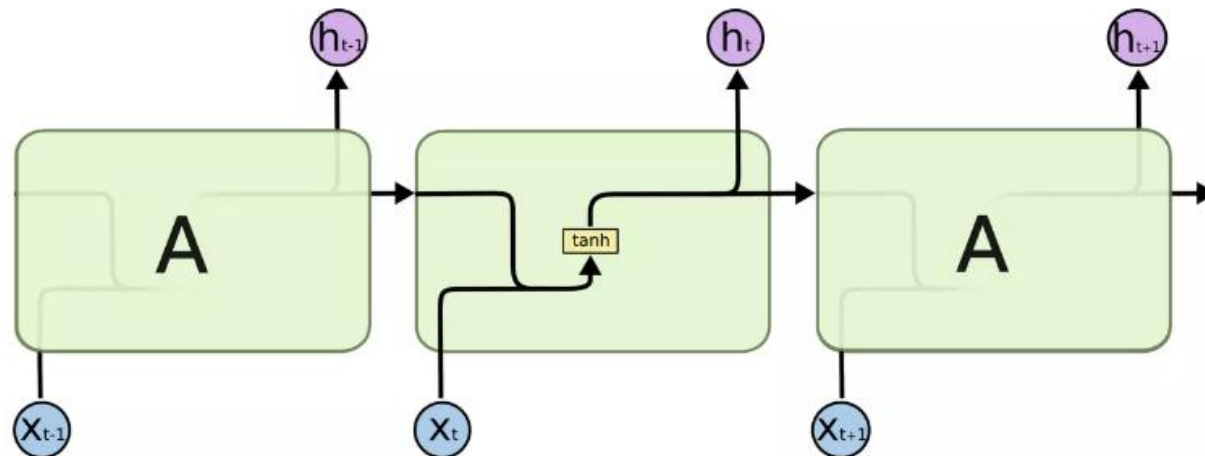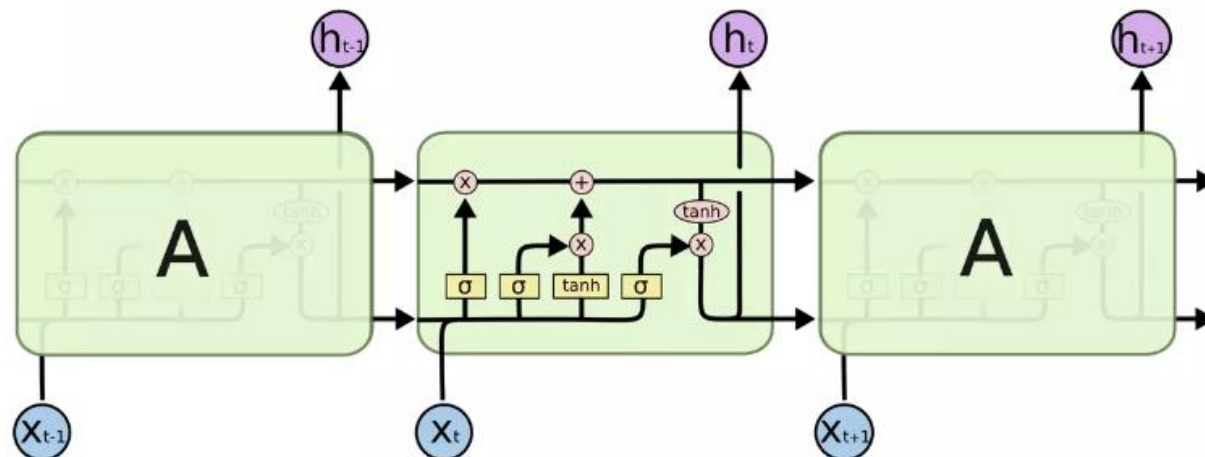# How do LSTMs work?

- LSTMs try to add long term memory to remember certain hidden states more than others. This allows them to retain knowledge over longer sequences.

- They have 2 outputs instead of 1, the hidden state and the cell state. Their computation is a bit more complex than RNNs
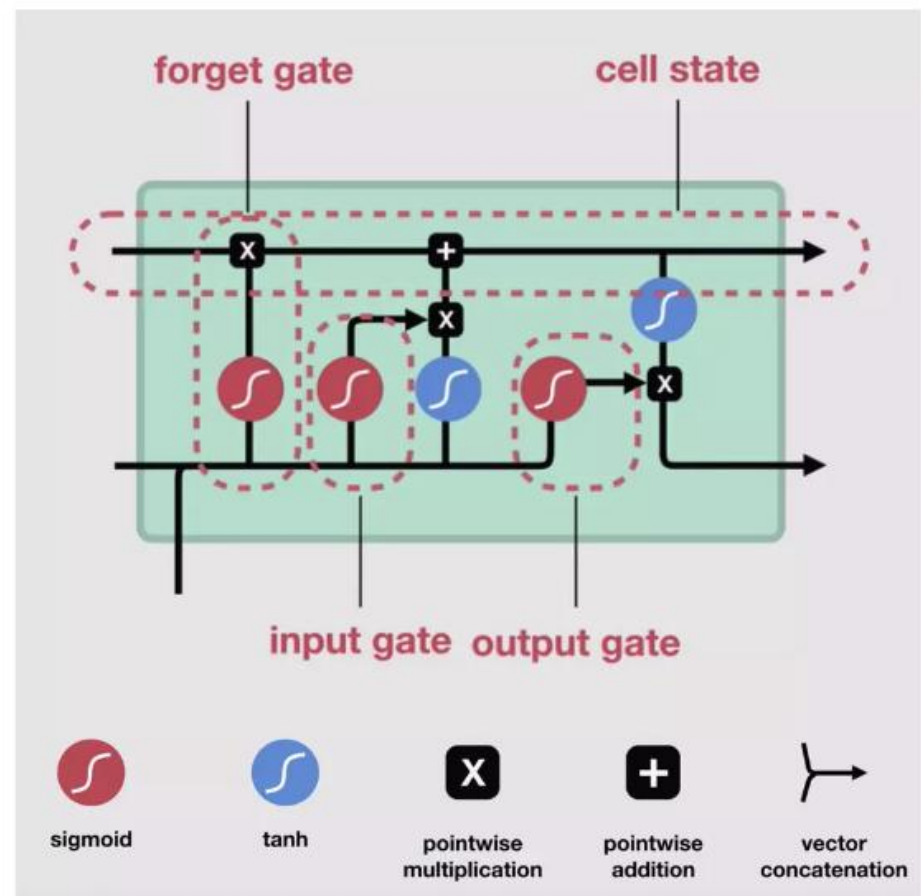
RNN Chain

LSTM Chain

LSTM cell architecture

- An LSTMs architecture consists of 3 gates - Forget gate, Input gate, Output gate

- Tanh acts as a squashing function while Sigmoid acts as a decision function (gate)

- Cell state is a channel that runs along the LSTM chain carrying information from one time-step to another freely
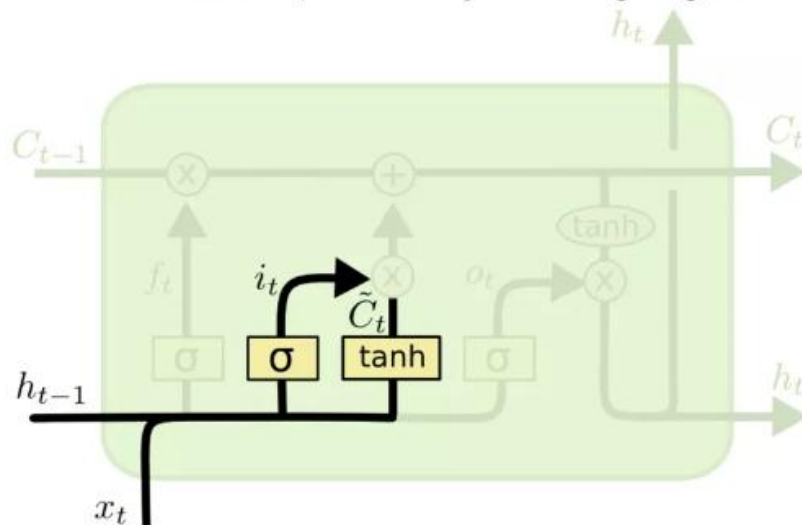
The Cell state

A cell state is a conveyor belt that can carry information
from one time step to another. The three gates add
information to the cell state. Whether to add information
or not is dependent on the Sigmoid function. 0 means
add no information, 1 means add complete information.
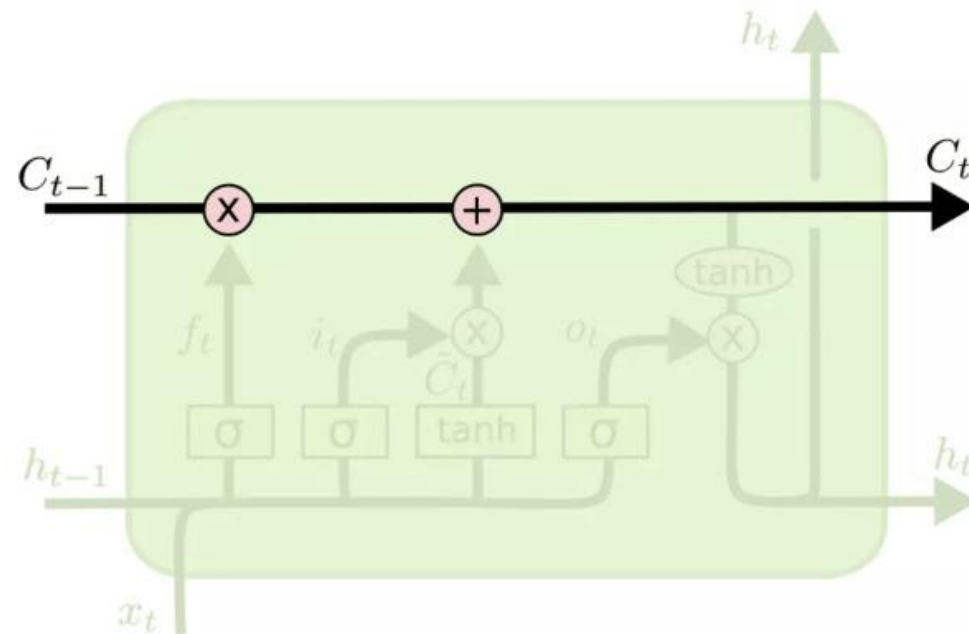
The Input Gate

Input gate decides what information needs to be saved to the cell state. It simply does the same operation as a forget gate but instead of writing it onto the cell state, it combines (multiplies) it with the Tanh (squashed) of the concatenated vector of hidden state and input (plus bias). This is then added to the cell state, which has been updated by the forget gate already.

$$i_t = \sigma\left(W_i \cdot [h_{t-1}, x_t] + b_i\right)$$

$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C)$$
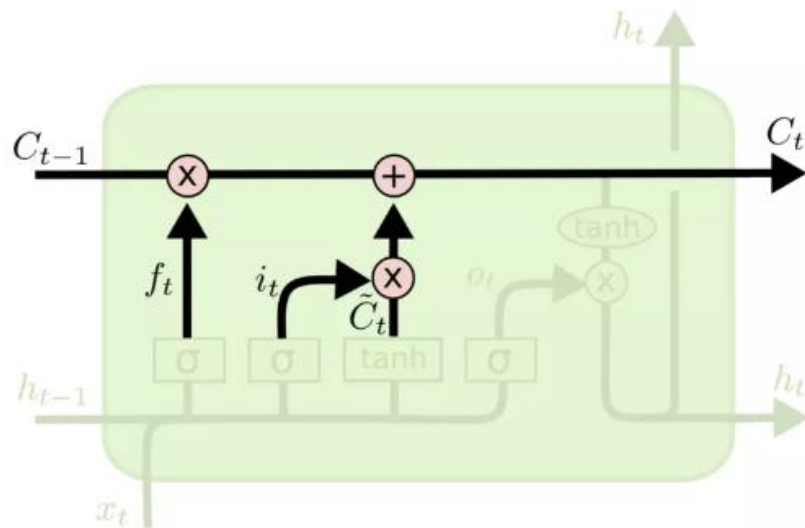
$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t$$

The Output Gate

Finally, we decide what is the output of the LSTM cell (other than the cell state, which becomes the hidden state for the next LSTM cell). This is done simply by applying a sigmoid function on the concatenation of the previous hidden state and current input. But we then multiply it with the squashed (tanh) version of the cell state which contains what to remember and what to forget.
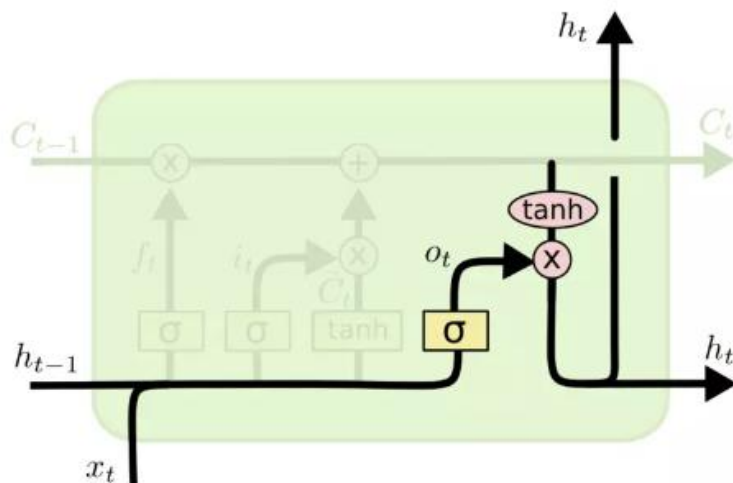
$$o_t = \sigma\left(W_o\left[h_{t-1}, x_t\right] + b_o\right)$$

$$h_t = o_t * \tanh\left(C_t\right)$$

Machine Translation

Using LSTMs as an encoder and decoder for machine translation or Question-Answering bot.