# Backward Propagation

**$\mathbf{R}^{(3)}$**

$$r^{(3)} = \frac{\partial E(w)}{\partial a^{(3)}} = \frac{-t}{e^{ta^{(3)}} + 1}$$

**$\mathbf{R}^{(2)}$**

$$r_q^{(2)} = \frac{\partial E(w)}{\partial a_q^{(2)}} = \frac{\partial E(w)}{\partial a^{(3)}} \frac{\partial a^{(3)}}{\partial a_q^{(2)}}$$

**L**

$$\begin{aligned} r_{L,q}^{(2)} &= r^{(3)} \frac{w_q^{(3)} \cdot a_{LR,q}^{(2)} \cdot e^{a_{R,q}^{(2)} + a_{L,q}^{(2)}}}{(e^{a_{R,q}^{(2)}} + 1)(e^{a_{L,q}^{(2)}} + 1)^2} \\ &= r^{(3)} w_q^{(3)} \sigma(a_{R,q}^{(2)}) \frac{\partial \sigma(a_{L,q}^{(2)})}{\partial a_{L,q}^{(2)}} \\ \mathbf{r}_L^{(2)} &= r^{(3)} \mathbf{w}^{(3)^T} \operatorname{diag}(\sigma(\mathbf{a}_R^{(2)}) \operatorname{diag}(\frac{\partial \sigma(\mathbf{a}_L^{(2)})}{\partial \mathbf{a}_L^{(2)}})) \end{aligned}$$

**R**

$$\begin{aligned} r_{R,q}^{(2)} &= r^{(3)} \frac{w_q^{(3)} \cdot a_{LR,q}^{(2)} \cdot e^{a_{R,q}^{(2)} + a_{L,q}^{(2)}}}{(e^{a_{L,q}^{(2)}} + 1)(e^{a_{R,q}^{(2)}} + 1)^2} \\ &= r^{(3)} w_q^{(3)} \sigma(a_{L,q}^{(2)}) \frac{\partial \sigma(a_{R,q}^{(2)})}{\partial a_{R,q}^{(2)}} \\ \mathbf{r}_R^{(2)} &= r^{(3)} \mathbf{w}^{(3)^T} \operatorname{diag}(\sigma(\mathbf{a}_L^{(2)}) \operatorname{diag}(\frac{\partial \sigma(\mathbf{a}_R^{(2)})}{\partial \mathbf{a}_R^{(2)}})) \end{aligned}$$

**LR**

$$\begin{aligned} r_{LR,q}^{(2)} &= r^{(3)} \frac{w_q^{(3)}}{(e^{-a_{L,q}^{(2)}} + 1)(e^{-a_{R,q}^{(2)}} + 1)} \\ &= r^{(3)} w_q^{(3)} \sigma(a_{L,q}^{(2)}) \sigma(a_{R,q}^{(2)}) \\ \mathbf{r}_{LR}^{(2)} &= r^{(3)} \mathbf{w}^{(3)^T} \operatorname{diag}(\sigma(\mathbf{a}_L^{(2)}) \operatorname{diag}(\sigma(\mathbf{a}_R^{(2)}))) \end{aligned}$$

**L**

$$r^{(1)}_{L,q} = \frac{\partial E(w)}{\partial a^{(1)}_{L,q}} = \sum_k \left(\frac{\partial E(w)}{\partial a^{(2)}_{L,k}}\frac{\partial a^{(2)}_{L,k}}{\partial a^{(1)}_{L,q}} + \frac{\partial E(w)}{\partial a^{(2)}_{R,k}}\frac{\partial a^{(2)}_{R,k}}{\partial a^{(1)}_{L,q}} + \frac{\partial E(w)}{\partial a^{(2)}_{LR,k}}\frac{\partial a^{(2)}_{LR,k}}{\partial a^{(1)}_{L,q}}\right)$$

$$= \sum_k (r^{(2)}_{L,k} \cdot w^{(2)}_{L,k,q} \cdot sech(a^{(1)}_{L,q})^2 + r^{(2)}_{LR,k} \cdot w^{(2)}_{LR,k,q} \cdot sech(a^{(1)}_{L,q}))^2$$

$$= sech(a^{(1)}_{L,q})^2 \sum_k (r^{(2)}_{L,k} \cdot w^{(2)}_{L,k,q} + r^{(2)}_{LR,k} \cdot w^{(2)}_{LR,k,q})$$

$$\mathbf{r}^{(1)}_L = \mathrm{diag}(sech(\mathbf{a}^{(1)}_L)^2)({\mathbf{W}^{(2)}_L}^T \mathbf{r}^{(2)}_L + {\mathbf{W}^{(2)}_{LR}}^T \mathbf{r}^{(2)}_{LR})$$

**R**

$$r^{(1)}_{R,q} = \frac{\partial E(w)}{\partial a^{(1)}_{R,q}} = \sum_k \left(\frac{\partial E(w)}{\partial a^{(2)}_{L,k}}\frac{\partial a^{(2)}_{L,k}}{\partial a^{(1)}_{R,q}} + \frac{\partial E(w)}{\partial a^{(2)}_{R,k}}\frac{\partial a^{(2)}_{R,k}}{\partial a^{(1)}_{R,q}} + \frac{\partial E(w)}{\partial a^{(2)}_{LR,k}}\frac{\partial a^{(2)}_{LR,k}}{\partial a^{(1)}_{R,q}}\right)$$

$$= \sum_k (r^{(2)}_{R,k} \cdot w^{(2)}_{R,k,q} \cdot sech(a^{(1)}_{R,q})^2 + r^{(2)}_{LR,k} \cdot w^{(2)}_{LR,k,q} \cdot sech(a^{(1)}_{R,q}))^2$$

$$= sech(a^{(1)}_{R,q})^2 \sum_k (r^{(2)}_{R,k} \cdot w^{(2)}_{R,k,q} + r^{(2)}_{LR,k} \cdot w^{(2)}_{LR,k,q})$$

$$\mathbf{r}^{(1)}_R = \mathrm{diag}(sech(\mathbf{a}^{(1)}_R)^2)({\mathbf{W}^{(2)}_R}^T \mathbf{r}^{(2)}_R + {\mathbf{W}^{(2)}_{LR}}^T \mathbf{r}^{(2)}_{LR})$$