Sergei Volodin

1. $s_t = \tanh(W \cdot s_{t-1} + U \cdot x_t)$
2. $z_{tk} = \text{SoftMax}_k(V \cdot s_t)$
3. $E_t = -y_t \cdot \ln z_t$

Then

$$\frac{\partial E_t}{\partial V} = \frac{\partial}{\partial V}(-\sum_k y_{tk} \ln z_{tk}) = (-\sum_k y_{tk} \frac{\partial}{\partial V} \ln z_{tk}) = -\sum_k \frac{y_{tk}}{z_{tk}} \frac{\partial z_{tk}}{\partial V}$$

Consider

$$\frac{\partial z_{tk}}{\partial V_{ij}} = \frac{\partial}{\partial V_{ij}} \frac{e^{(V \cdot s_t)_k}}{\sum\limits_p e^{(V \cdot s_t)_p}} \boxed{=}$$

Where $(V \cdot s_t)_k = \sum\limits_s V_{ks} s_{ts}$, $\frac{\partial (V \cdot s_t)_k}{\partial V_{ij}} = \delta_{ki} s_{tj}$, $\frac{\partial e^{(V \cdot s_t)_k}}{\partial V_{ij}} = \delta_{ki} s_{tj} e^{(V \cdot s_t)_k}$, $\frac{\partial \sum\limits_p e^{(V \cdot s_t)_p}}{\partial V_{ij}} = \sum\limits_p \delta_{pi} s_{tj} e^{(V \cdot s_t)_p} = s_{tj} e^{(V \cdot s_t)_i}$

$$\boxed{=} \frac{\delta_{ki} s_{tj} e^{(V \cdot s_t)_k} \sum\limits_p e^{(V \cdot s_t)_p} - e^{(V \cdot s_t)_k} s_{tj} e^{(V \cdot s_t)_i}}{(\sum\limits_p e^{(V \cdot s_t)_p})^2} = \delta_{ki} s_{tj} z_{tk} - z_{tk} z_{ti} s_{tj} = z_{tk} s_{tj} (\delta_{ki} - z_{ti})$$

Note that $\sum\limits_k y_{tk} = 1$.

Then

$$\frac{\partial E_t}{\partial V_{ij}} = -\sum_k \frac{y_{tk}}{z_{tk}} z_{tk} s_{tj} (\delta_{ki} - z_{ti}) = -\sum_k y_{tk} s_{tj} \delta_{ki} + \sum_k y_{tk} s_{tj} z_{ti} = -y_{ti} s_{tj} + \sum_k y_{tk} s_{tj} z_{ti} = -(y_t s_t^T)_{ij} + (z_t s_t^T)_{ij}$$

Then $\frac{\partial E_t}{\partial V} = -y_t s_t^T + z_t s_t^T = \boxed{(z_t - y_t) s_t^T}$