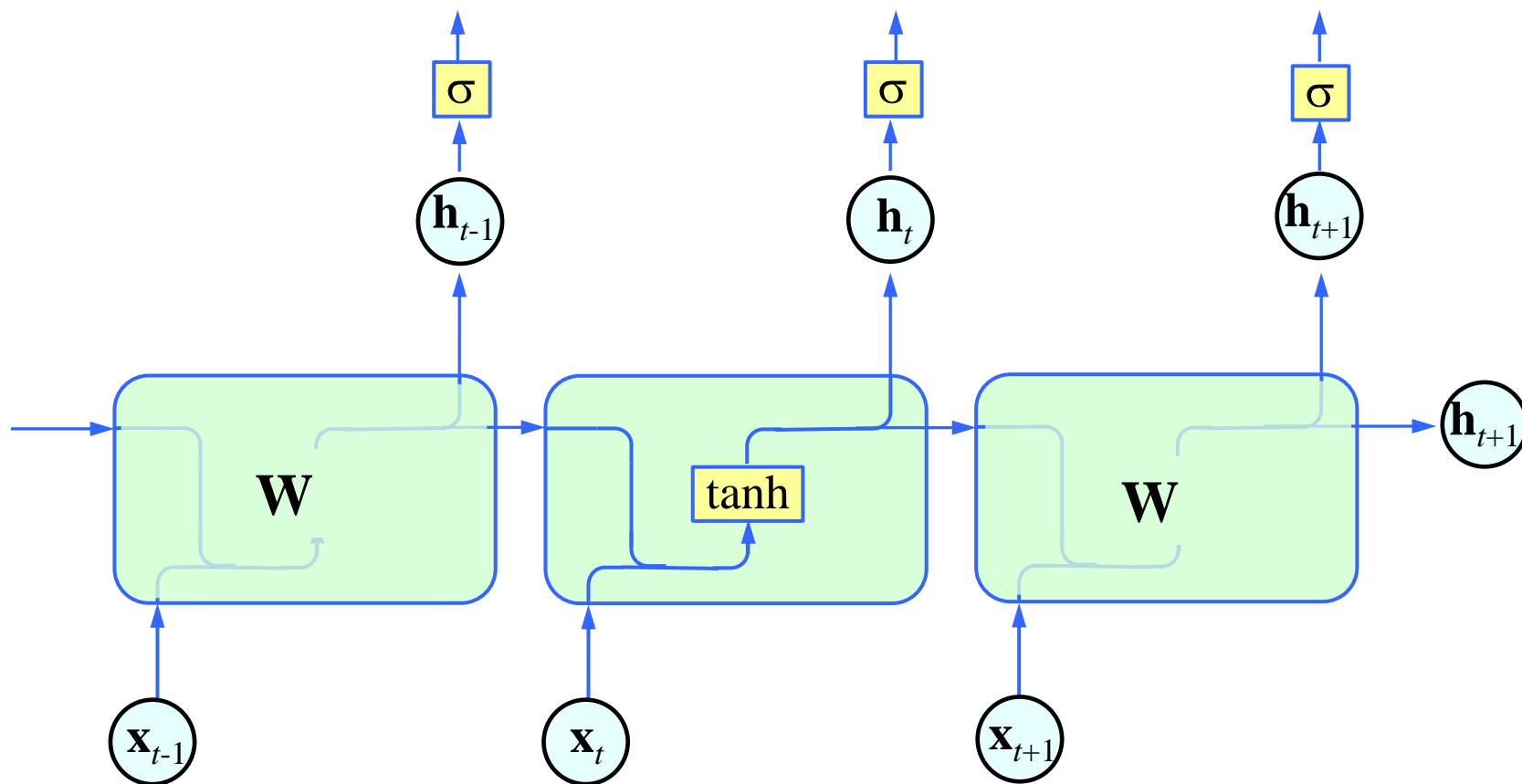


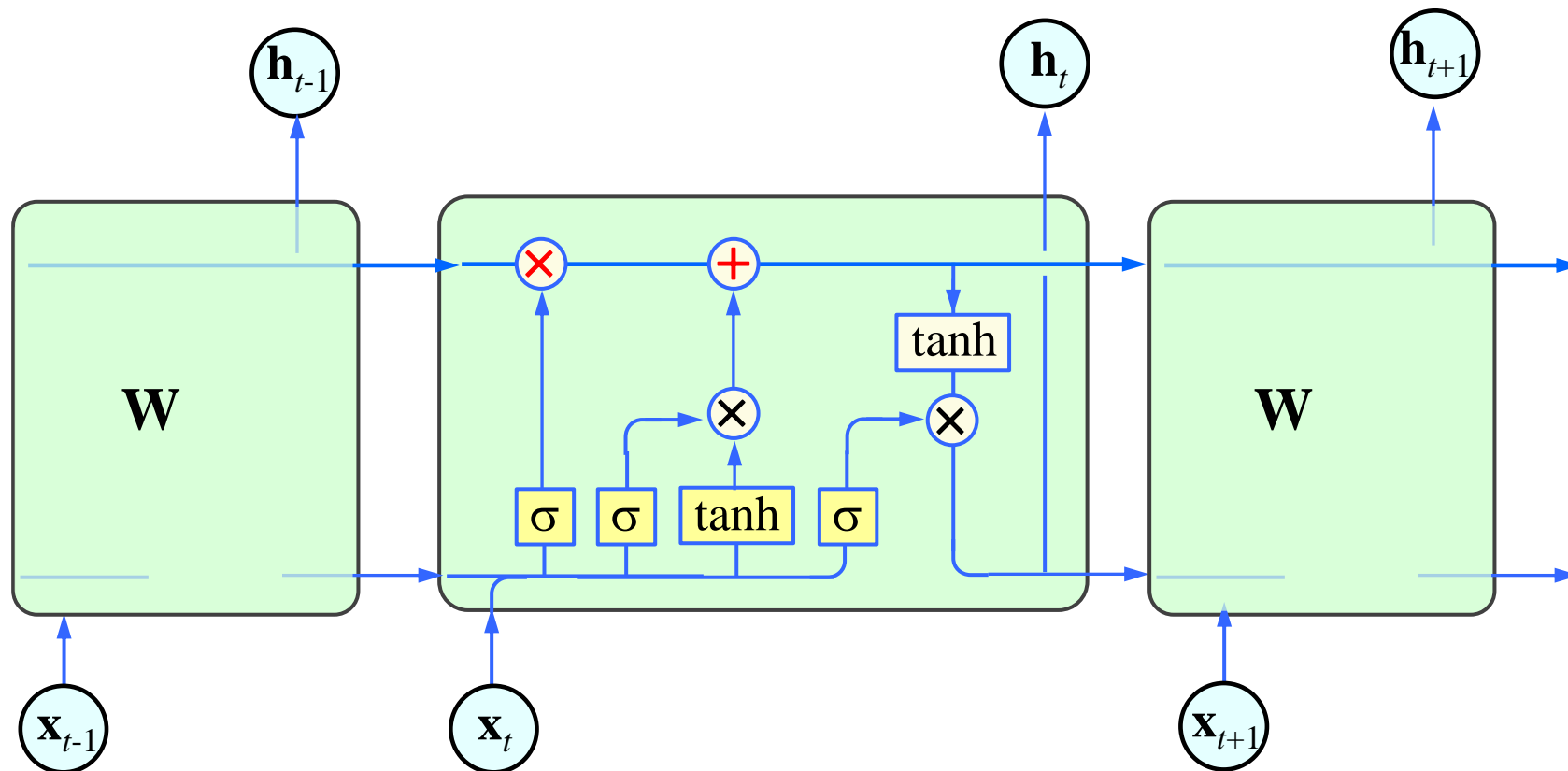
$$\frac{\partial L}{\partial \mathbf{s}_k} \approx \left(f'(\mathbf{s}) \mathbf{W} \right)^t \frac{\partial L}{\partial \mathbf{s}_{k+t}}$$

5.7 From RNN to LSTM:



标准 RNN 中的重复模块包含单一的层

5.7 From RNN to LSTM



LSTM 中的重复模块包含四个交互的层

■ : 神经网络层

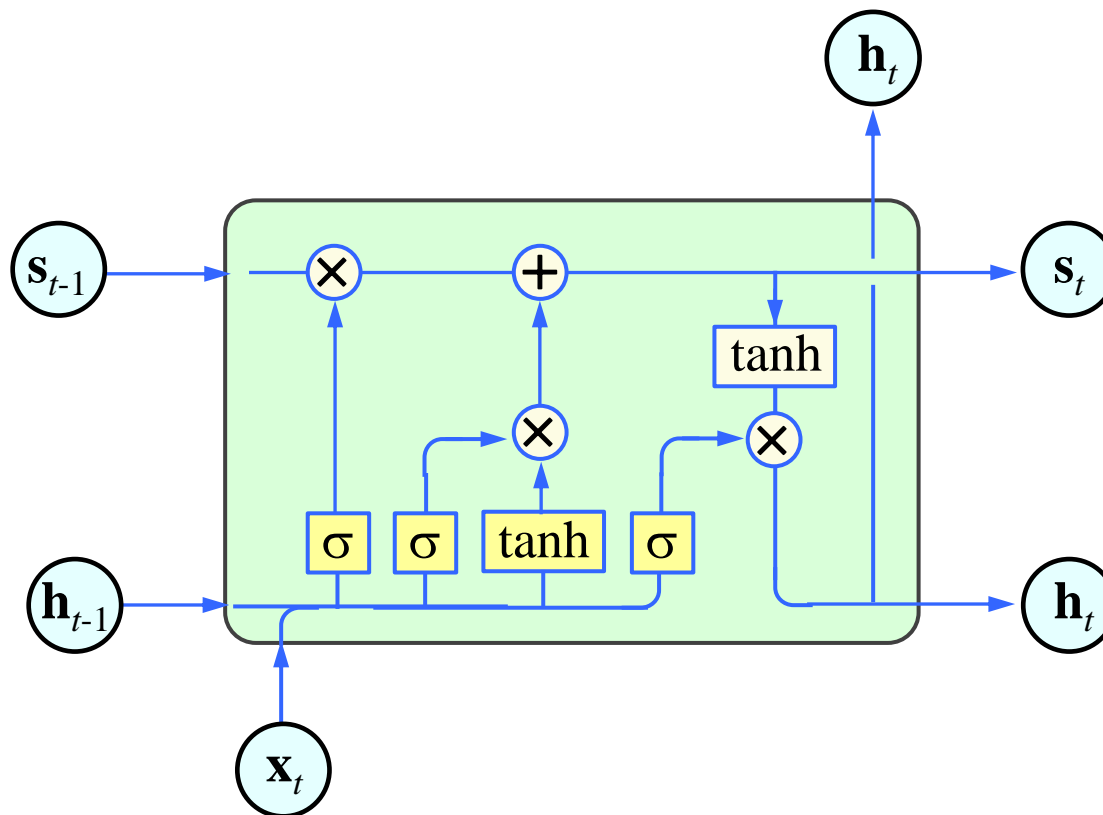
○ : pointwise operation

→ : vector transfer

⤵ : concatenate

⤴ : copy

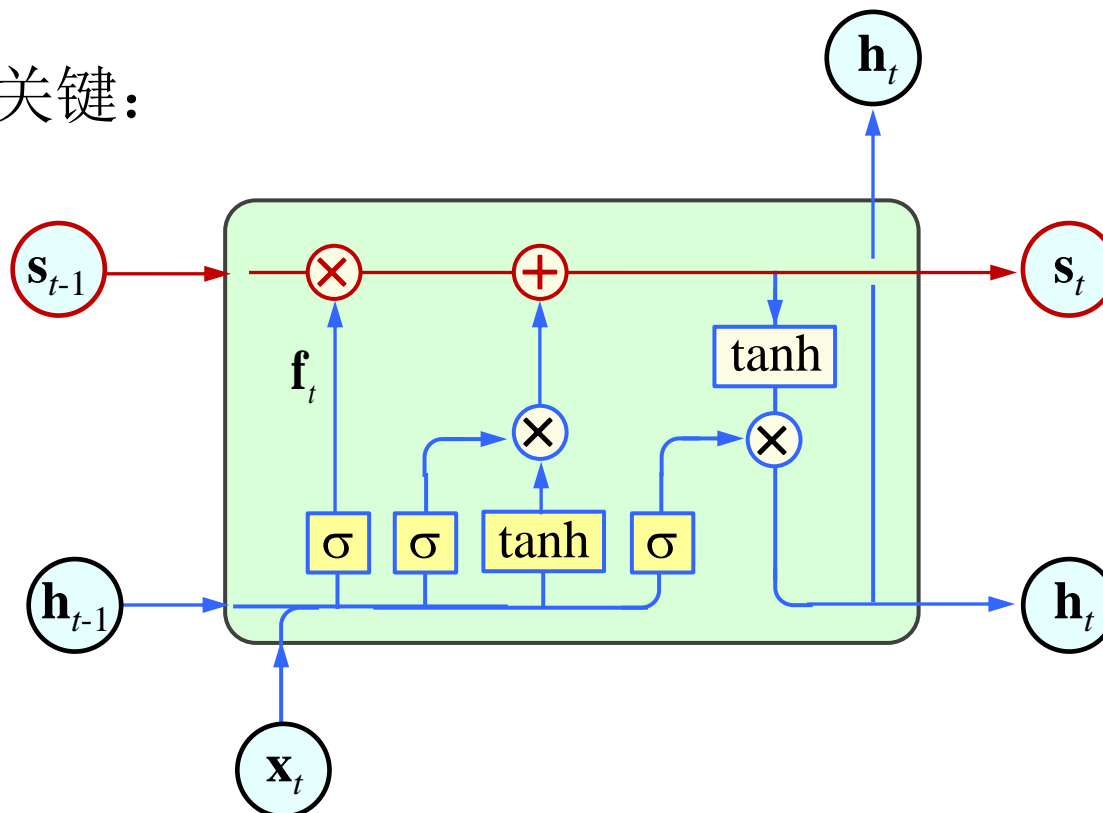
5.7 From RNN to LSTM



LSTM 中的一个时序单元的输入输出

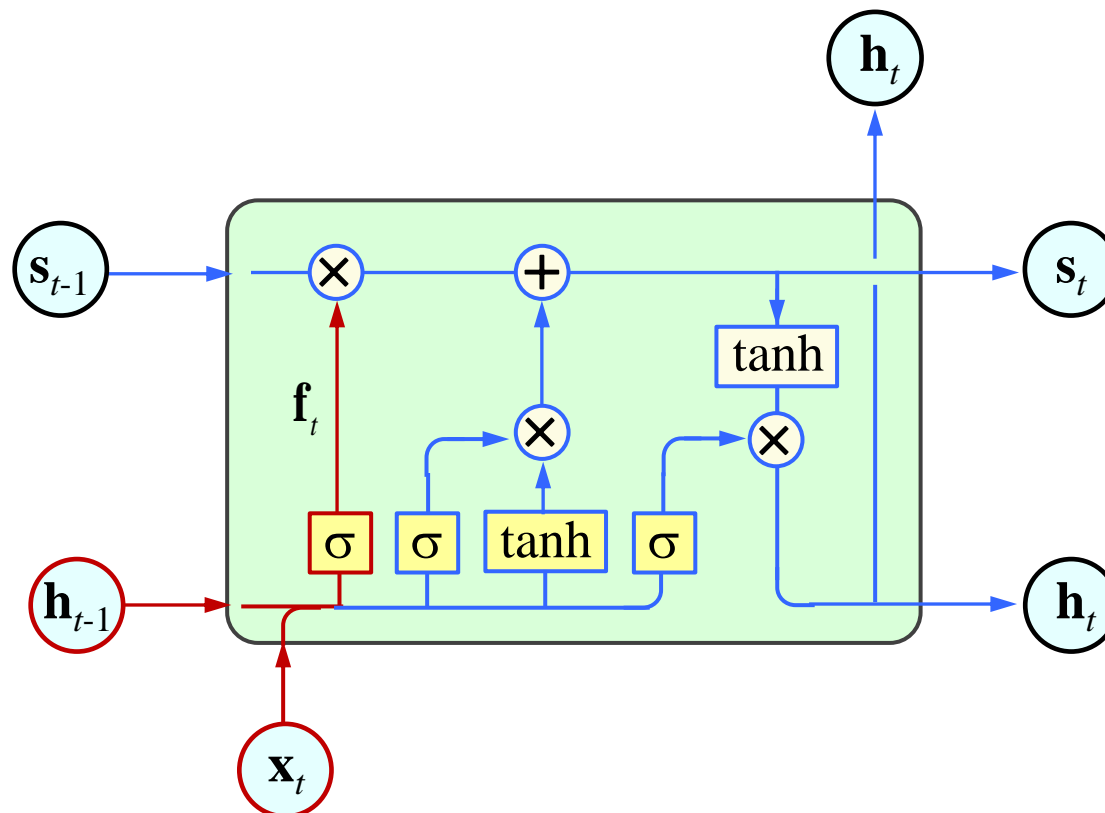
5.7 From RNN to LSTM

LSTM的关键：



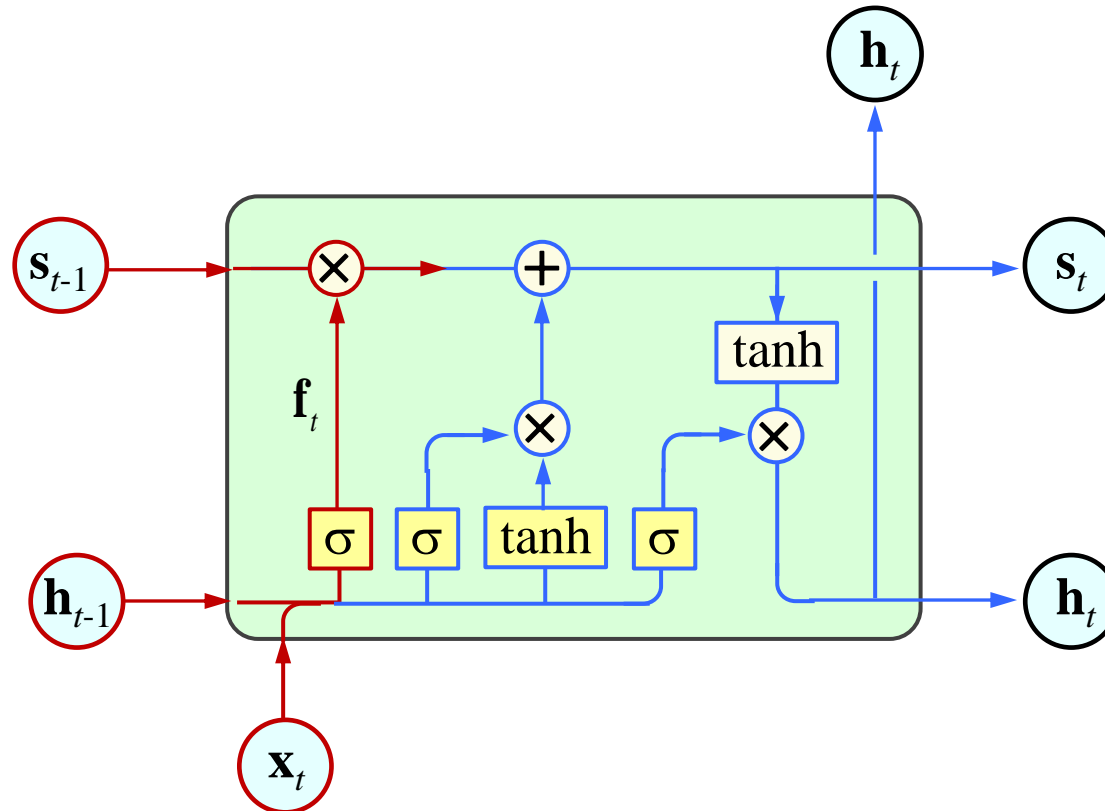
LSTM 的关键就是细胞状态，水平线在图上方贯穿运行。细胞状态类似于传送带。直接在整个链上运行，只有一些少量的线性交互。信息在上面流传保持不变会很容易。

5.7 From RNN to LSTM



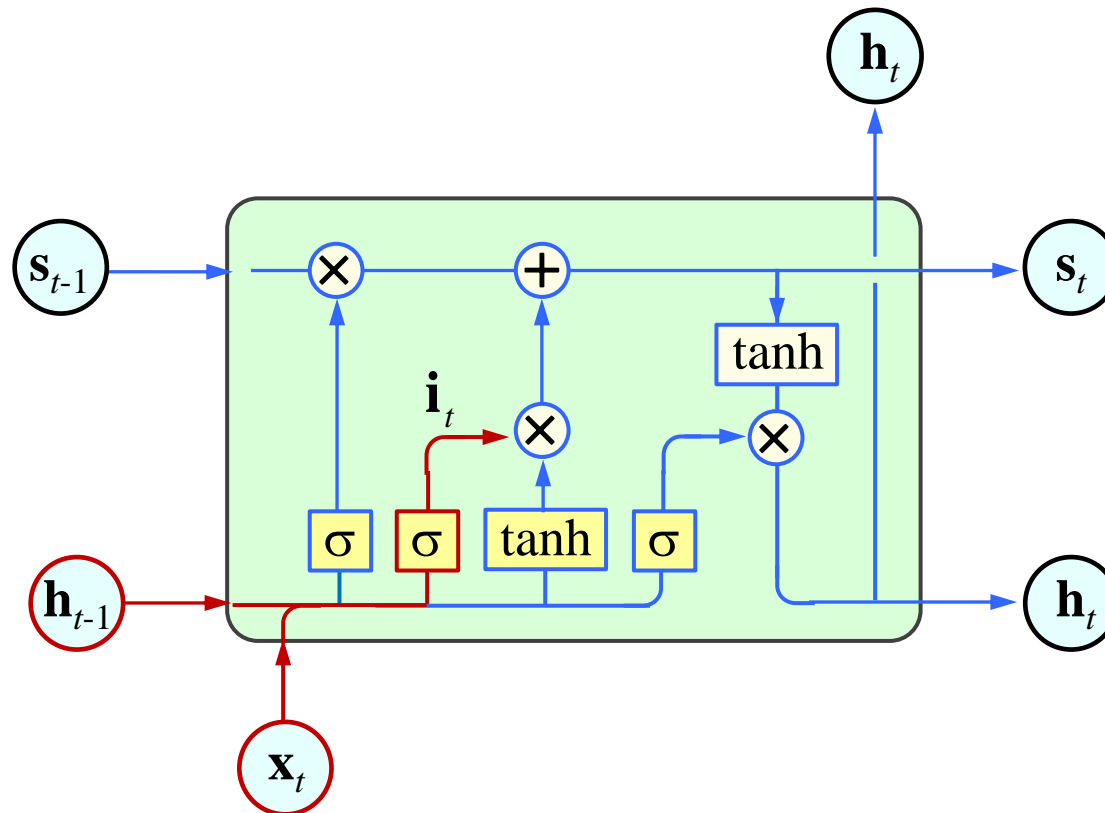
$$\mathbf{f}_t = \text{sigmoid}(\mathbf{b}_f + \mathbf{U}_f \mathbf{x}_t + \mathbf{W}_f \mathbf{h}_{t-1})$$

5.7 From RNN to LSTM



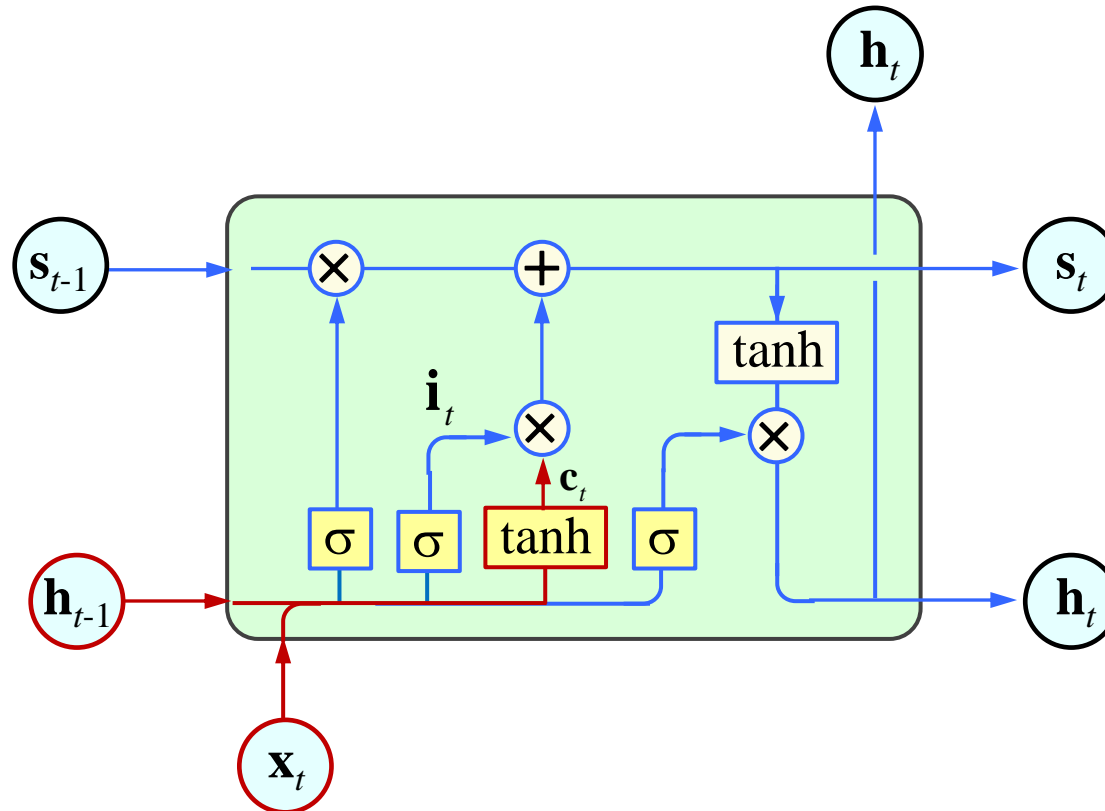
$$\mathbf{f}_t = \text{sigmod}(\mathbf{b}_f + \mathbf{U}_f \mathbf{x}_t + \mathbf{W}_f \mathbf{h}_{t-1}) \longrightarrow \mathbf{f}_t \otimes \mathbf{s}_{t-1}$$

5.7 From RNN to LSTM



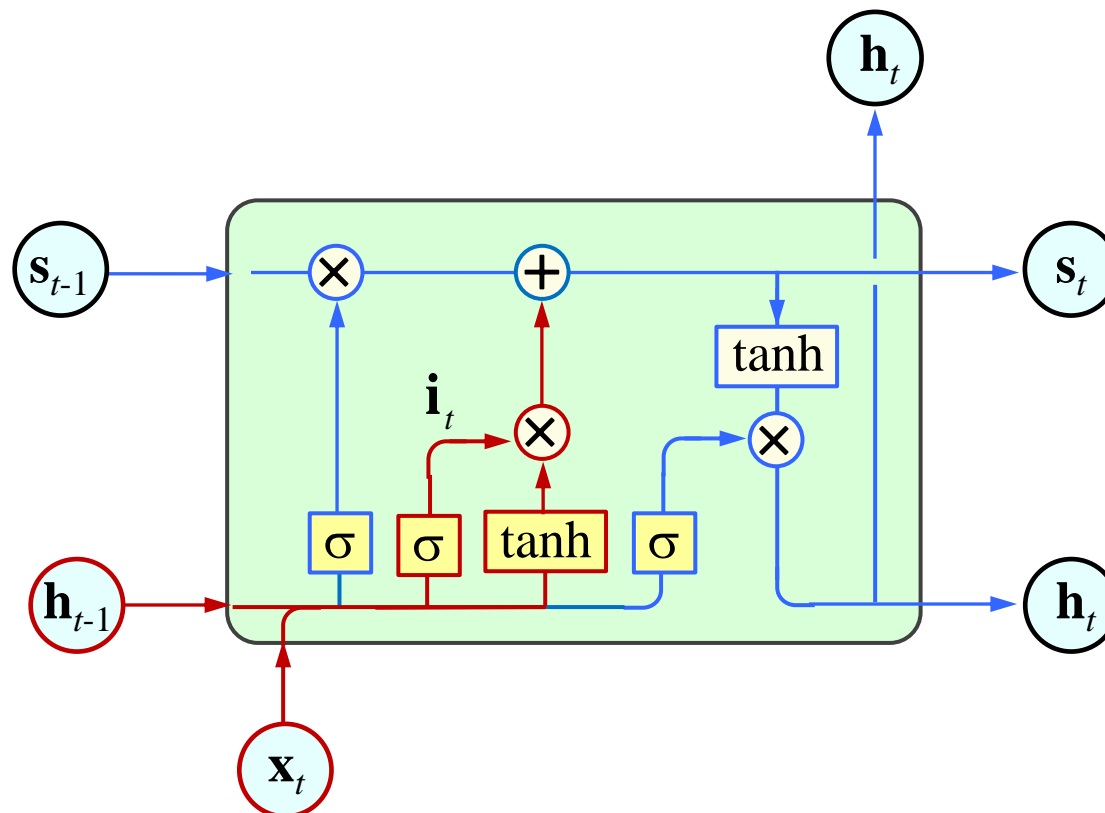
$$\mathbf{i}_t = \text{sigmoid}(\mathbf{b}_{in} + \mathbf{U}_{in} \mathbf{x}_t + \mathbf{W}_{in} \mathbf{h}_{t-1})$$

5.7 From RNN to LSTM



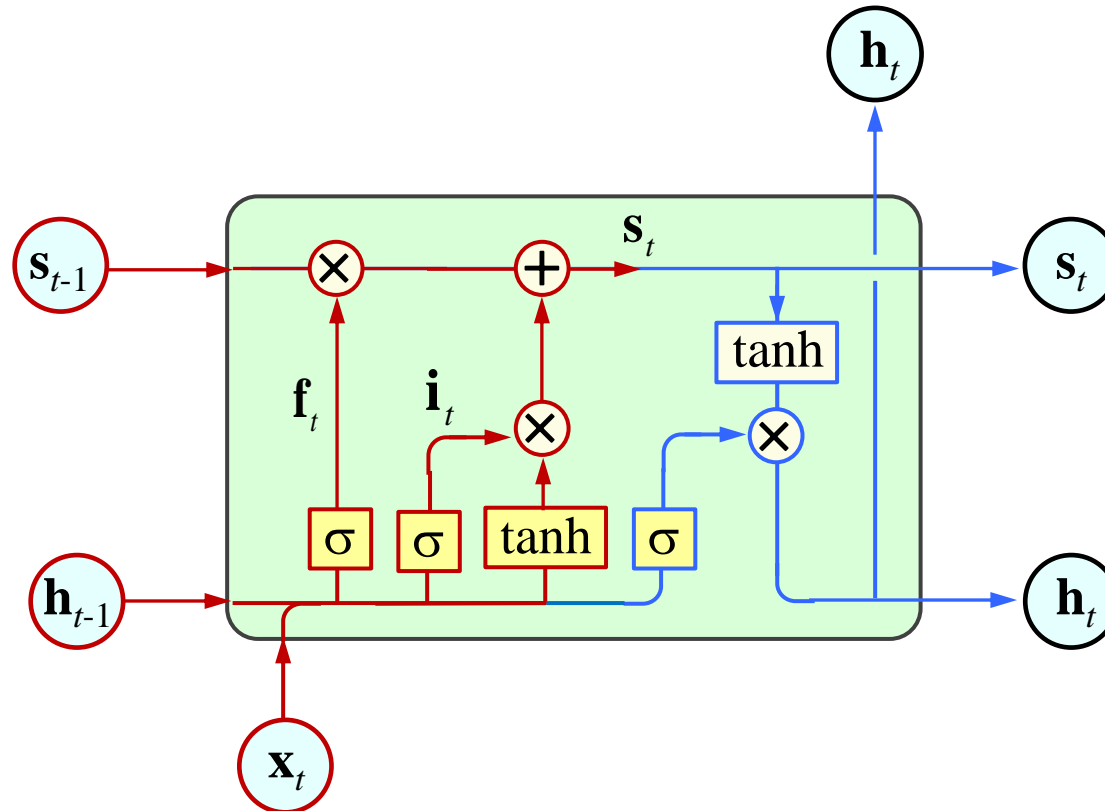
$$\mathbf{c}_t = \tanh(\mathbf{b} + \mathbf{U}\mathbf{x}_t + \mathbf{W}\mathbf{h}_{t-1})$$

5.7 From RNN to LSTM



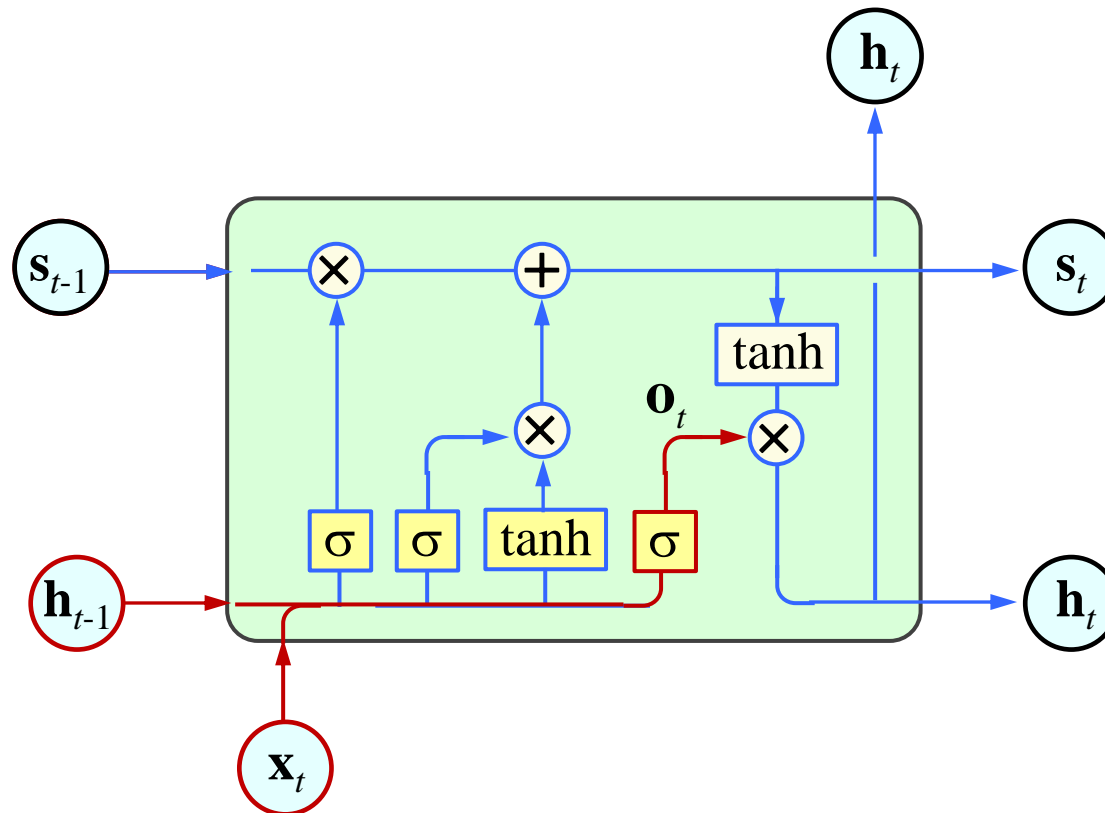
$$\mathbf{i}_t \otimes \mathbf{c}_t$$

5.7 From RNN to LSTM



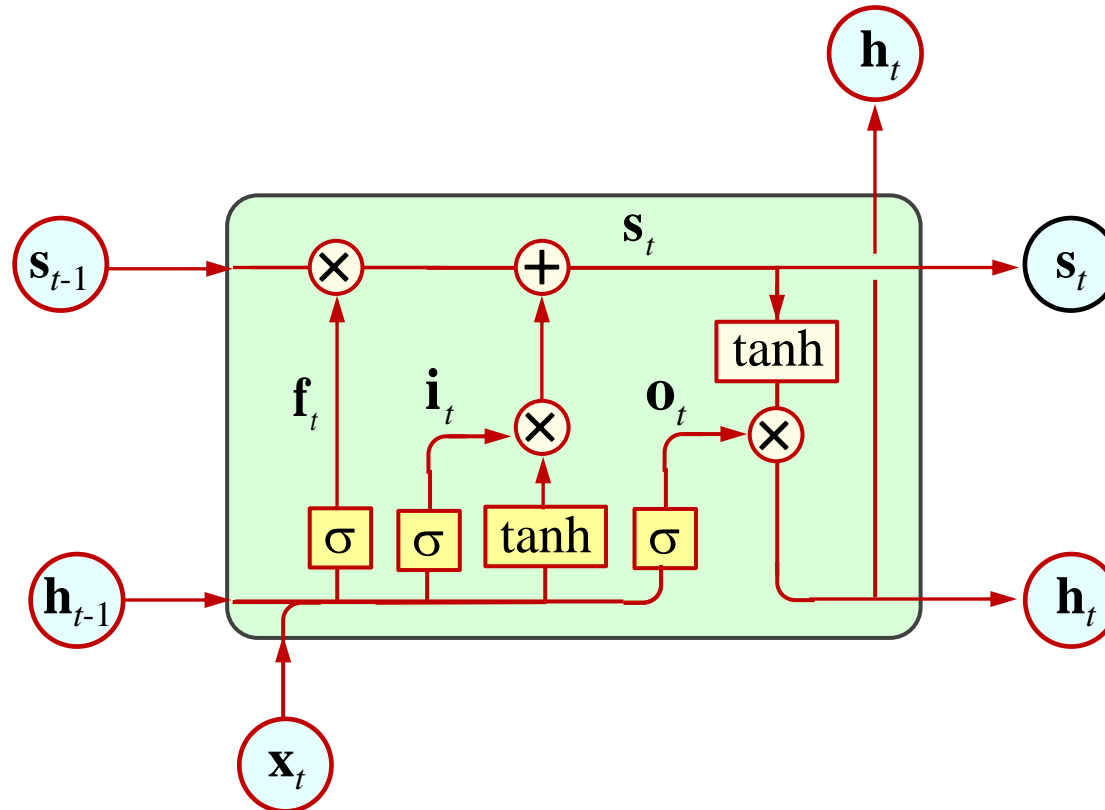
$$s_t = f_t \otimes s_{t-1} + i_t \otimes c_t$$

5.7 From RNN to LSTM



$$\mathbf{o}_t = \text{sigmoid}(\mathbf{b}_o + \mathbf{U}_o \mathbf{x}_t + \mathbf{W}_o \mathbf{h}_{t-1})$$

5.7 From RNN to LSTM



$$\mathbf{o}_t = \text{sigmoid}(\mathbf{b}_o + \mathbf{U}_o \mathbf{x}_t + \mathbf{W}_o \mathbf{h}_{t-1})$$