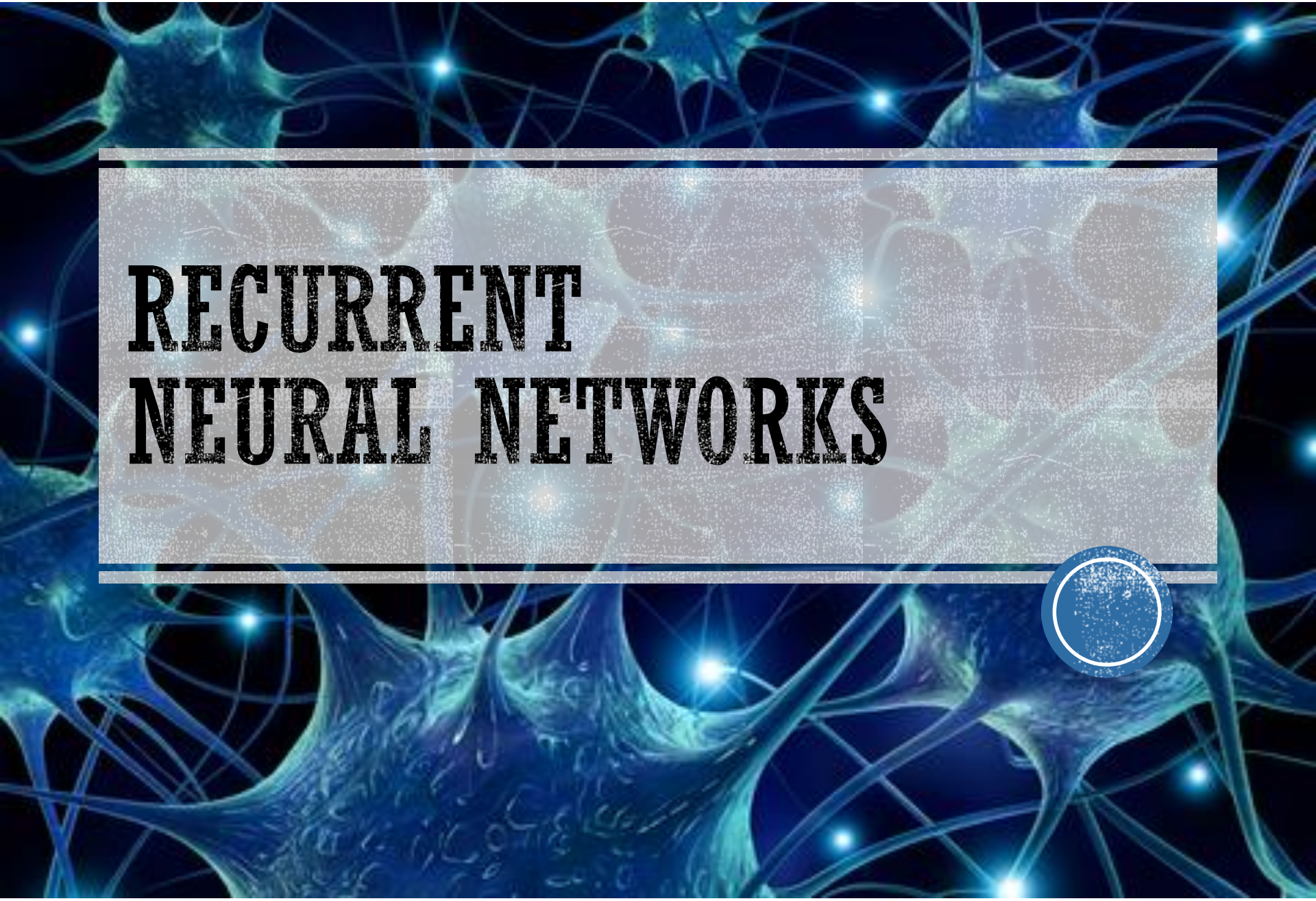
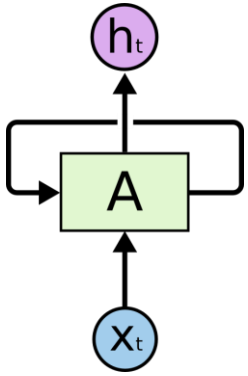

RECURRENT
NEURAL NETWORKS

RECURRENT NEURAL NETWORKS

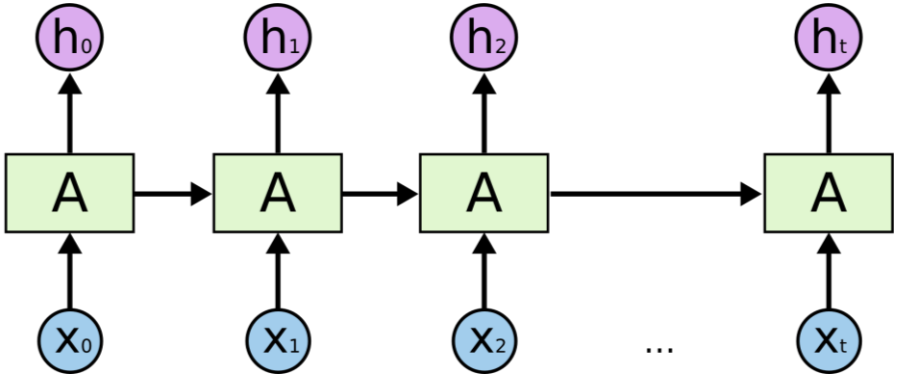
Definition.

Neural networks where the connections form a directed cycle. As opposed to feedforward neural networks.



Redefinition.

May be unrolled as a feedforward network with shared weights propagating through time.

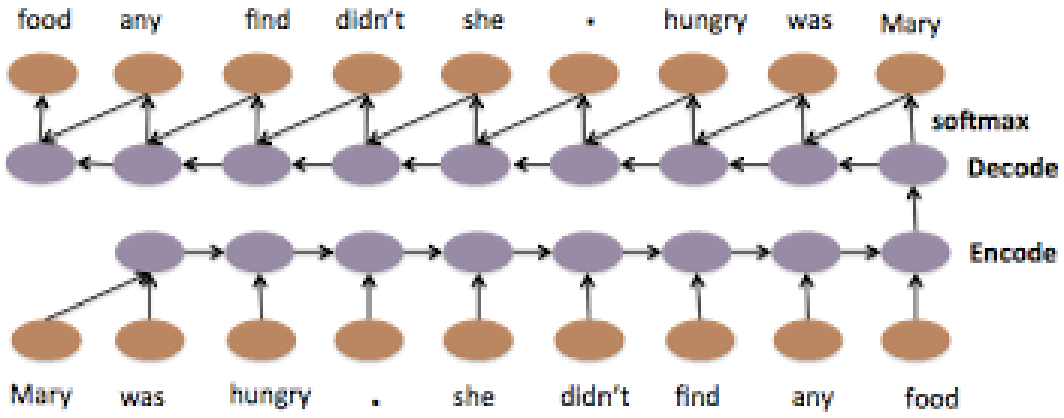


http://colah.github.io/posts/2015-08-Understanding-LSTMs/

RECURRENT NEURAL NETWORKS

Temporal State.

- RNN dynamics encode some time-varying state.
- Perfect for modeling sequences, e.g. speech recognition, natural language processing, financial modeling.



PROBLEM WITH VERY DEEP NETWORKS

Vanishing or Exploding Gradient Problem.

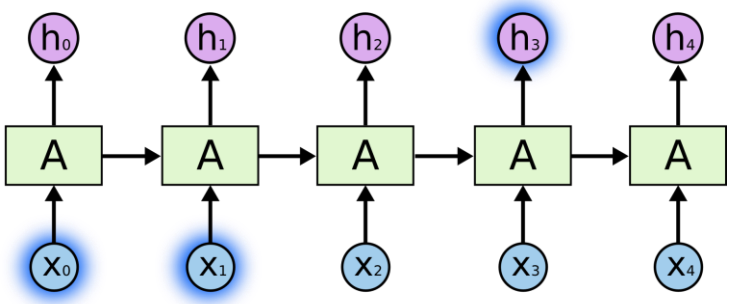
- Difficult to train very deep networks because of issues with backpropagation algorithm
- Error signals $\delta^{(l)}$ either shrink exponentially, or grow exponentially with the number of layers

Recurrent Neural Networks are Very Deep!

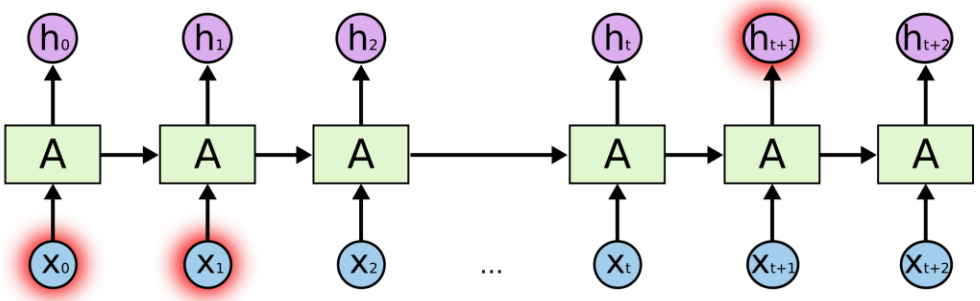
- We want RNNs to learn long-term dependencies
- But for a long time, nobody knew how to train them

LONG-TERM DEPENDENCIES

"the clouds are in the *sky*"

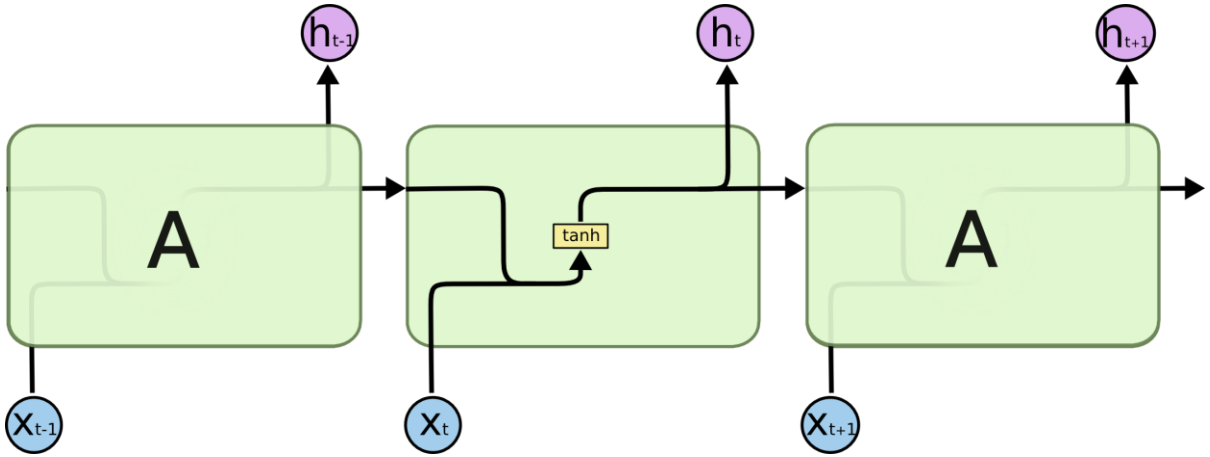


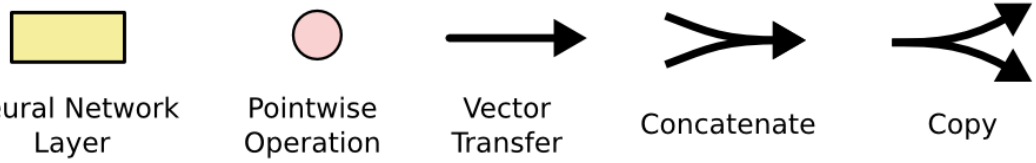
I grew up in France… I speak fluent *French*.



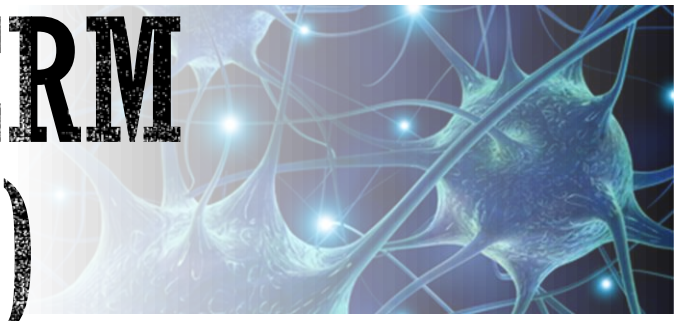
STANDARD RNN

Single Layer RNN.



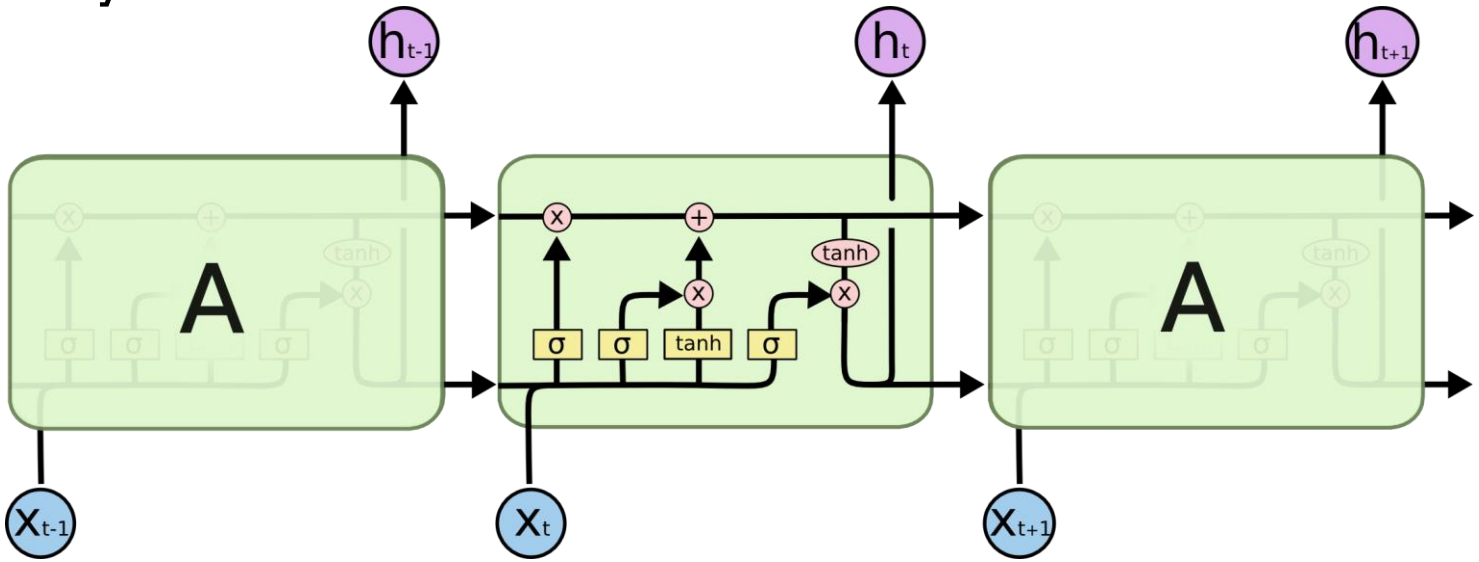


LONG SHORT TERM MEMORY (LSTM)



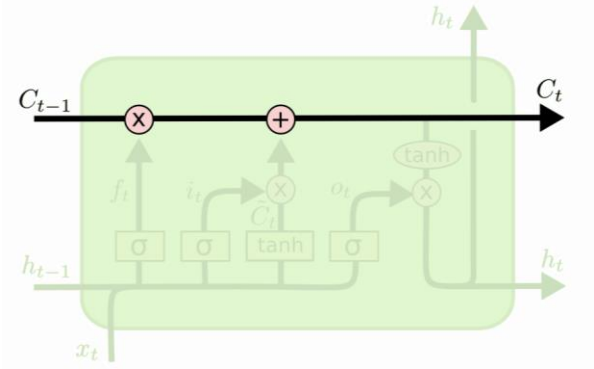
LONG SHORT TERM MEMORY

Four-Layer RNN.

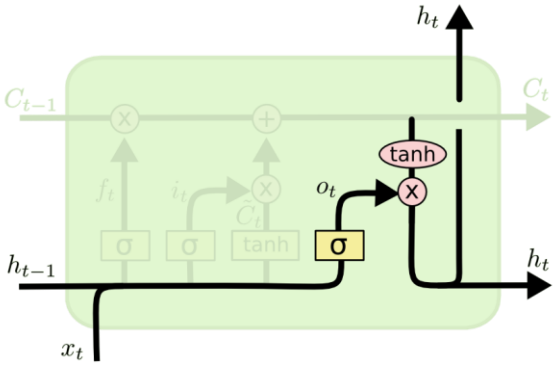


STATES AND GATES

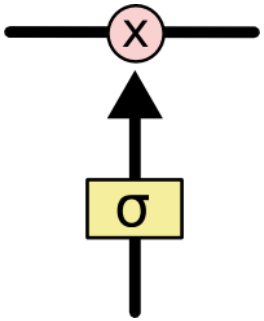
Cell State



Output State



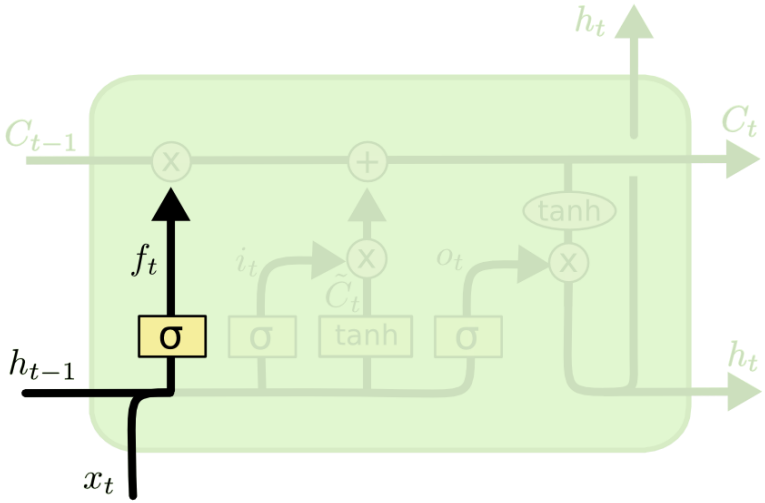
Gates



Output of sigmoid
is between 0 and 1
0 – forget past state
1 – keep past state

LAYER 1 – FORGET PAST STATE

Example. Forget gender of subject when subject changes.

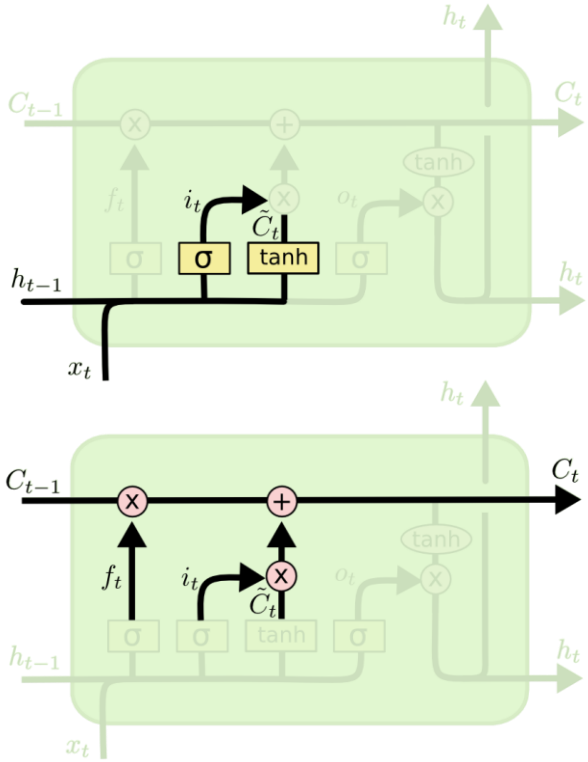


How much to forget?

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

LAYER 2,3 – ADD NEW INFORMATION

Example. Add gender of new subject.



How much to add?

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

What info to add?

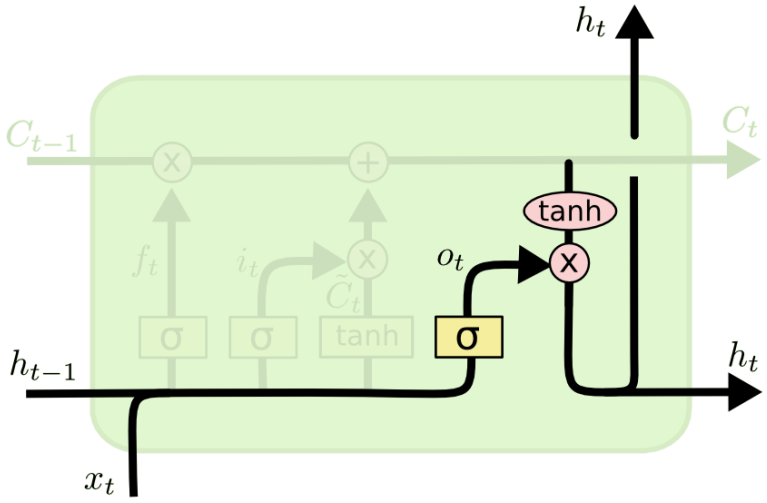
$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] \; + \; b_C)$$

New cell state.

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

LAYER 4 – SELECTIVE OUTPUT

Example. Singular or plural form of verb? Output and forward.



How much to output?

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

What to output?

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