* RNN Backward propagation ⟶

$$a^{\langle t \rangle} = \tanh(z^{\langle t \rangle})$$

$$z^{\langle t \rangle} = W_{ax}\, x^{\langle t \rangle} + W_{aa}\, a^{\langle t-1 \rangle} + b_a$$

$$\frac{\partial a_{rs}^{\langle t \rangle}}{\partial z_{ij}^{\langle t \rangle}} = (1 - \tanh^2(z_{rs}))\, \delta_{r,i}\, \delta_{s,j}$$

$$\frac{\partial z_{rs}^{\langle t \rangle}}{\partial W_{ax_{ij}}} = \delta_{r,i}\, x_{js} \quad ; \quad \frac{\partial z_{rs}^{\langle t \rangle}}{\partial x_{ij}^{\langle t \rangle}} = \delta_{s,j}\, W_{r,i}$$

$$\frac{\partial a_{ke}^{\langle t \rangle}}{\partial W_{ax_{ij}}} = \sum_{r=1}^{n_a} \sum_{s=1}^{M} \frac{\partial a_{ke}^{\langle t \rangle}}{\partial z_{rs}^{\langle t \rangle}} \frac{\partial z_{rs}^{\langle t \rangle}}{\partial W_{ax_{ij}}} =$$

$$\sum_{r=1}^{n_a} \sum_{s=1}^{M} (1 - \tanh^2(z_{ke}))\, \delta_{r,k}\, \delta_{e,s} \cdot \delta_{r,i}\, x_{j,s} =$$

$$= (1 - \tanh^2(z_{ke}))\, x_{jl}\, \delta_{k,i}$$

$$\frac{\partial a_{ke}^{\langle t \rangle}}{\partial x_{ij}^{\langle t \rangle}} = \sum_{r=1}^{n_a} \sum_{s=1}^{M} \frac{\partial a_{ke}^{\langle t \rangle}}{\partial z_{rs}^{\langle t \rangle}} \cdot \frac{\partial z_{rs}^{\langle t \rangle}}{\partial x_{ij}^{\langle t \rangle}} =$$

$$\sum_{r=1}^{n_a} \sum_{s=1}^{M} (1 - \tanh^2(z_{ke}))\, \delta_{k,r}\, \delta_{e,s} \cdot W_{ri}\, \delta_{s,j} =$$

$$= (1 - \tanh^2(z_{ke}))\, W_{ki}\, \delta_{e,j}$$

$$\frac{\partial L}{\partial Wax_{ij}} = \sum_{o=1}^{n_a} \sum_{p=1}^{M} \frac{\partial L}{\partial a_{op}^{<t>}} \cdot \frac{\partial a_{op}^{<t>}}{\partial Wax_{ij}} = \sum_{o=1}^{n_a} \sum_{p=1}^{M} \frac{\partial L}{\partial a_{op}^{<t>}} \cdot (1 - \tanh^2(z_{op}))\, x_{jp} \cdot \delta_{o,i}$$

$$= \sum_{p=1}^{M} \frac{\partial L}{\partial a_{ip}^{<t>}} (1 - \tanh^2(z_{ip}))\, x_{jp} \quad \Longrightarrow \quad dWax = (dA * dz) \odot X.T$$

$$\frac{\partial L}{\partial x_{ij}^{<t>}} = \sum_{o=1}^{n_a} \sum_{p=1}^{M} \frac{\partial L}{\partial a_{op}^{<t>}} \cdot \frac{\partial a_{op}^{<t>}}{\partial x_{ij}^{<t>}} = \sum_{o=1}^{n_a} \sum_{p=1}^{M} \frac{\partial L}{\partial a_{op}^{<t>}} \cdot (1 - \tanh^2(z_{op}))\, W_{oi}\, \delta_{pj}$$

$$= \sum_{o=1}^{n_a} (1 - \tanh^2(z_{oj}))\, W_{oi} \frac{\partial L}{\partial a_{oj}^{<t>}} \quad \Longrightarrow \quad dX = W.T \odot (dA * dz)$$

$a^{<t>}, a^{<t-1>}, x^{<t>}, W, b$