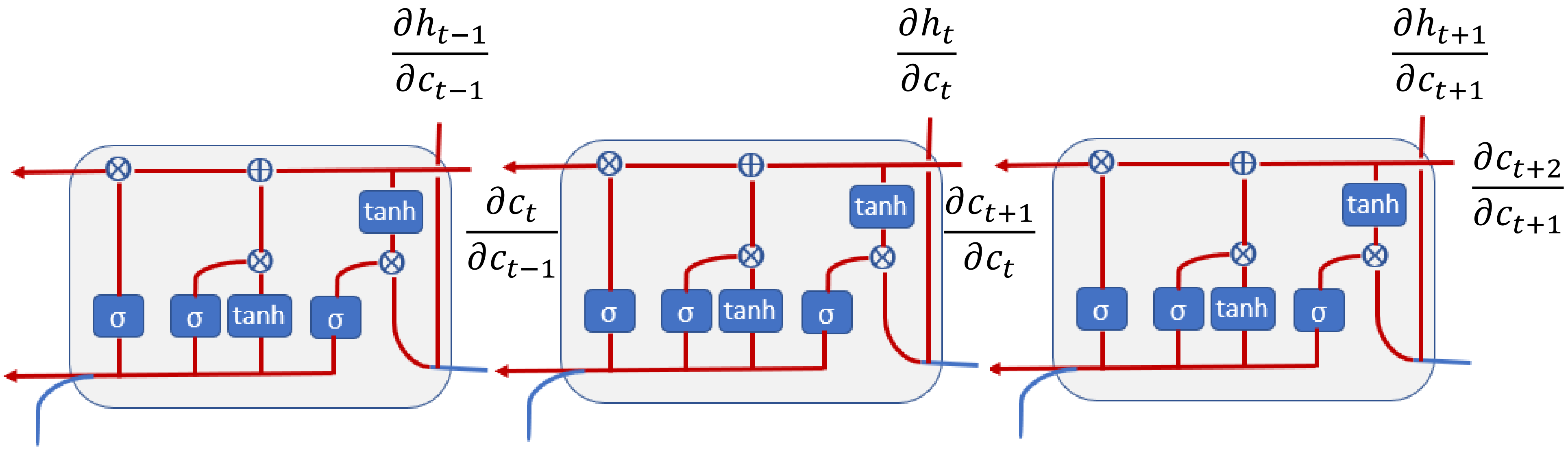
Backpropagation in LSTM

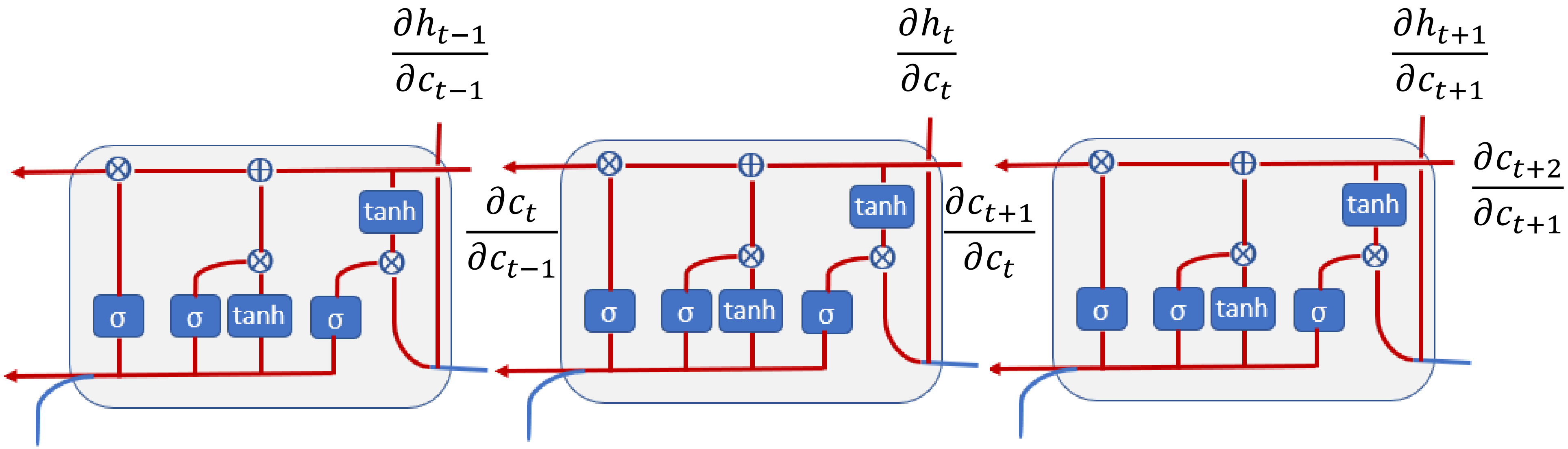


$$Z_t = \begin{bmatrix} \tilde{c}_t \\ f_t \\ i_t \\ o_t \end{bmatrix} = \begin{bmatrix} W_c & U_c \\ W_f & U_f \\ W_i & U_i \\ W_o & U_o \end{bmatrix} \cdot \begin{bmatrix} \boldsymbol{x_t} \\ \boldsymbol{h_{t-1}} \end{bmatrix} + \begin{bmatrix} b_c \\ b_f \\ b_i \\ b_o \end{bmatrix} = W \cdot \boldsymbol{I_t} + \boldsymbol{b}$$

Topics

- Bias
- RNN Architecture
- RNN Forward Pass
- RNN Backpropagation
- LSTM Architecture
- LSTM Forward Pass
- GRU Architecture
- LSTM Back propagation

Backpropagation in LSTM



$$\frac{\partial E}{\partial W} = \sum_{t=1}^{T} \frac{\partial E_t}{\partial W}$$

$$\frac{\partial E_t}{\partial W} = \frac{\partial E}{\partial h_t} \frac{\partial h_t}{\partial c_t} \frac{\partial c_t}{\partial c_{t-1}} \cdots \frac{\partial c_2}{\partial c_1} \frac{\partial c_1}{\partial W} = \frac{\partial E}{\partial h_t} \frac{\partial h_t}{\partial c_t} \left(\prod_{t=2}^{T} \frac{\partial c_t}{\partial c_{t-1}} \right) \frac{\partial c_1}{\partial W}$$

Backpropagation in LSTM

$$c_t = c_{t-1} \otimes f_t \oplus i_t \otimes \widetilde{c_t}$$

$$\frac{\partial c_t}{\partial c_{t-1}} = \frac{\partial}{\partial c_{t-1}} (c_{t-1} \otimes f_t \oplus i_t \otimes \widetilde{c_t})$$

$$= \frac{\partial}{\partial c_{t-1}} [c_{t-1} \otimes f_t] \oplus [i_t \otimes \widetilde{c_t}]$$

$$= \frac{\partial c_{t-1}}{\partial c_{t-1}} f_t + \frac{\partial f_t}{\partial c_{t-1}} c_{t-1} + \frac{\partial \widetilde{c_t}}{\partial c_{t-1}} i_t + \frac{\partial i_t}{\partial c_{t-1}} \widetilde{c_t}$$

$$= f_t + \frac{\partial f_t}{\partial c_{t-1}} c_{t-1} + \frac{\partial \widetilde{c_t}}{\partial c_{t-1}} i_t + \frac{\partial i_t}{\partial c_{t-1}} \widetilde{c_t}$$

Backpropagation in LSTM

$$c_t = c_{t-1} \otimes f_t \oplus i_t \otimes \widetilde{c_t}$$

$$\frac{\partial c_t}{\partial c_{t-1}} = \frac{\partial}{\partial c_{t-1}}(c_{t-1} \otimes f_t \oplus i_t \otimes \widetilde{c_t})$$

$$= \frac{\partial}{\partial c_{t-1}}[c_{t-1} \otimes f_t] \oplus [i_t \otimes \widetilde{c_t}]$$

$$= \frac{\partial c_{t-1}}{\partial c_{t-1}} f_t + \frac{\partial f_t}{\partial c_{t-1}} c_{t-1} + \frac{\partial \widetilde{c_t}}{\partial c_{t-1}} i_t + \frac{\partial i_t}{\partial c_{t-1}} \widetilde{c_t}$$

$$= f_t + \frac{\partial f_t}{\partial c_{t-1}} c_{t-1} + \frac{\partial \widetilde{c_t}}{\partial c_{t-1}} i_t + \frac{\partial i_t}{\partial c_{t-1}} \widetilde{c_t}$$

p q r s

$$f_t = \sigma\left(W_f \cdot x_t + U_f \cdot h_{t-1} + b_i\right)$$

$$i_t = \sigma(W_i \cdot x_t + U_i \cdot h_{t-1} + b_i)$$

$$\widetilde{c_t} = tanh(W_c \cdot x_t + U_c \cdot h_{t-1} + b_c)$$

$$h_t = tanh(c_t) \otimes o_t$$

$$\frac{\partial E_t}{\partial W} = \frac{\partial E}{\partial h_t}\frac{\partial h_t}{\partial c_t}\left(\prod_{t=2}^{T}(p+q+r+s)\right)\frac{\partial c_1}{\partial W}$$