Applied Text Analytics & Natural Language Processing

with Dr. Mahdi Roozbahani
& Wafa Louhichi

Deep Learning

Long Short-Term Memory (LSTM) – Part 2

Some of the slides are based on Ming Li (University of Waterloo – Deep Learning Part)
with some modifications

Forget Gate Layer

$$f_t = \text{sigmoid}(x_t\theta^{\{input_f\}} + h_{t-1}\theta^{\{hidden_f\}} + b_f)$$

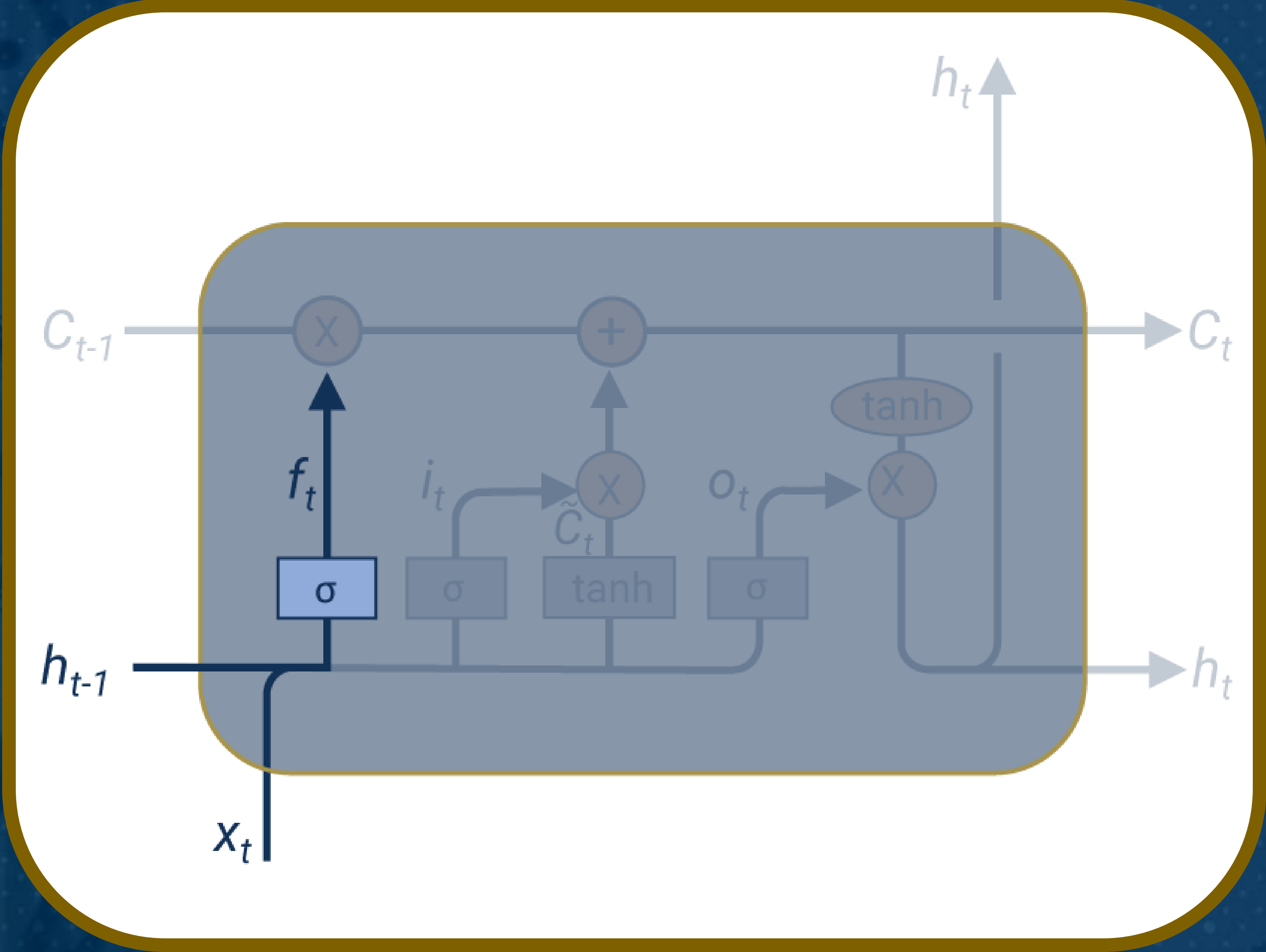
OR

$$f_t = \sigma(x_t\theta^{\{input_f\}} + h_{t-1}\theta^{\{hidden_f\}} + b_f)$$

f_t: forget gate's activation vector

$\theta^{\{input\}}$: Similar to RNN; the input matrix parameter has a size $d \times m$

$\theta^{\{hidden\}}$: Similar to RNN; the hidden layer matrix parameter has a size $m \times m$



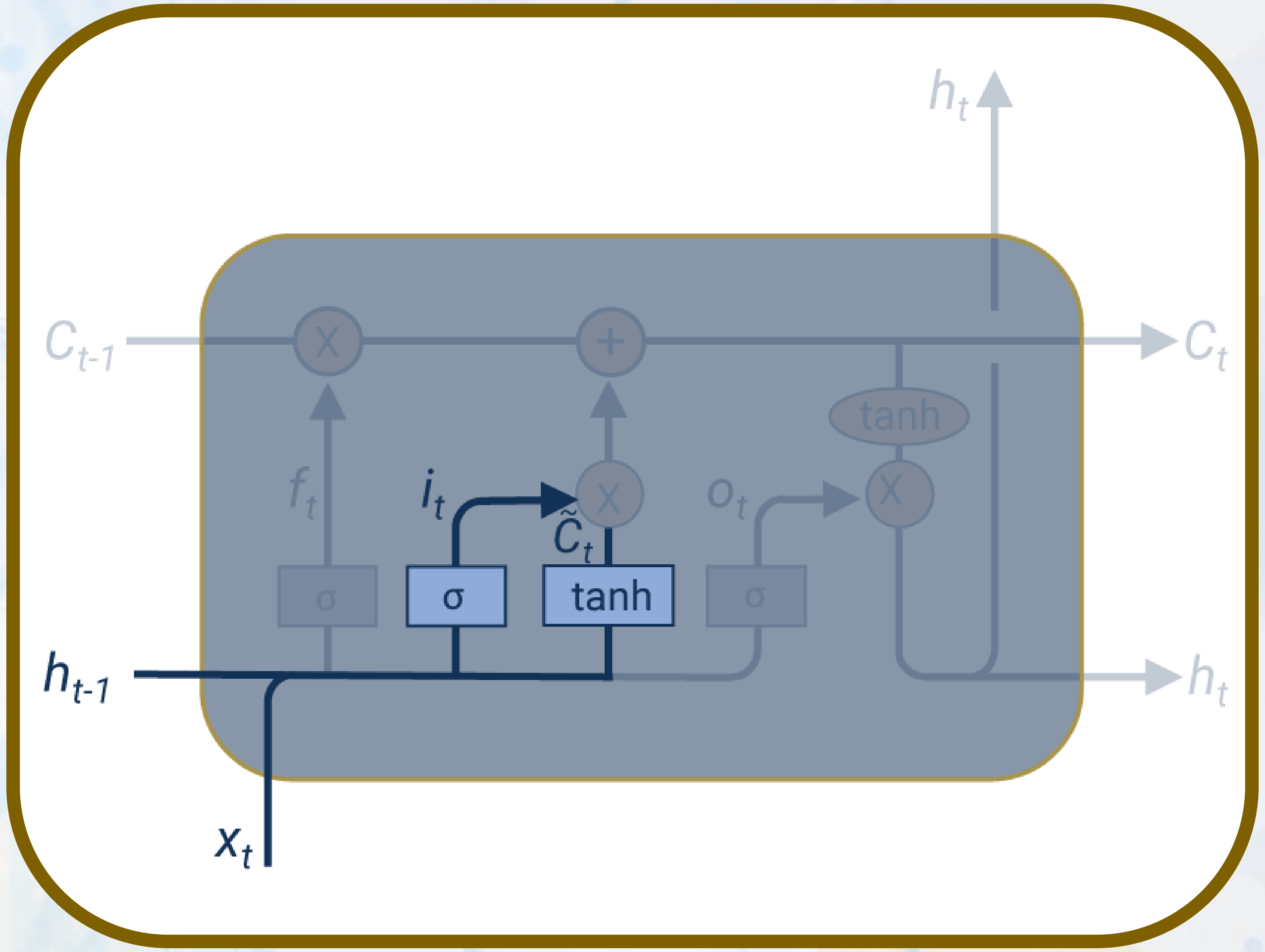
Input Gate Layer to Update Cell State:

$$i_t = \sigma(x_t\theta^{\{input_i\}} + h_{t-1}\theta^{\{hidden_i\}} + b_i)$$

$$\widehat{C}_t = tanh(x_t\theta^{\{input_{\hat{C}}\}} + h_{t-1}\theta^{\{hidden_{\hat{C}}\}} + b_{\hat{C}})$$

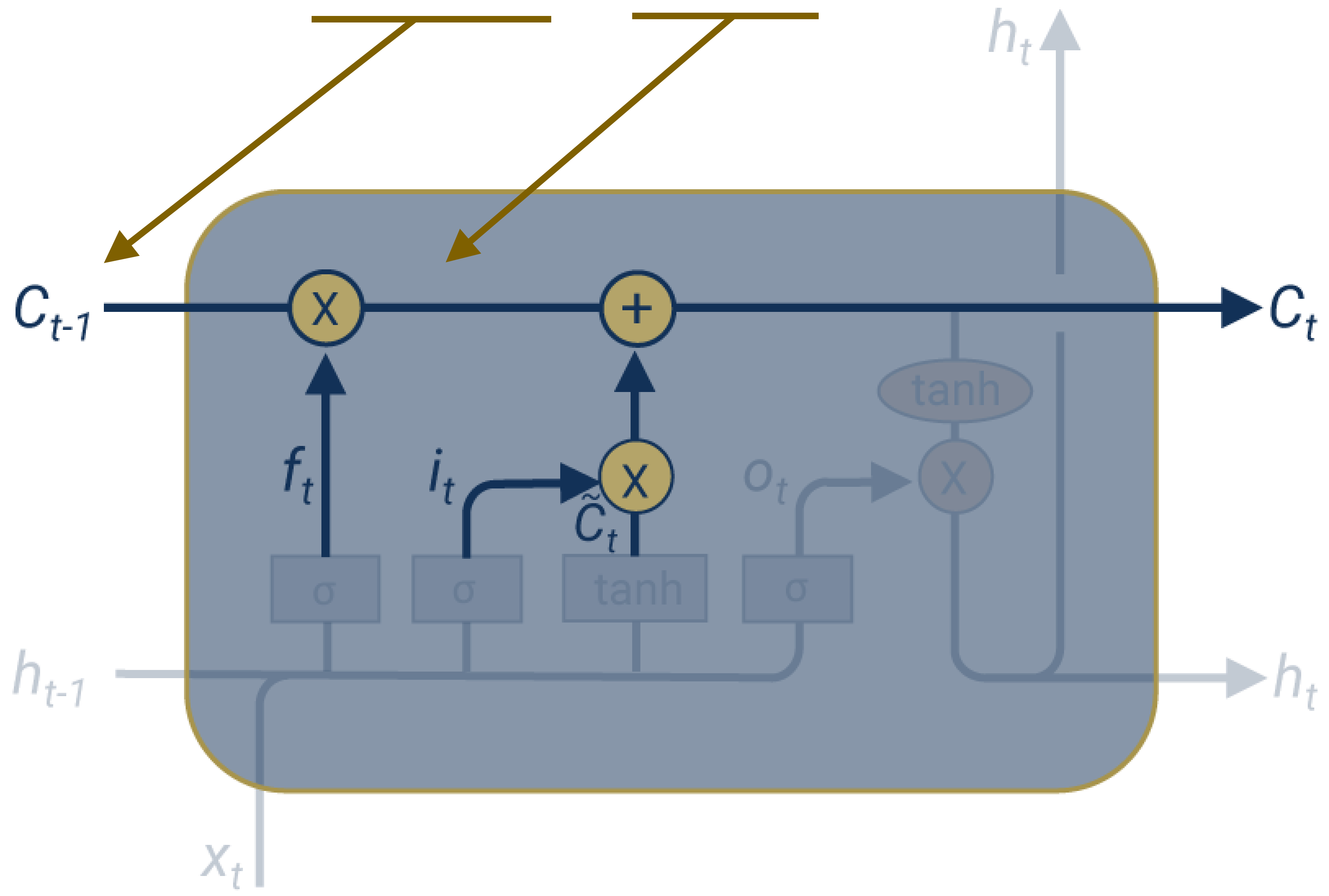
i_t: input/update gate's activation vector

$\widehat{C}_t$: cell input activation vector



Updating the Old Cell State

$$C_t = \underline{f_t * \mathrm{C}_{t-1}} + \underline{\mathrm{i}_t * \widehat{C}_t}$$



C_t: The new Cell state vector

This will update the old cell state (C_{t-1}) to the new one (C_t)

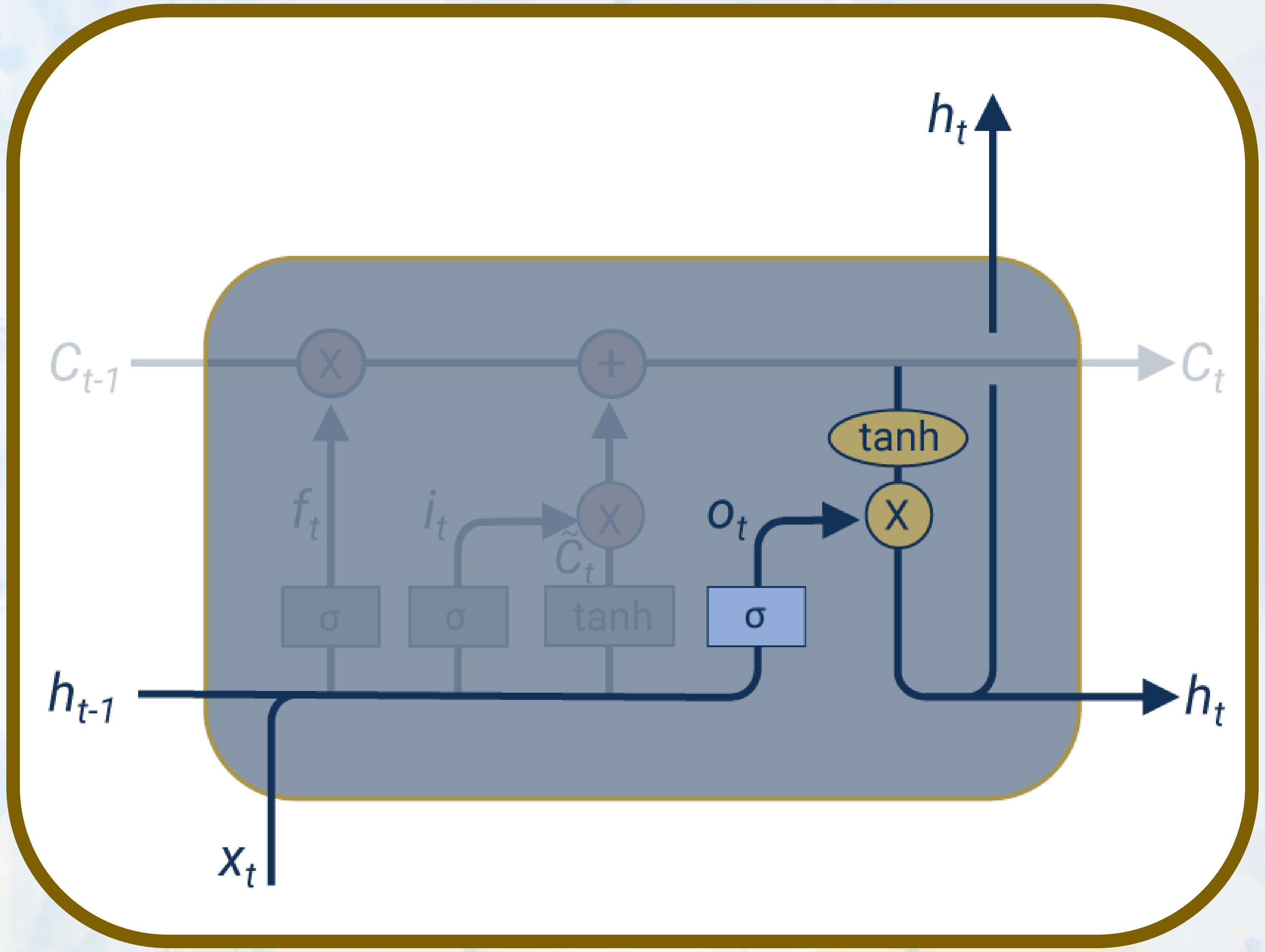
Calculating the Output (h_t) Based on New C_t

$$o_t = \sigma(x_t\theta^{\{input_O\}} + h_{t-1}\theta^{\{hidden_O\}} + b_O)$$

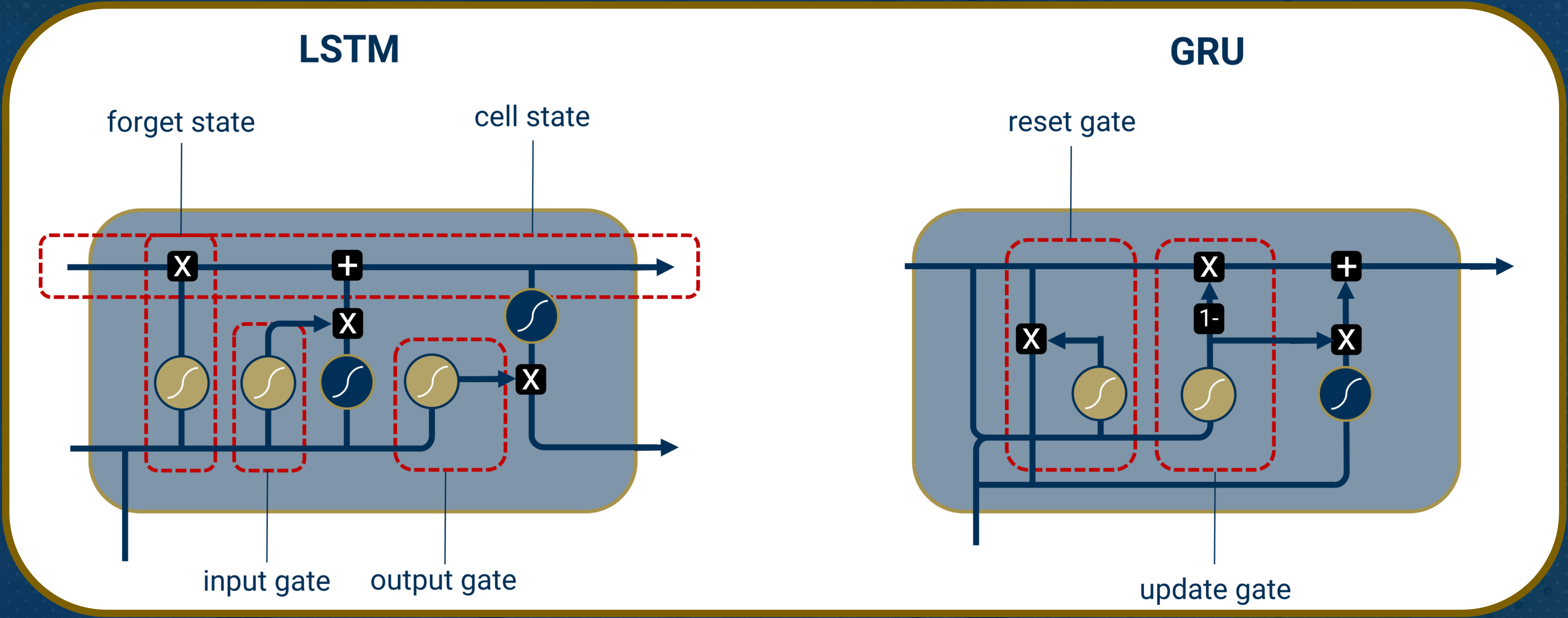
$$h_t = o_t * \tanh(C_t)$$

o_t: output gate's activation vector

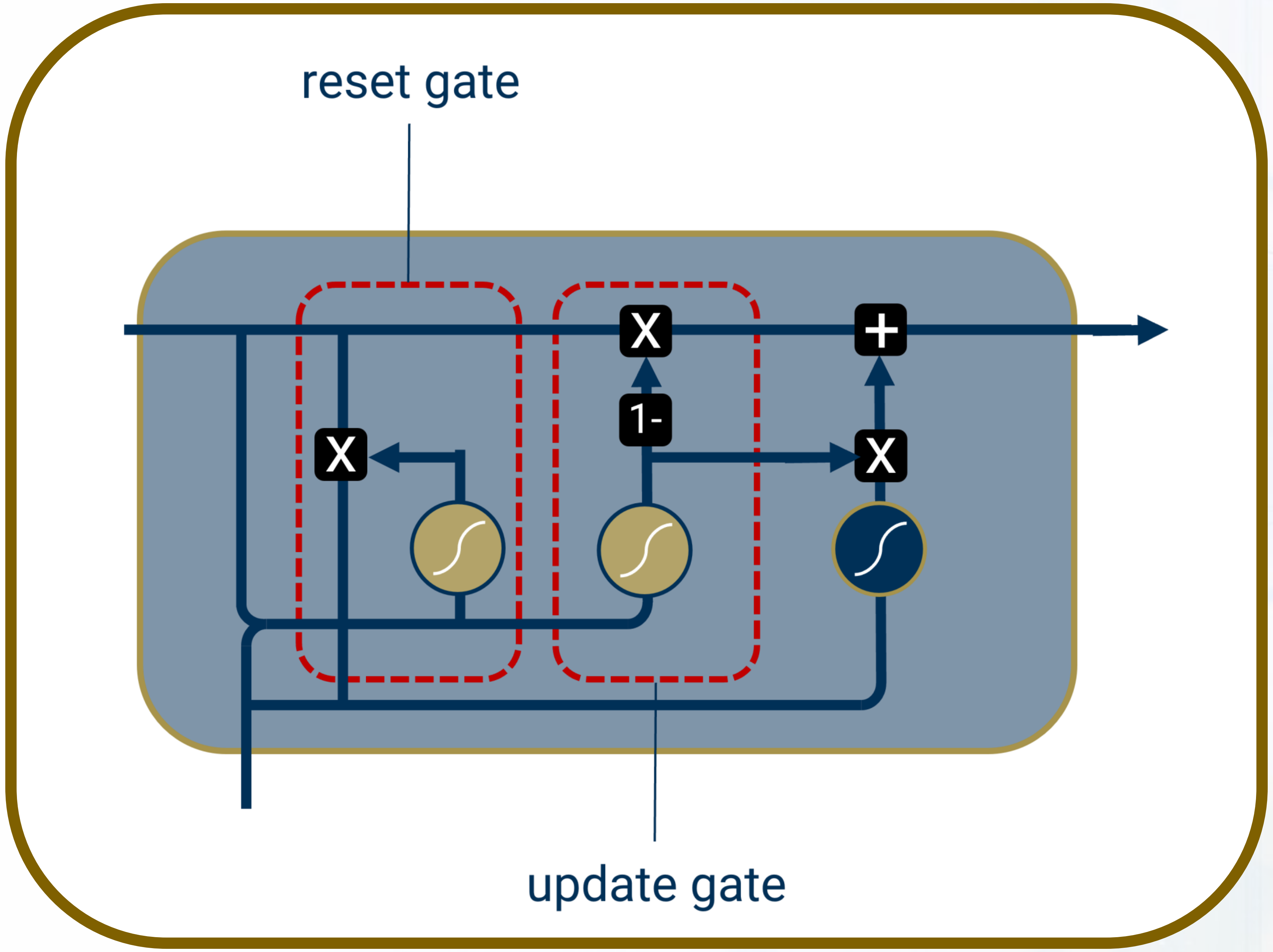
h_t: hidden state vector also known as output vector of the LSTM unit



Gated Recurrent Units (GRU): An update to LSTM



GRU Update and Reset Gates:



Update Gate
The update gate acts similar to the forget and input gate of an LSTM. It decides what information to throw away and what new information to add.

Reset Gate
The reset gate is another gate that is used to decide how much past information to forget.

Summary

- We learned about LSTM
- The long-term dependency
- Different gates (forget, input, cell, and output)
- GRU