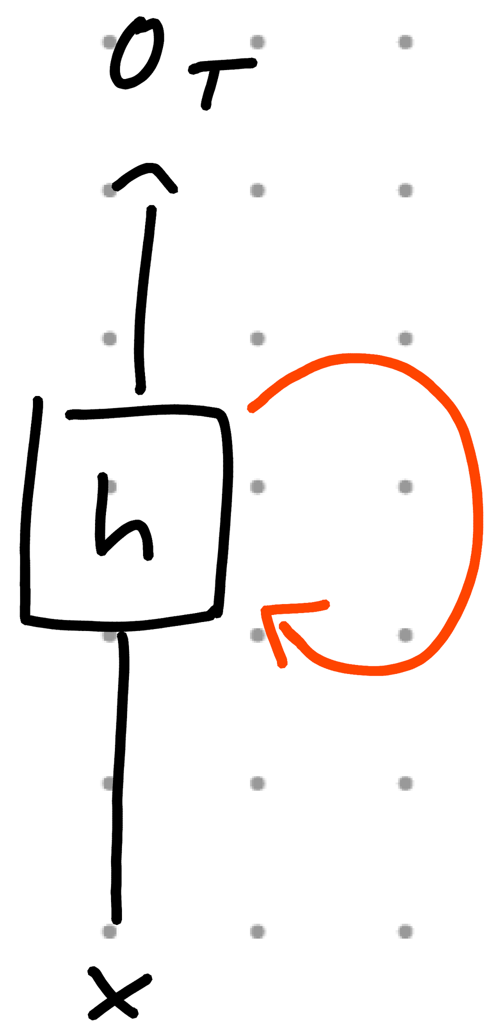

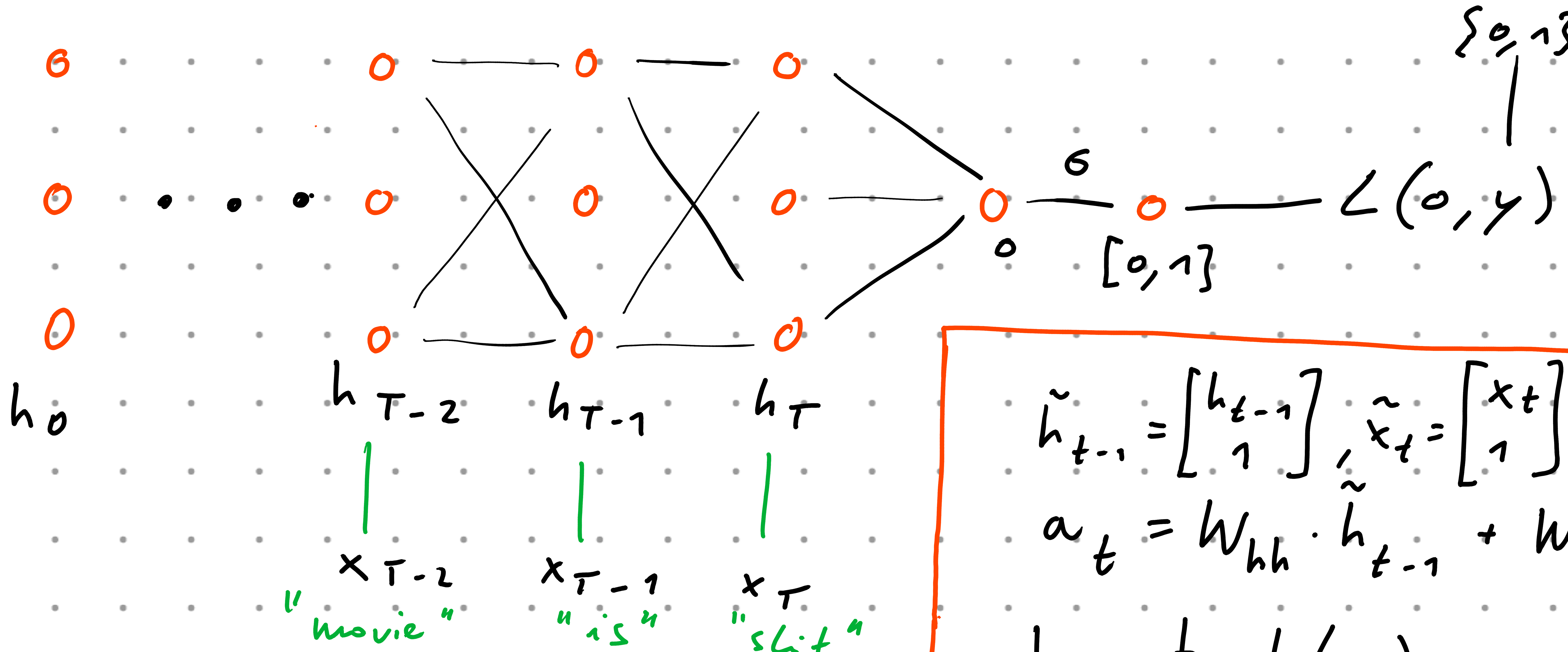

$$\tilde{h}_{t-1} = \begin{bmatrix} h_{t-1} \\ 1 \end{bmatrix}, \quad \tilde{x}_t = \begin{bmatrix} x_t \\ 1 \end{bmatrix}$$

$$a_t = W_{hh} \cdot \tilde{h}_{t-1} + W_{hx} \cdot \tilde{x}_t ,$$

$$h_t = \tanh(a_t)$$

$$o_T = W_{oh} \cdot \tilde{h}_T$$

$$z_T = \sigma(o_T)$$

INFERENCE

$$L = -\left(y \cdot \log(z) + (1-y)\log(1-z)\right)$$

$$\frac{\partial L}{\partial o} = e_T = z_T - y_T$$

$$\boxed{\frac{\partial L}{\partial W_{oh}} = \frac{\partial L}{\partial o} \cdot \frac{\partial o}{\partial W_{oh}} = e_T \cdot \tilde{h}_T^T}$$

$$\frac{\partial L}{\partial h_T} = W_{oh}^T \cdot e_T$$

$$\delta_t = \frac{\partial L}{\partial a_t} = \frac{\partial L}{\partial h_t} \odot (1 - h_t \odot h_t)$$

For $t = T-1, \dots, 0$

$$\frac{\partial L}{\partial h_t} = (W_{hh})^T \cdot \delta_{t+1}$$

$$\boxed{\frac{\partial L}{\partial W_{hh}} = \sum_{t=1}^{T} \delta_t \cdot \tilde{h}_{t-1}^T}$$

$$\boxed{\frac{\partial L}{\partial W_{hx}} = \sum_{t=1}^{T} \delta_t \cdot \tilde{x}_{t-1}^T}$$