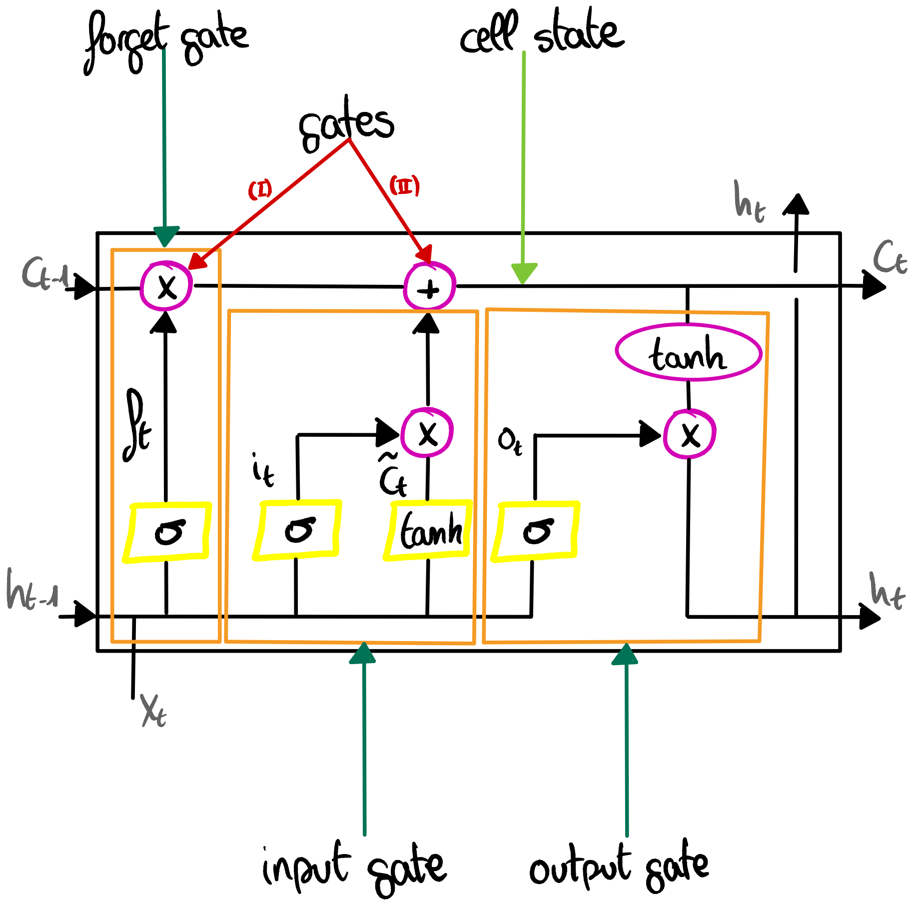
A closer look at LSTM



cell state "The information that flows through the path"

gates "Let pass or not information to the cell state"
(I) 1 → nothing is forgotten ; 0 → all is forgotten
(II) 1 → information is added ; 0 → information is not added

sigmoid (σ) Sigmoid can output 0 to 1, it can be used to forget or remember the information.

tanh (tanh) To overcome the vanishing gradient problem. tanh's second derivative can sustain for a long range before going to zero.

forget gate "What information to remember, what information to forget"

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f)$$

Outputs a number between 0 and 1 for each number in the cell state. 0 completely forget and 1 to keep all information.

input gate "What new information will be stored in the cell state"

$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i)$ ← sigmoid layer decides which values are updated.

$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C)$ ← tanh layer gives weights to the values to be added to the state

output gate "Decide what part of the current cell makes to the output"

$o_t = \sigma(W_o [h_{t-1}, x_t] + b_o)$ ← sigmoid layer decides which part of cell state is selected for output.

$h_t = o_t * \tanh(C_t)$ ← tanh layer gives weights to the values (-1 to 1)