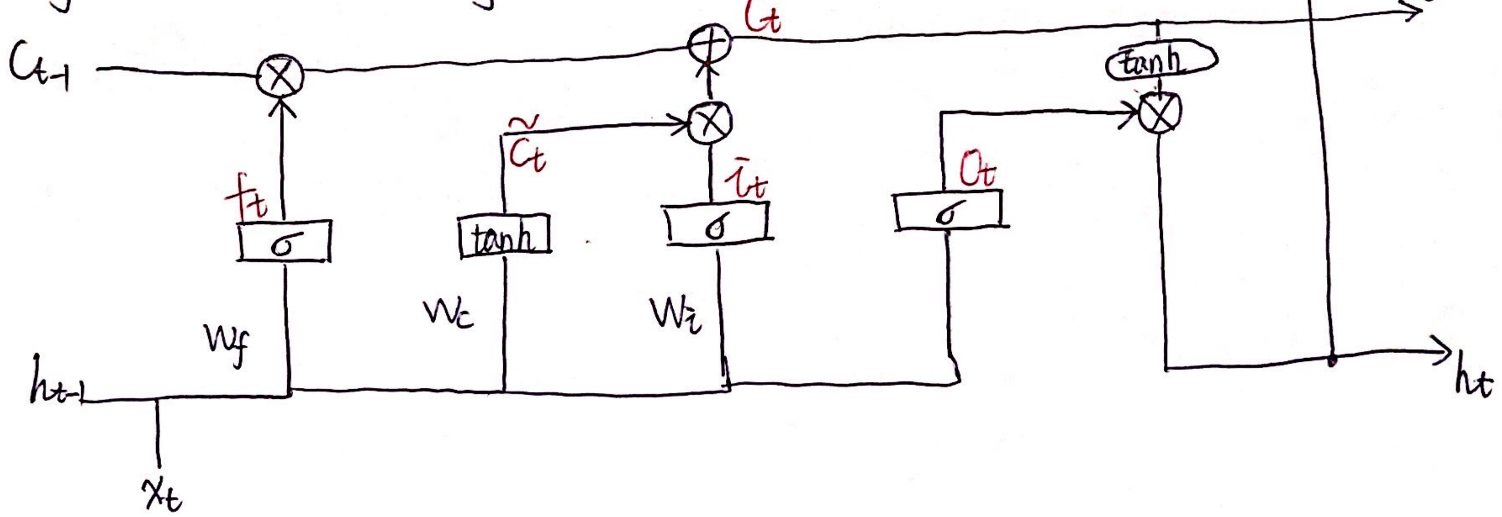
Long Short-Term Memory (LSTM)



C_t ~~ht~~ : knowledge encoded in C_t captures long-term dependencies and relations in the sequential ~~order~~ data

h_t : predictive vectors (hidden state)

3 gates: forget (f_t) / input (i_t) / output (O_t) gates

$f_t = \sigma(W_f \cdot [h_{t-1}, x_t])$ $\qquad$ $\tilde{C}_t = \tanh(W_c \cdot [h_{t-1}, x_t])$

$i_t = \sigma(W_i \cdot [h_{t-1}, x_t])$ $\qquad$ $C_t = f_t \otimes C_{t-1} \oplus i_t \otimes \tilde{C}_t$

$O_t = \sigma(W_o \cdot [h_{t-1}, x_t])$ $\qquad$ $h_t = O_t \otimes \tanh(C_t)$

BPTT:

$$\frac{\partial E_k}{\partial W} = \frac{\partial E_k}{\partial h_k} \cdot \frac{\partial h_k}{\partial C_k} \cdot \left(\prod_{t=2}^{k} \frac{\partial C_t}{\partial C_{t-1}}\right) \cdot \frac{\partial C_1}{\partial W} \quad ②$$

$$\frac{\partial C_t}{\partial C_{t-1}} = \frac{\partial}{\partial C_{t-1}}(C_{t-1} \otimes f_t) + \frac{\partial}{\partial C_{t-1}}(\tilde{C}_t \otimes i_t)$$

$$= \boxed{\frac{\partial f_t}{\partial C_{t-1}} C_{t-1}}_{①} + \boxed{\frac{\partial C_{t-1}}{\partial C_{t-1}} \cdot f_t}_{②} + \boxed{\frac{\partial i_t}{\partial C_{t-1}} \cdot \tilde{C}_t}_{③} + \boxed{\frac{\partial \tilde{C}_t}{\partial C_{t-1}} \cdot i_t}_{④}$$

$$= \sigma'(W_f \cdot [h_{t-1}, x_t]) \cdot W_f \cdot O_{t-1} \otimes \tanh'(C_{t-1}) \cdot C_{t-1} \quad ① = A$$

$$+ f_t \quad ② = B$$

$$+ \sigma'(W_i \cdot [h_{t-1}, x_t]) \cdot W_i \cdot O_{t-1} \otimes \tanh'(C_{t-1}) \cdot \tilde{C}_t \quad ③ = C$$

$$+ \tanh'(W_c \cdot [h_{t-1}, x_t]) \cdot W_c \cdot O_{t-1} \otimes \tanh'(C_{t-1}) \cdot i_t \quad ④ = D$$