y_j^i}{\partial s_b^a} = \delta^{kjb} d^{iak} = y_j^i (\delta_b^a - y_b^a) \delta_{jb}$$
$$\frac{\partial l}{\partial s_b^a} = \frac{\partial l}{\partial y_j^i} \frac{\partial y_j^i}{\partial s_b^a}$$
$$\frac{\partial l}{\partial s_b^a} = \Delta_j^i \delta^{jbk} d^{iak}$$
$$\frac{\partial l}{\partial s_b^a} = \sum_{i,j,k} \Delta_j^i \delta^{jbk} d^{iak}$$
$$\frac{\partial l}{\partial s_b^a} = \Delta_b^i d^{iab}$$