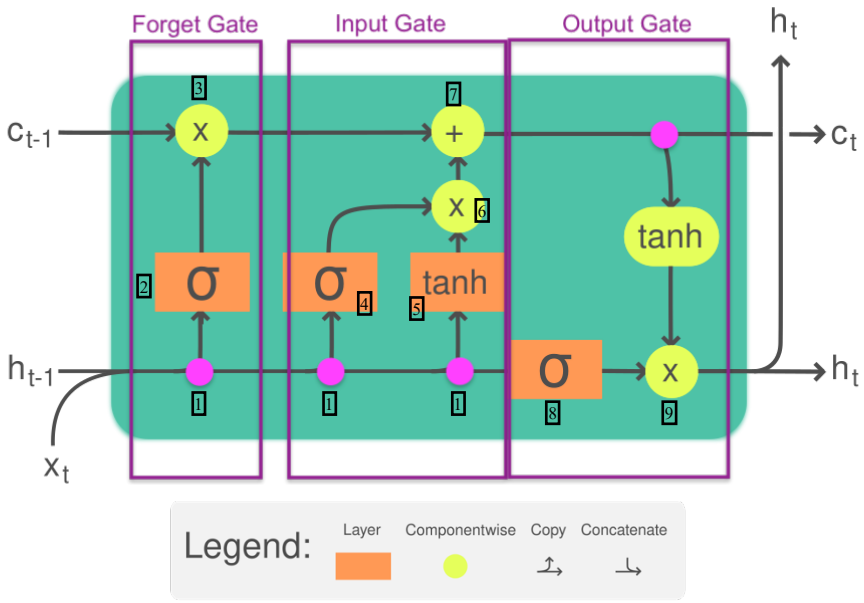
LSTM
Long Short-Term Memory



1. Concatenation: $[h_{t-1}, x_t]$

Forget Gate: Controls what to discard from the previous cell state.

2. Sigmoid Activation (f_t) : $f_t = \sigma(W_f \cdot [h_{t-1}, x_t + b_f)$
3. $f_t \odot c_{t\text{-}1}$

Sigmoid = Gate Controller, regulating the flow of information by outputting values between 0 (block/discard) and 1 (allow/keep).

Input Gate: Decides what new information to add.

4. $i_t = \sigma(W_i \cdot [h_{t-1}, x_t + b_f)$
5. Candidate Cell State $\tilde{C}_t = tanh(W_c \cdot [h_{t-1}, x_t + b_c)$
6. $i_t \odot \tilde{C}_t$

Tanh = Value Rescaler that brings values into [−1,+1], ensuring the cell state and hidden state remain bounded + Provides nonlinearity, enabling the network to capture complex patterns.

Cell State Update: Combines old and new information to update the memory.

7. $c_t = f_t \odot c_{t-1} + i_t \odot \tilde{C}_t =$ Retained memory + New memory

Output Gate: Determines what part of the memory to output as the hidden state.

8. $o_t = \sigma(W_o \cdot [h_{t-1}, x_t + b_O)$
9. $h_t = o_t \odot tanh(c_t)$