* LSTM Backprop →

$$\frac{\partial a_{ke}^{\langle t\rangle}}{\partial \Gamma_{o_{ij}}^{\langle t\rangle}} = \tanh\left(C_{ke}^{\langle t\rangle}\right)\delta_{k,i}\,\delta_{e,j} \;;\; \frac{\partial a_{ke}^{\langle t\rangle}}{\partial G_{ij}^{\langle t\rangle}} = \Gamma_{o_{ke}}^{\langle t\rangle}\left(1-\tanh^2\left(C_{ke}^{\langle t\rangle}\right)\right)\delta_{k,i}\,\delta_{e,j}$$

$$\frac{\partial C_{ke}^{\langle t\rangle}}{\partial \Gamma_{f_{ij}}^{\langle t\rangle}} = C_{ke}^{\langle t-1\rangle}\,\delta_{k,i}\,\delta_{e,j} \;;\; \frac{\partial C_{ke}^{\langle t\rangle}}{\partial \Gamma_{\mu_{ij}}^{\langle t\rangle}} = \hat{C}^{\langle t\rangle}\,\delta_{k,i}\,\delta_{e,j}$$

$$\frac{\partial C_{ke}^{\langle t\rangle}}{\partial C_{ij}^{\langle t-1\rangle}} = \Gamma_{f_{ke}}^{\langle t\rangle}\,\delta_{k,i}\,\delta_{e,j} \;;\; \frac{\partial C_{ke}^{\langle t\rangle}}{\partial \hat{C}^{\langle t\rangle}} = \Gamma_{\mu_{ij}}^{\langle t\rangle}\,\delta_{k,i}\,\delta_{e,j}$$

$$\frac{\partial h}{\partial \Gamma_{o_{ij}}^{\langle t\rangle}} = \sum_{k=1}^{n_a}\sum_{e=1}^{M}\frac{\partial h}{\partial a_{ke}^{\langle t\rangle}}\cdot\frac{\partial a_{ke}^{\langle t\rangle}}{\partial \Gamma_{o_{ij}}^{\langle t\rangle}} = \sum_{k=1}^{n_a}\sum_{e=1}^{M}\frac{\partial h}{\partial a_{ke}^{\langle t\rangle}}\,\delta_{e,j}\cdot\tanh\left(C_{ke}^{\langle t\rangle}\right)\cdot\delta_{k,i}$$

$$= \frac{\partial h}{\partial a_{ij}^{\langle t\rangle}}\tanh\left(C_{ij}^{\langle t\rangle}\right)$$

$$\frac{\partial h}{\partial C_{ij}^{\langle t\rangle}} = \left(\sum_{k=1}^{n_a}\sum_{e=1}^{M}\frac{\partial h}{\partial a_{ke}^{\langle t\rangle}}\cdot\frac{\partial a_{ke}^{\langle t\rangle}}{\partial G_{ij}^{\langle t\rangle}}\right) + \frac{\partial h}{\partial G_{ij}^{\langle t\rangle}} =$$

$$= \frac{\partial h}{\partial a_{ij}^{\langle t\rangle}}\cdot\Gamma_{o_{ij}}^{\langle t\rangle}\left(1-\tanh^2\left(C_{ij}^{\langle t\rangle}\right)\right) + \frac{\partial h}{\partial G_{ij}^{\langle t\rangle}}$$

$$\frac{\partial h}{\partial \Gamma_{f_{ij}}^{\langle t\rangle}} = \sum_{k=1}^{n_a}\sum_{e=1}^{M}\frac{\partial h}{\partial G_{ke}^{\langle t\rangle}}\cdot\frac{\partial G_{ke}^{\langle t\rangle}}{\partial \Gamma_{f_{ij}}^{\langle t\rangle}} = \sum_{k=1}^{n_a}\sum_{e=1}^{M}\frac{\partial h}{\partial G_{ke}^{\langle t\rangle}}\cdot C_{ke}^{\langle t-1\rangle}\,\delta_{k,i}\,\delta_{e,j}$$

$$= \frac{\partial h}{\partial G_{ij}^{\langle t\rangle}}\cdot C_{ij}^{\langle t-1\rangle} =$$

$$\left(\frac{\partial h}{\partial a_{ij}^{\langle t\rangle}} * \Gamma_{o_{ij}}^{\langle t\rangle} * \left(1-\tanh^2\left(C_{ij}^{\langle t\rangle}\right)\right) + \frac{\partial h}{\partial G_{ij}^{\langle t\rangle}}\right) * C_{ij}^{\langle t-1\rangle}$$

$$\frac{\partial h}{\partial C_{ij}^{u}\langle t-1\rangle} = \sum_{k=1}^{n_c}\sum_{c=1}^{M} \frac{\partial h}{\partial C_{kc}\langle t\rangle} \cdot \Gamma_{fkc}\langle t\rangle \cdot \delta_{ki}\,\delta_{cj} =$$

$$= \left(\frac{\partial h}{\partial a_{ij}\langle t\rangle} * \Gamma_{oj}\langle t\rangle * (1-\tanh^2(C_{ij}\langle t\rangle)) + \frac{\partial h}{\partial C_{ij}\langle t\rangle}\right) \Gamma_{ij}\langle t\rangle$$

$$\frac{\partial h}{\partial \Gamma_{uij}\langle t\rangle} = \left(\frac{\partial h}{\partial a_{ij}\langle t\rangle} * \Gamma_{oj}\langle t\rangle * (1-\tanh^2(C_{ij}\langle t\rangle)) + \frac{\partial h}{\partial C_{ij}\langle t\rangle}\right) * \hat{C}_{ij}\langle t\rangle$$

$$\frac{\partial h}{\partial \hat{C}_{ij}\langle t\rangle} = \left(\frac{\partial h}{\partial a_{ij}\langle t\rangle} * \Gamma_{oj}\langle t\rangle * (1-\tanh^2(C_{ij}\langle t\rangle)) + \frac{\partial h}{\partial C_{ij}\langle t\rangle}\right) * \Gamma_{uij}\langle t\rangle$$

$\longleftrightarrow$

$$\Gamma = \sigma(z) \;;\; z = WX \longrightarrow \frac{\partial \Gamma_{kc}}{\partial z_{ij}} = \Gamma_{kc}(1-\Gamma_{kc})\,\delta_{ki}\,\delta_{cj}$$

$$\frac{\partial z_{kc}}{\partial W_{ij}} = \delta_{ki}\, X_{jc} \longrightarrow \frac{\partial \Gamma_{kc}}{\partial W_{ij}} = \sum_{r=1}^{n_x}\sum_{s=1}^{M} \frac{\partial \Gamma_{kc}}{\partial z_{rs}} \cdot \frac{\partial z_{rs}}{\partial W_{ij}}$$

$$\frac{\partial \Gamma_{kc}}{\partial W_{ij}} = \sum_{r=1}^{n_x}\sum_{s=1}^{M} \Gamma_{kc}(1-\Gamma_{kc})\,\delta_{ki}\,\delta_{cj} \cdot \delta_{ki}\, X_{js} = \delta_{ki}\,\Gamma_{kc}(1-\Gamma_{kc})\,X_{jc}$$

$\longleftrightarrow$

$$\frac{\partial h}{\partial W_{fij}} = \sum_{k=1}^{n_c}\sum_{c=1}^{M} \frac{\partial h}{\partial \Gamma_{fkc}} \frac{\partial \Gamma_{fkc}}{\partial W_{fij}} = \sum_{k=1}^{n_c}\sum_{c=1}^{M} \frac{\partial h}{\partial \Gamma_{fkc}} \cdot \Gamma_{kc}(1-\Gamma_{kc})\,X_{jc}\,\delta_{ki}$$

$$= \sum_{c=1}^{M} \frac{\partial h}{\partial \Gamma_{fic}}\,\Gamma_{fic}(1-\Gamma_{fic})\,X_{jc} \;,\longrightarrow\; X = [a\langle t-1\rangle, x\langle t\rangle]$$

$$\frac{\partial h}{\partial W_{uij}} = \sum_{c=1}^{M} \frac{\partial h}{\partial \Gamma_{uic}} \cdot \Gamma_{uic}(1-\Gamma_{uic})\,[a\langle t-1\rangle, x\langle t\rangle]_{jc}$$

$$\frac{\partial h}{\partial W_{o\,i,j}} = \sum_{c=1}^{m} \frac{\partial h}{\partial \Gamma_{o\,i,c}^{<t>}} \cdot \Gamma_{o\,i,c}^{<t>} \left(1 - \Gamma_{o\,i,c}^{<t>}\right) \left[a^{<t-1>}, x^{<t>}\right]_{j,c}$$

$$\frac{\partial h}{\partial W_{c\,i,j}} = \sum_{c=1}^{m} \frac{\partial h}{\partial \hat{c}_{i,c}^{<t>}} \left(1 - \tanh^2\left(W_c \left[a^{<t-1>}, x^{<t>}\right] + b_c\right)_{i,c}\right) \left[a^{<t-1>}, x^{<t>}\right]_{j,c}$$

$\longleftrightarrow$

$$z = Wx \longrightarrow \frac{\partial z_{k,c}}{\partial x_{i,j}} = \delta_{c,j} W_{k,i} \longrightarrow \frac{\partial \Gamma_{k,c}}{\partial x_{i,j}} = \delta_{c,j} \Gamma_{k,c} \left(1 - \Gamma_{k,c}\right) W_{k,i}$$

$\longleftrightarrow$

$$\frac{\partial h}{\partial \left[a^{<t-1>}, x^{<t>}\right]} = \sum_{k=1}^{n_c} \sum_{c=1}^{m} \left(\frac{\partial h}{\partial \Gamma_{f\,k,c}} \frac{\partial \Gamma_{k,c}}{\partial \left[a^{<t-1>}, x^{<t>}\right]_{i,j}} + \frac{\partial h}{\partial \Gamma_{u\,k,c}^{<t>}} \frac{\partial \Gamma_{u\,k,c}^{<t>}}{\partial \left[a^{<t-1>}, x^{<t>}\right]_{i,j}} \right.$$

$$\left. + \frac{\partial h}{\partial \Gamma_{o\,k,c}^{<t>}} \frac{\partial \Gamma_{o\,k,c}^{<t>}}{\partial \left[a^{<t-1>}, x^{<t>}\right]_{i,j}} + \frac{\partial h}{\partial \hat{c}_{k,c}^{<t>}} \frac{\partial \hat{c}_{k,c}^{<t>}}{\partial \left[a^{<t-1>}, x^{<t>}\right]_{i,j}} \right)$$

↑ concatenated, for the gradients, just update until index element.

$$= \sum_{k=1}^{n_c} \sum_{c=1}^{m} \delta_{c,j} \left[\frac{\partial h}{\partial \Gamma_{f\,k,c}} \Gamma_{f\,k,c} \left(1 - \Gamma_{f\,k,c}\right) W_{f\,k,i} + \frac{\partial h}{\partial \Gamma_{u\,k,c}^{<t>}} \Gamma_{u\,k,c}^{<t>} \left(1 - \Gamma_{u\,k,c}^{<t>}\right) W_{u\,k,i} + \frac{\partial h}{\partial \Gamma_{o\,k,c}^{<t>}} \Gamma_{o\,k,c}^{<t>} \left(1 - \Gamma_{o\,k,c}^{<t>}\right) W_{o\,k,i} \right.$$

$$\left. + \frac{\partial h}{\partial \hat{c}_{k,c}^{<t>}} \left(1 - \tanh^2\left(W_c \left[a^{<t-1>}, x^{<t>}\right] + b_c\right)_{k,c}\right) W_{c\,k,i} \right]$$

$$= \sum_{k=1}^{n_c} \left[\frac{\partial h}{\partial \Gamma_{f\,k,j}} \Gamma_{f\,k,j} \left(1 - \Gamma_{f\,k,j}\right) W_{f\,k,i} + \frac{\partial h}{\partial \Gamma_{u\,k,j}} \Gamma_{u\,k,j} \left(1 - \Gamma_{u\,k,j}\right) W_{u\,k,i} \right.$$

$$\left. + \frac{\partial h}{\partial \Gamma_{o\,k,j}} \Gamma_{o\,k,j} \left(1 - \Gamma_{o\,k,j}\right) W_{o\,k,i} + \frac{\partial h}{\partial \hat{c}_{k,j}} \left(1 - \tanh^2\left(W_c \left[a^{<t-1>}, x^{<t>}\right]_{k,j} + b_c\right)\right) W_{c\,k,i} \right]$$

- Recommended implementation.

$$d\Gamma_o^{\langle t\rangle} = da_{next} * \tanh(c_{next})$$

$$d\Gamma_f^{\langle t\rangle} = \left(da_{next} * \Gamma_o^{\langle t\rangle} * \left(1 - \tanh^2(c_{next})\right) + dc_{next}\right) * c^{\langle t-1\rangle}$$

($c^{\langle t-1\rangle}$: c_{prev})

$$d\Gamma_u^{\langle t\rangle} = \left(da_{next} * \Gamma_o^{\langle t\rangle} * \left(1 - \tanh^2(c_{next})\right) + dc_{next}\right) * \hat{c}^{\langle t\rangle}$$

$$d\hat{c}^{\langle t\rangle} = \left(da_{next} * \Gamma_o^{\langle t\rangle} * \left(1 - \tanh^2(c_{next})\right) + dc_{next}\right) * \Gamma_u^{\langle t\rangle}$$

(dot product)

$$dW_f = \left[d\Gamma_f^{\langle t\rangle} * \Gamma_f^{\langle t\rangle} * \left(1 - \Gamma_f^{\langle t\rangle}\right)\right] \cdot \begin{pmatrix} a_{prev} \\ x^{\langle t\rangle} \end{pmatrix}^T$$

$$dW_o = \left[d\Gamma_o^{\langle t\rangle} * \Gamma_o^{\langle t\rangle} * \left(1 - \Gamma_o^{\langle t\rangle}\right)\right] \cdot \begin{pmatrix} a_{prev} \\ x^{\langle t\rangle} \end{pmatrix}^T$$

$$dW_u = \left[d\Gamma_u^{\langle t\rangle} * \Gamma_u^{\langle t\rangle} * \left(1 - \Gamma_u^{\langle t\rangle}\right)\right] \cdot \begin{pmatrix} a_{prev} \\ x^{\langle t\rangle} \end{pmatrix}^T$$

$$dW_c = \left[d\tilde{c}^{\langle t\rangle} * \left(1 - \underbrace{\tanh^2\left(W_c\left[a^{\langle t-1\rangle}, x^{\langle t\rangle}\right] + b_c\right)}_{\tilde{c}^{\langle t\rangle}}\right)\right] \begin{pmatrix} a_{prev} \\ x^{\langle t\rangle} \end{pmatrix}^T$$

$$d\begin{bmatrix} a^{\langle t-1\rangle} \\ x^{\langle t\rangle} \end{bmatrix} = W_f^T\left[d\Gamma_f^{\langle t\rangle} * \Gamma_f^{\langle t\rangle} * \left(1 - \Gamma_f^{\langle t\rangle}\right)\right]$$

$$+ W_o^T\left[d\Gamma_o^{\langle t\rangle} * \Gamma_o^{\langle t\rangle} * \left(1 - \Gamma_o^{\langle t\rangle}\right)\right]$$

$$* + W_u^T\left[d\Gamma_u^{\langle t\rangle} * \Gamma_u^{\langle t\rangle} * \left(1 - \Gamma_u^{\langle t\rangle}\right)\right]$$

$$+ W_c^T\left[d\hat{c}^{\langle t\rangle} * \left(1 - \tanh^2\left(W_c\left[a^{\langle t-1\rangle}, x^{\langle t\rangle}\right] + b_c\right)\right)\right]$$

$$dc_{prev} = \left(da_{next} * \Gamma_o^{\langle t\rangle} * \left(1 - \tanh^2(c_{next})\right) + dc_{next}\right) * \Gamma_f^{\langle t\rangle}$$