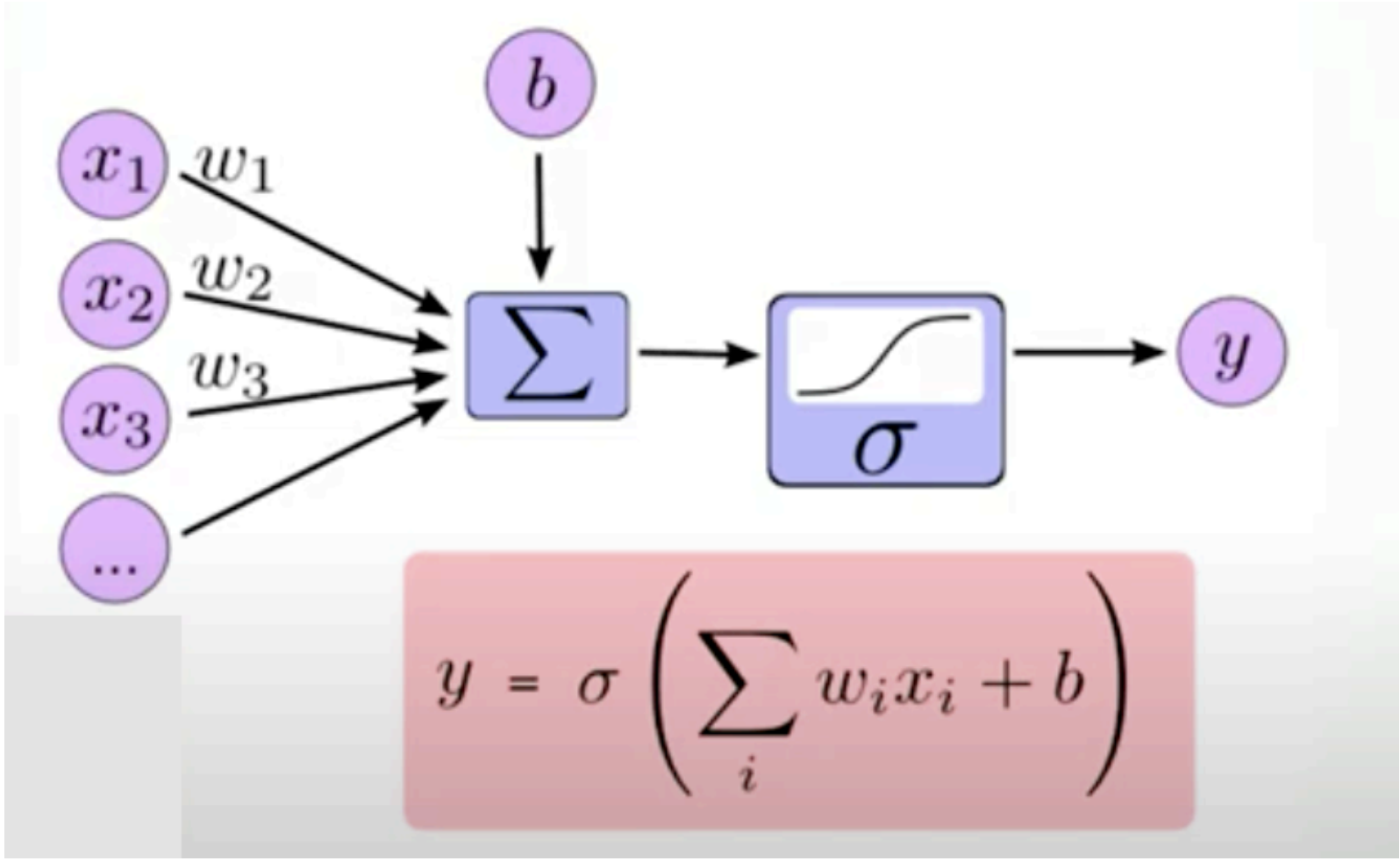
Understanding LSTM Concept

Anuar Assamidanov

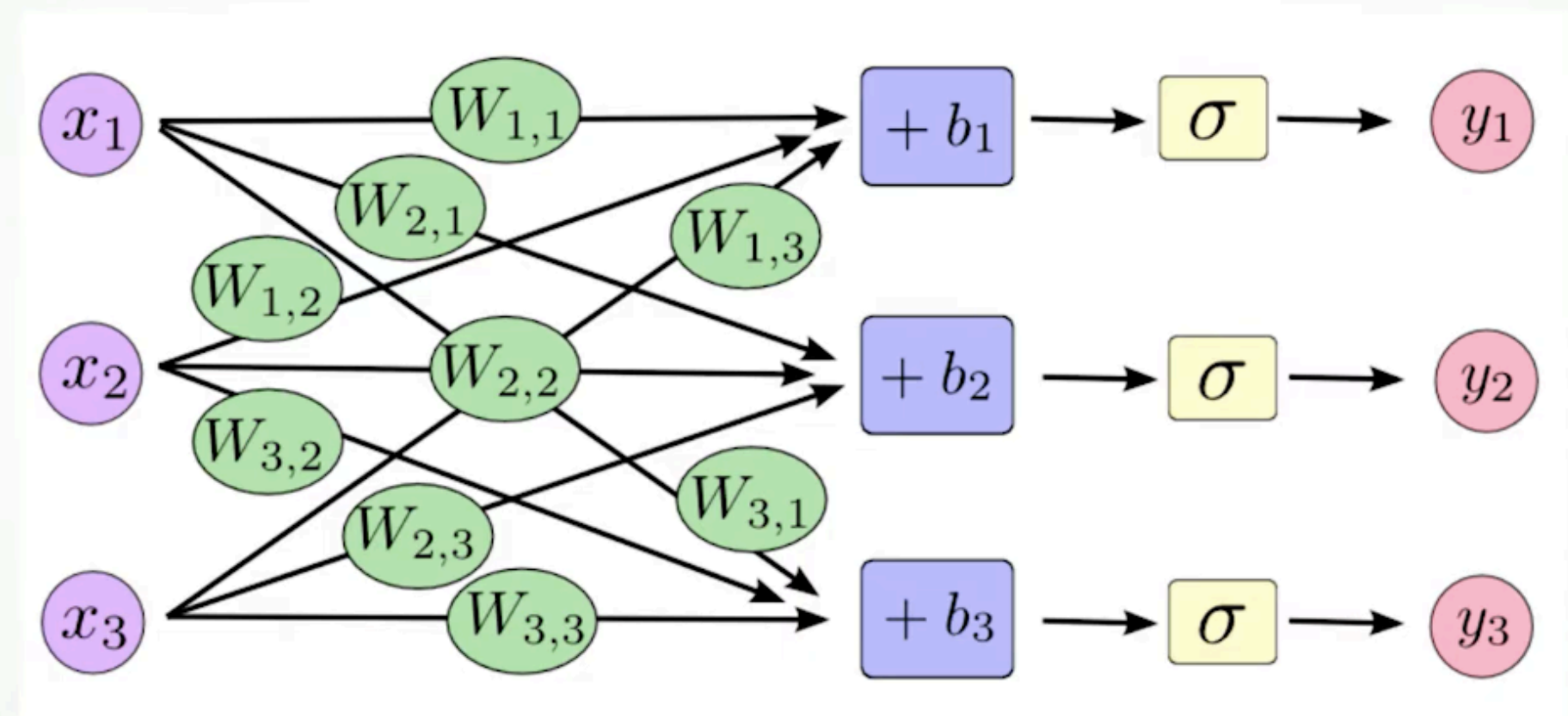
Overview

- Review of Neural Networks
- Recurrent Neural Network
- The Problem of Long-Term Dependencies
- The Core Idea Behind LSTMs
- Step-by-Step LSTM Walk Through
- Variants on Long Short Term Memory
- Demo with TensorFlow

Review of Neural Networks



Layers


x_1
x_2
x_3
$W_{1,1}$
$W_{2,1}$
$W_{1,2}$
$W_{2,2}$
$W_{3,2}$
$W_{2,3}$
$W_{3,3}$
$W_{1,3}$
$W_{3,1}$
$+ b_1$
$+ b_2$
$+ b_3$
σ
σ
σ
y_1
y_2
y_3

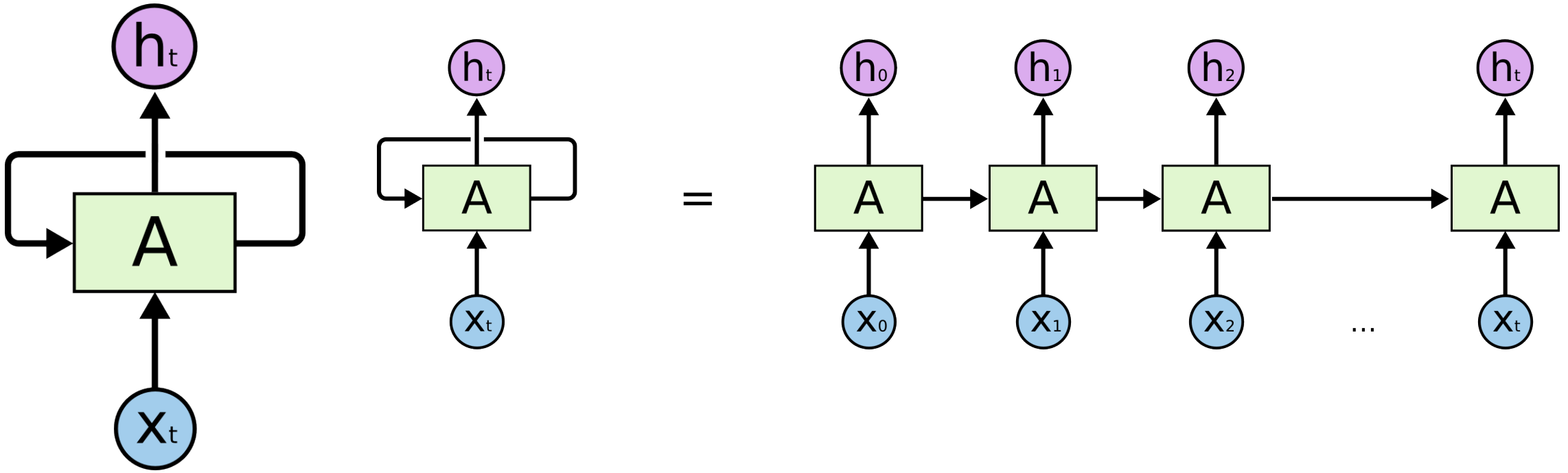
Neural Networks

$$
\begin{aligned}
y_1 &= \sigma(W_{1,1}x_1 + W_{1,2}x_2 + W_{1,3}x_3 + b_1) \\
y_2 &= \sigma(W_{2,1}x_1 + W_{2,2}x_2 + W_{2,3}x_3 + b_2) \\
y_3 &= \sigma(W_{3,1}x_1 + W_{3,2}x_2 + W_{3,3}x_3 + b_3)
\end{aligned}
$$

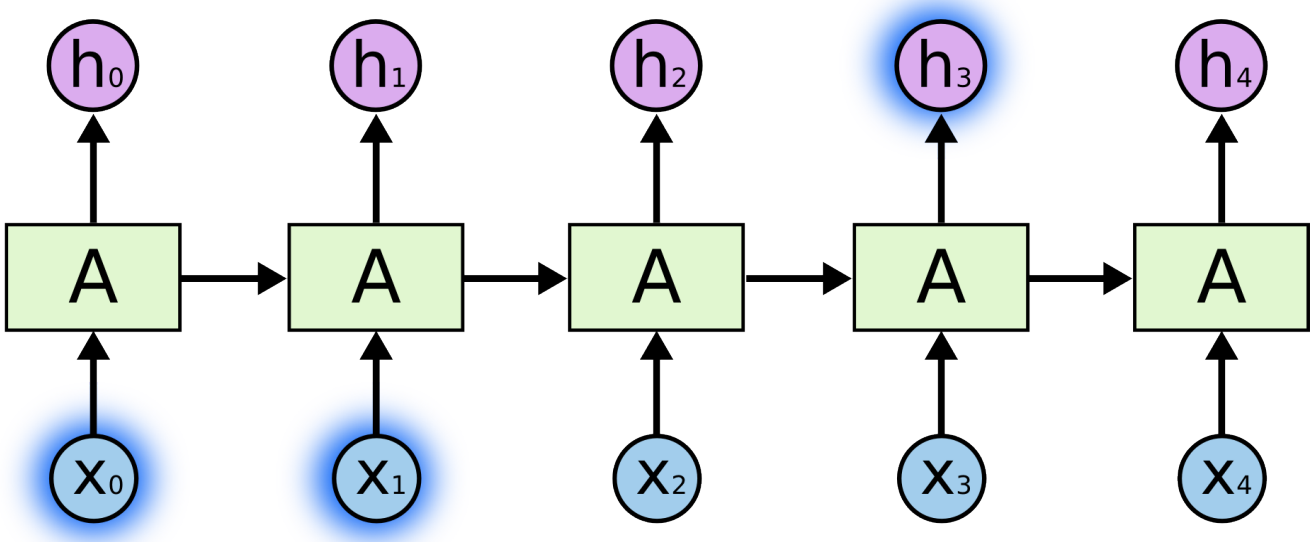
$$
\begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \sigma\left(\begin{bmatrix} W_{1,1} & W_{1,2} & W_{1,3} \\ W_{2,1} & W_{2,2} & W_{2,3} \\ W_{3,1} & W_{3,2} & W_{3,3} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} + \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix} \right)
$$

Recurrent Neural Network(RNN)

Recurrent Neural Networks (RNN)



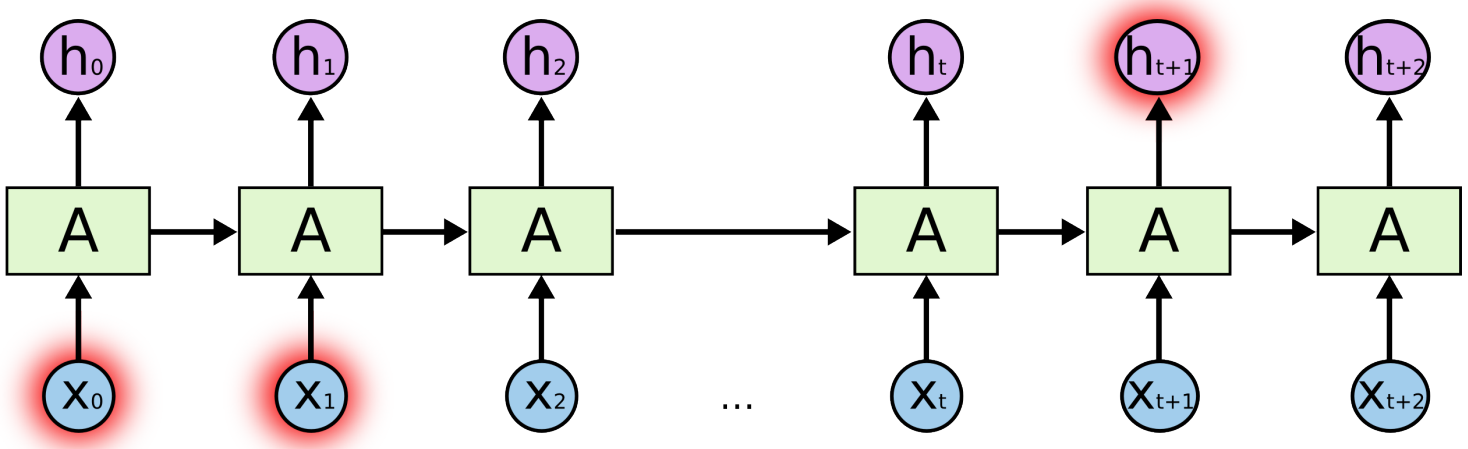
The Problem of Long-Term Dependencies



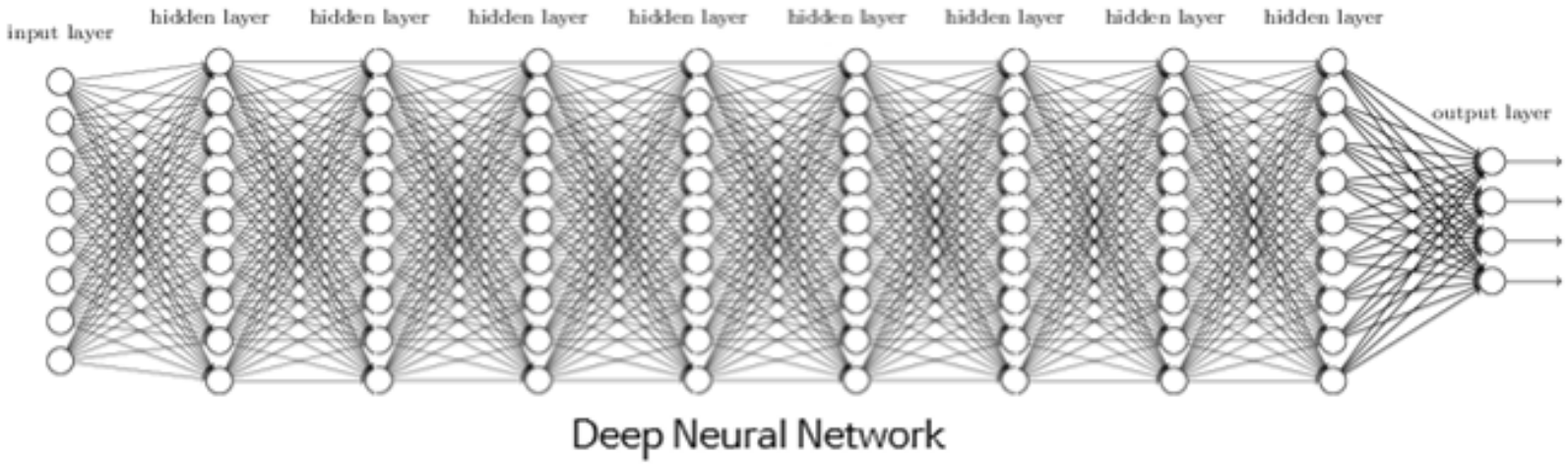
“We eat breakfast in the morning”

The Problem of Long-Term Dependencies

"I live in Europe…. I speak fluent Spanish"

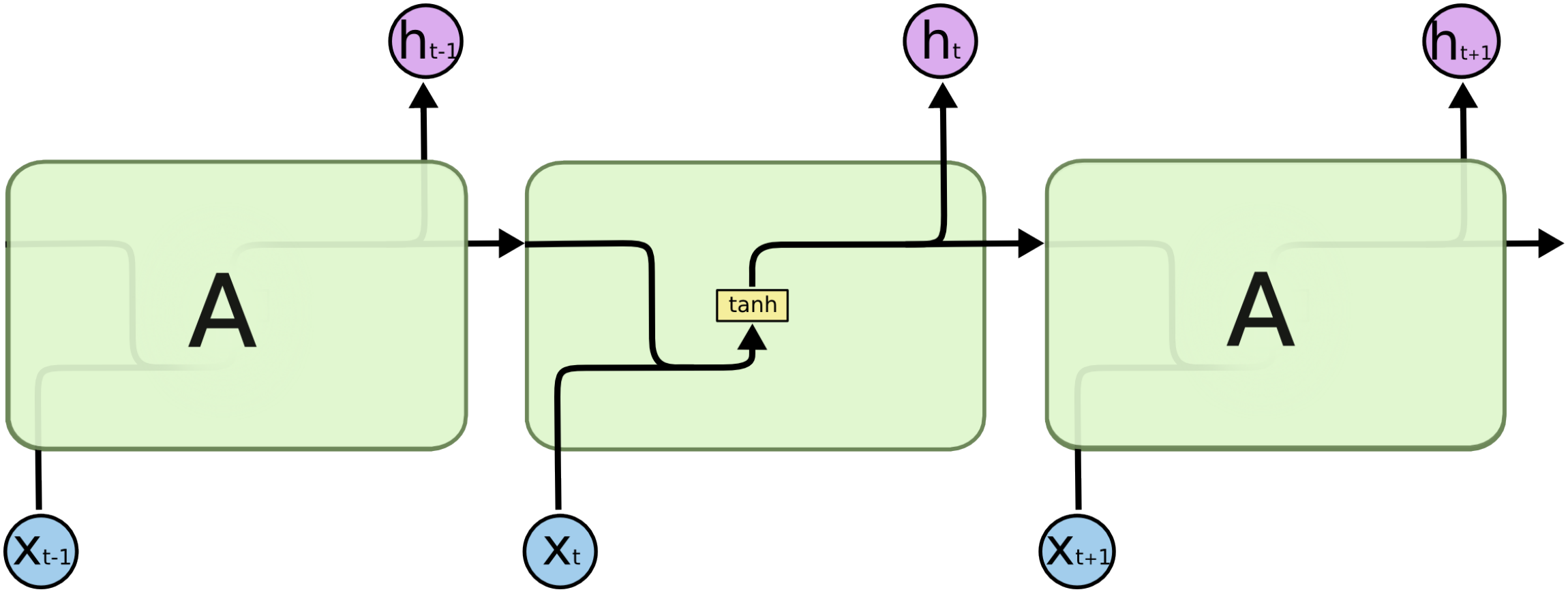


Vanishing Gradient Problem

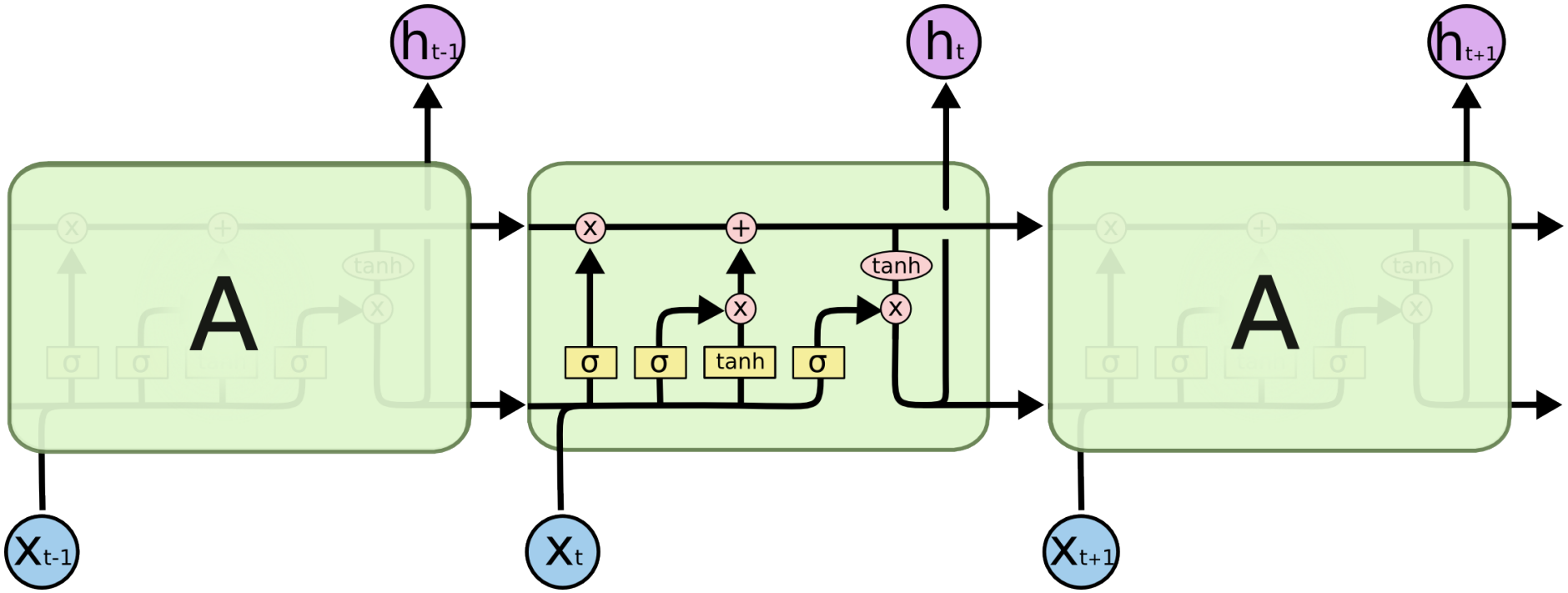


Good news is LSTMs don’t have this problem

The repeating module in a standard RNN contains a single layer.



The repeating module in an LSTM contains four interacting layers.



Neural Network Layer

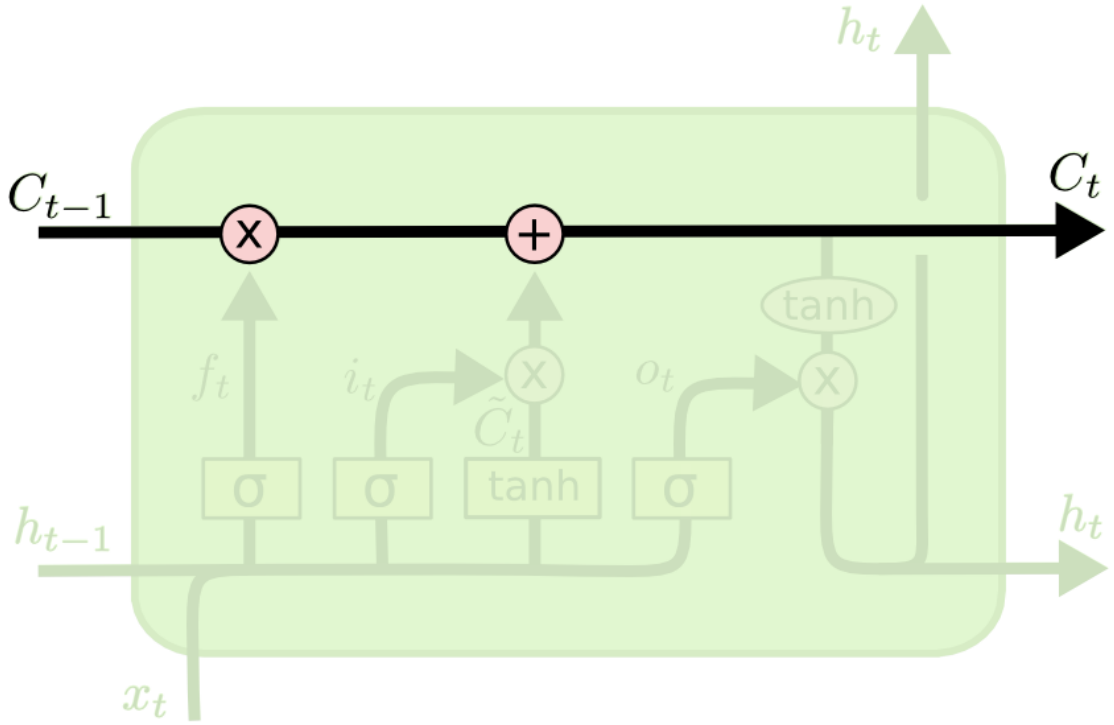
Pointwise Operation

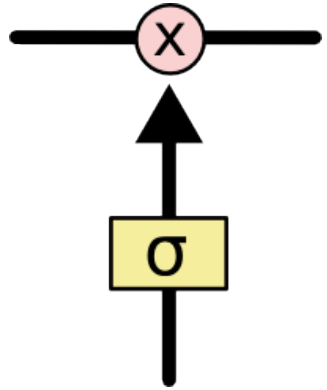
Vector Transfer

Concatenate

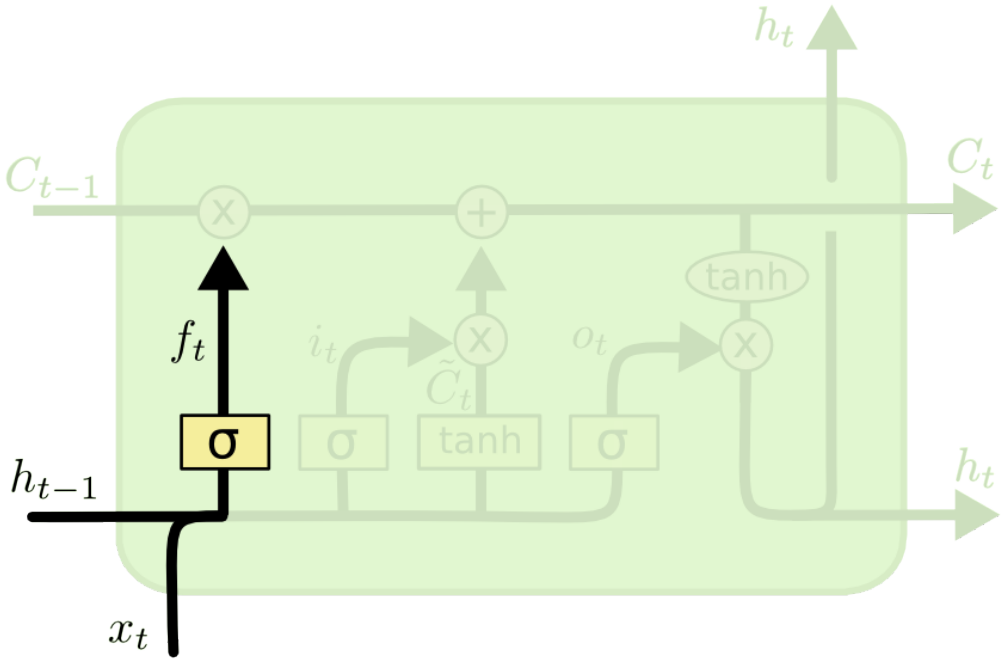
Copy

The Core Idea Behind LSTMs



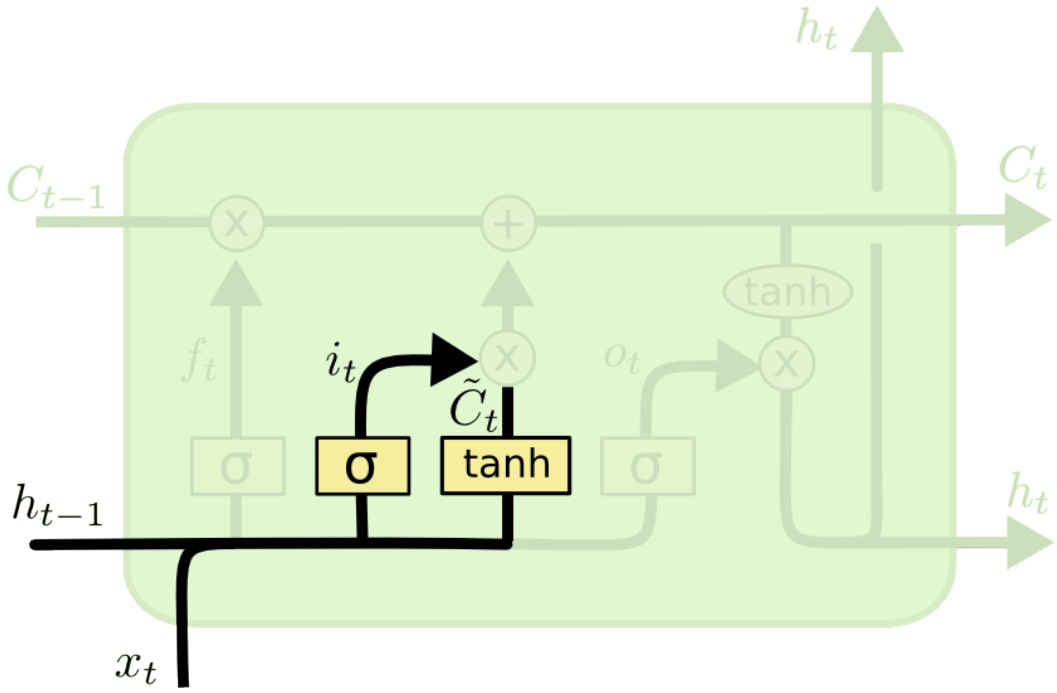


Forget Gate Layer



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

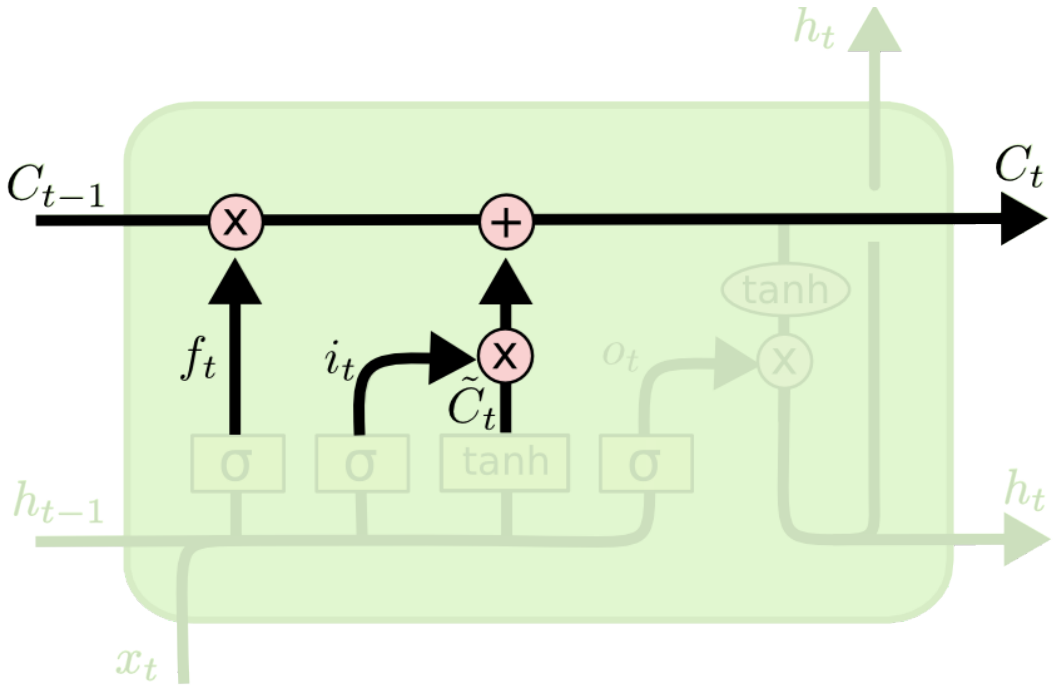
Input Gate Layer



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

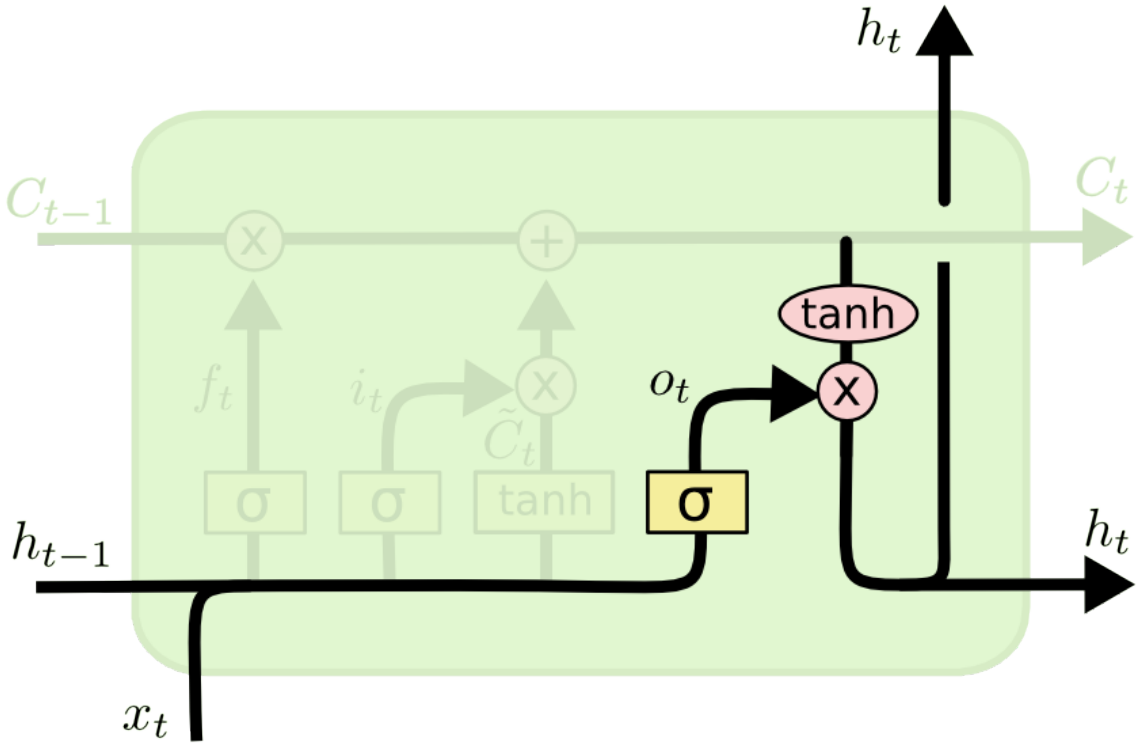
$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] \; + \; b_C)$$

Update Cell State



$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t$$

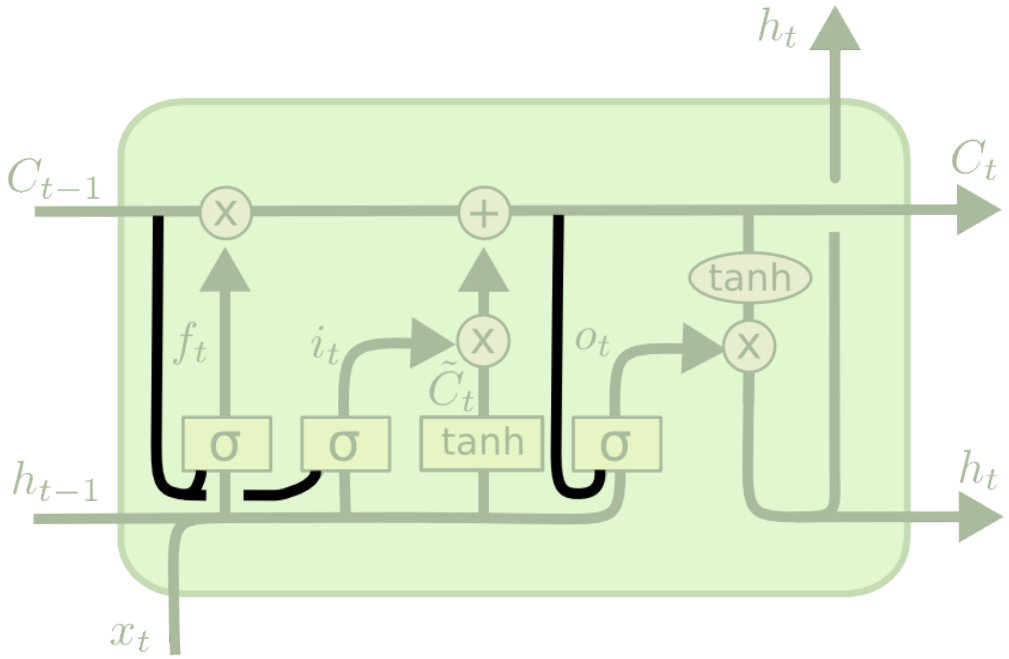
Output Gate Layer



$$o_t = \sigma\left(W_o\left[h_{t-1}, x_t\right] + b_o\right)$$

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

Variants on Long Short Term Memory

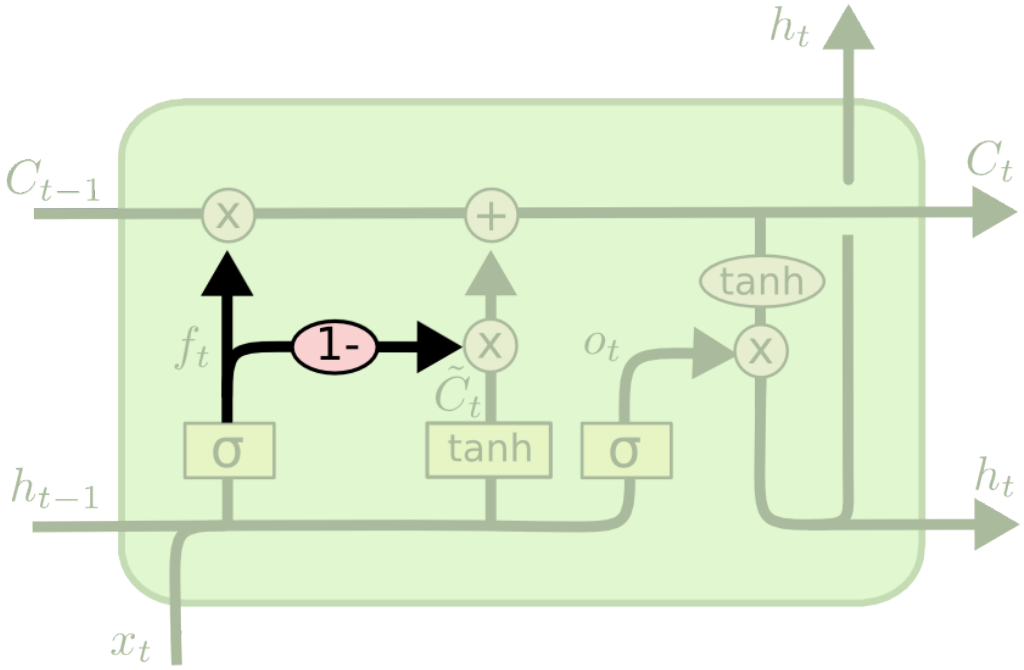


$$f_t = \sigma\left(W_f \cdot [\boldsymbol{C_{t-1}}, h_{t-1}, x_t] \; + \; b_f\right)$$

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

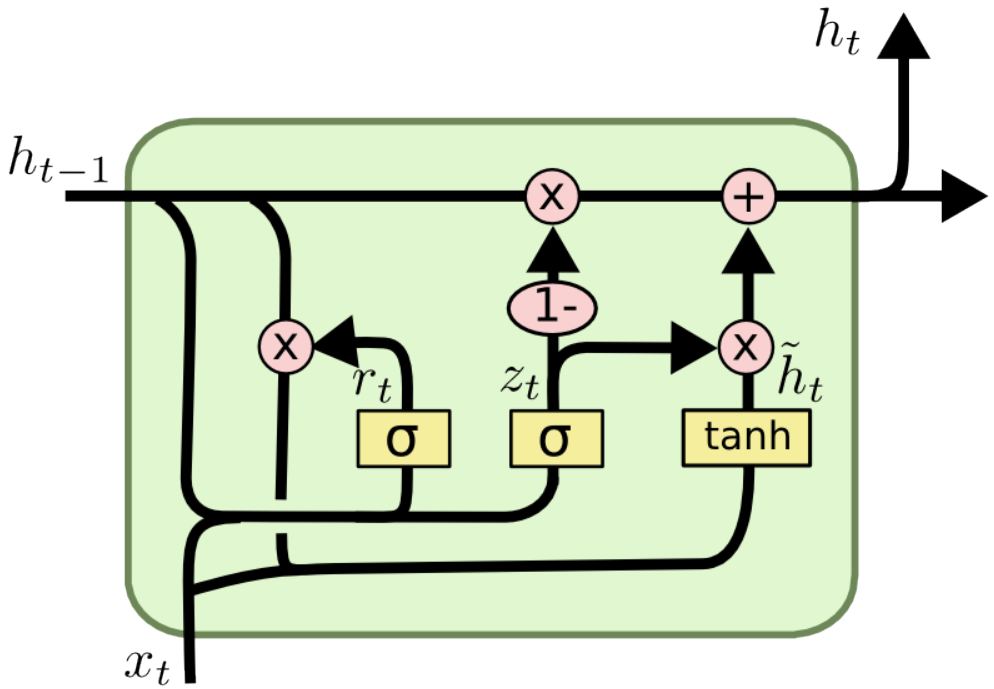
$$o_t = \sigma\left(W_o \cdot [\boldsymbol{C_t}, h_{t-1}, x_t] \; + \; b_o\right)$$

Variants on Long Short Term Memory



$$C_t = f_t * C_{t-1} + (1 - \boldsymbol{f_t}) * \tilde{C}_t$$

Variants on Long Short Term Memory



$$z_t = \sigma\left(W_z \cdot [h_{t-1}, x_t]\right)$$

$$r_t = \sigma\left(W_r \cdot [h_{t-1}, x_t]\right)$$

$$\tilde{h}_t = \tanh\left(W \cdot [r_t * h_{t-1}, x_t]\right)$$

$$h_t = (1 - z_t) * h_{t-1} + z_t * \tilde{h}_t$$

Demo of LSTM