Time will explain.

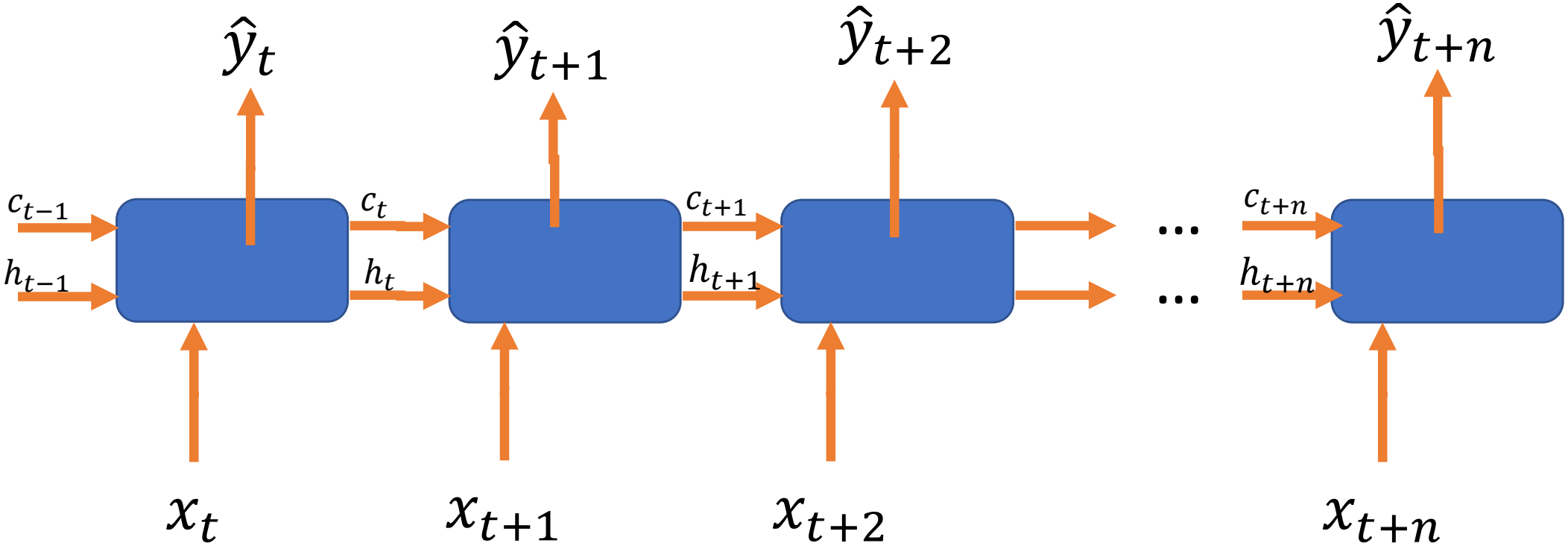
Jane Austen, Persuasion

Neural Language Models

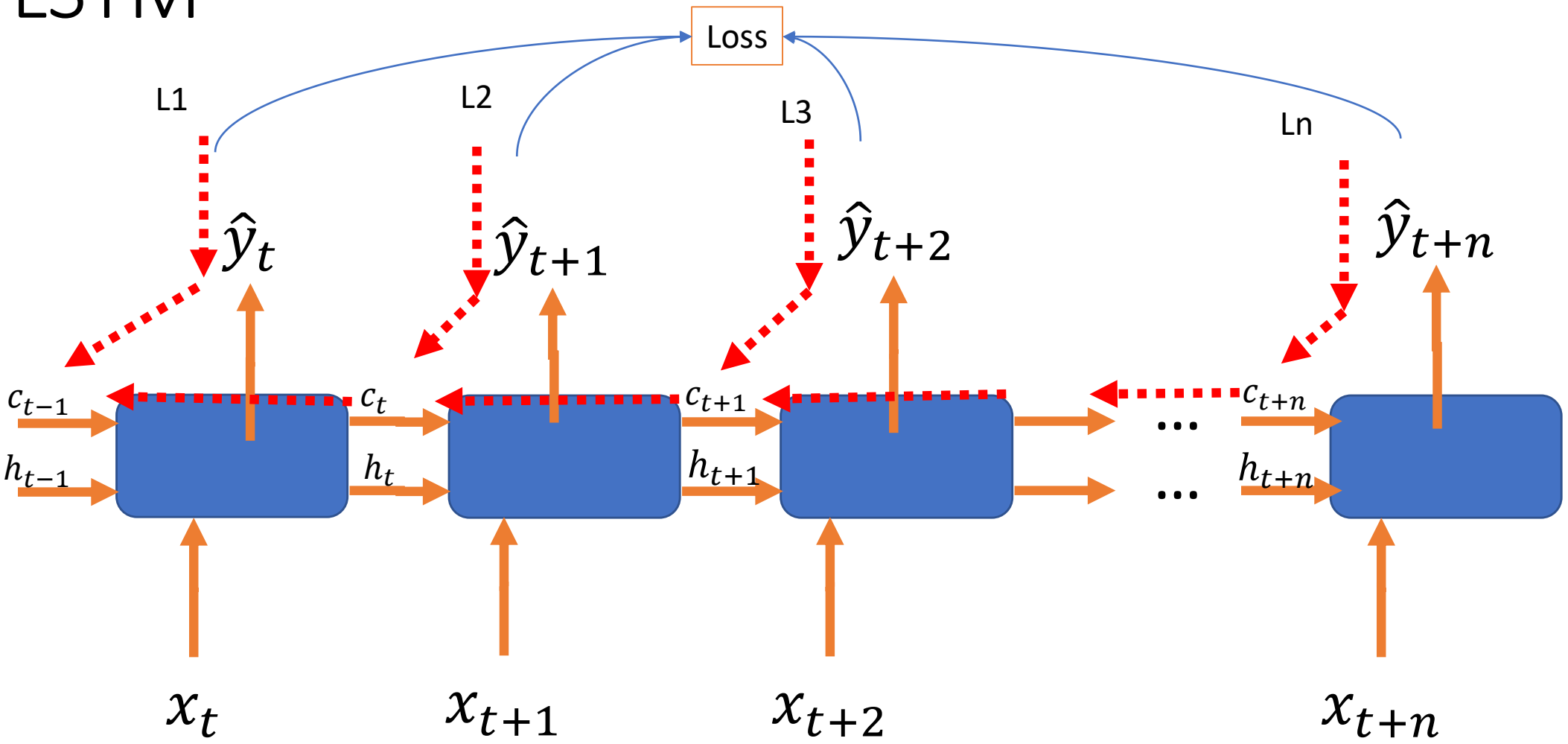
LSTM

Dr. Uzair Ahmad

Long Short-Term Memories (LSTM)



LSTM


Loss
L1
L2
L3
Ln
$\hat{y}_t$
$\hat{y}_{t+1}$
$\hat{y}_{t+2}$
$\hat{y}_{t+n}$
c_{t-1}
c_t
c_{t+1}
c_{t+n}
h_{t-1}
h_t
h_{t+1}
h_{t+n}
x_t
x_{t+1}
x_{t+2}
x_{t+n}

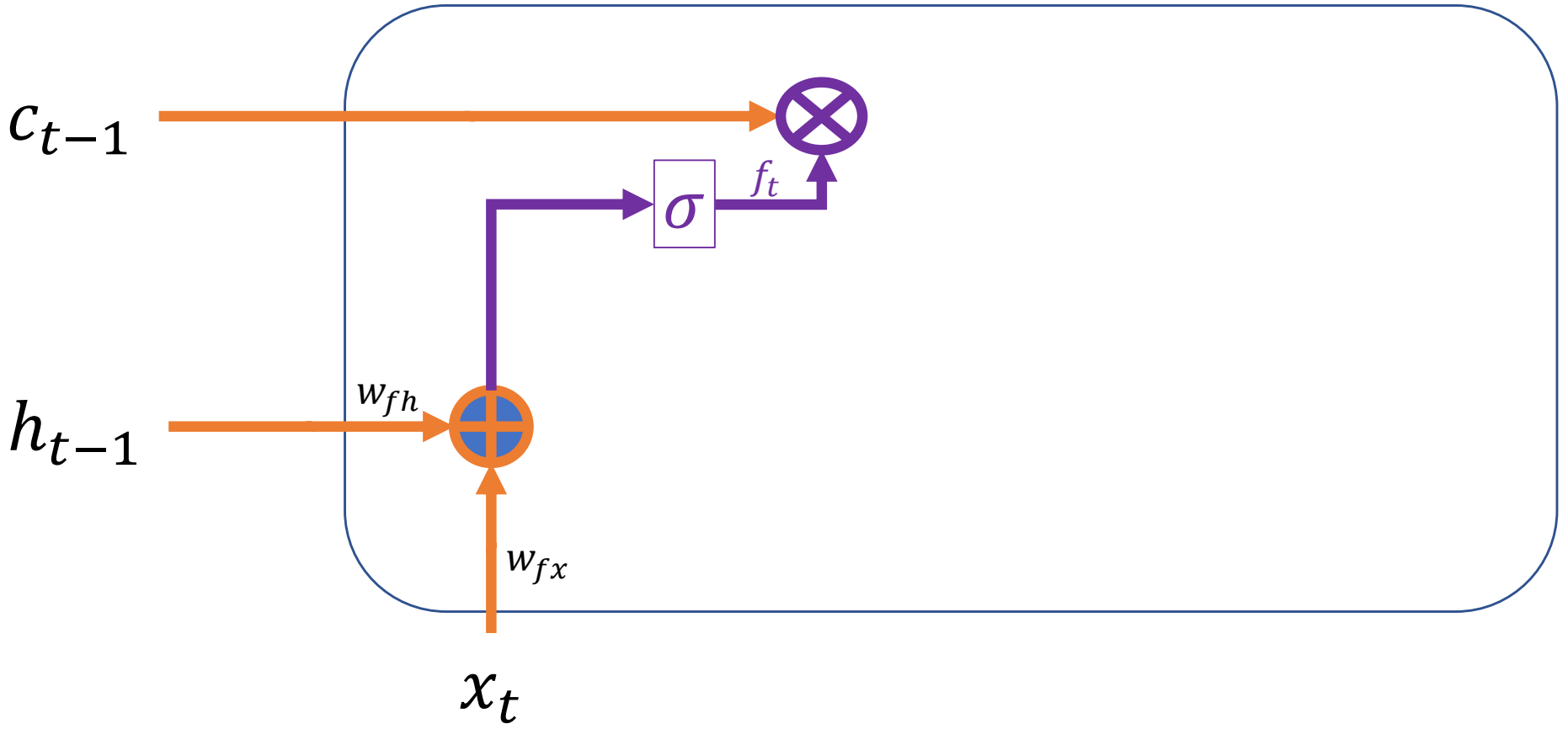
LSTM

- Forget
- Input / Store
- Cell State Update
- Output

LSTM

- Forget

$$f_t = \sigma(w_{fx}x_t + w_{fh}h_t)$$

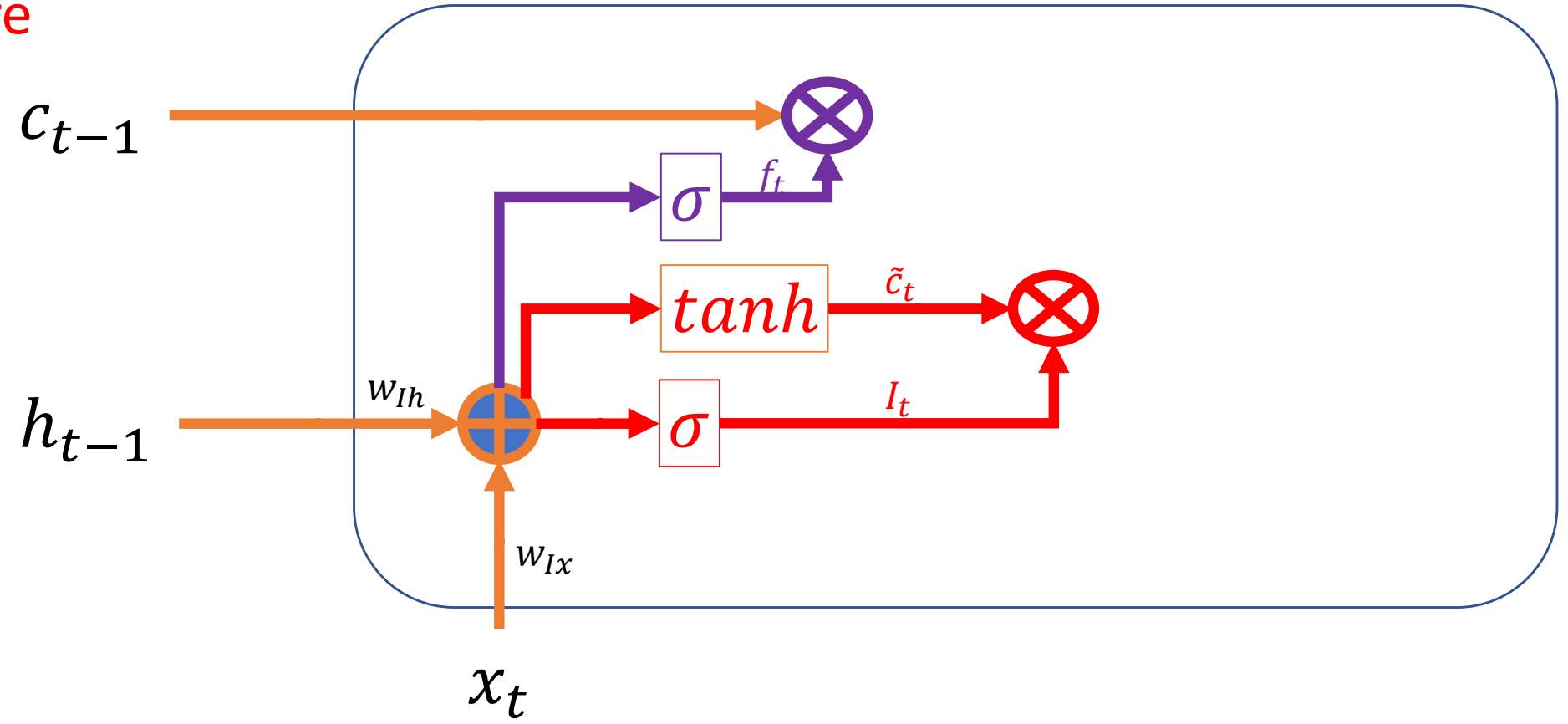


LSTM

- Forget
- Input / Store

$$\tilde{c}_t = \sigma(w_{cx}x_t + w_{ch}h_t)$$

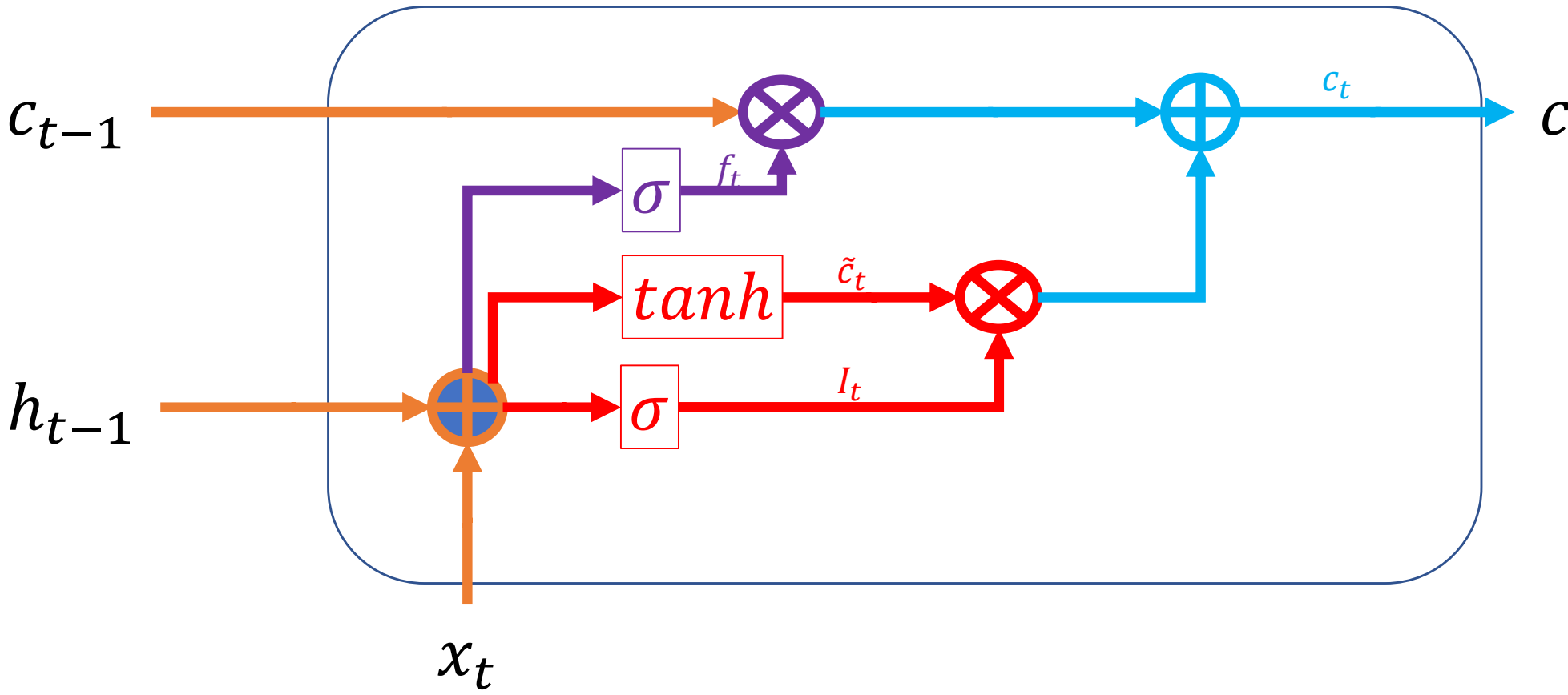
$$I_t = \sigma(w_{Ix}x_t + w_{Ih}h_t)$$



LSTM

- Forget
- Store
- Cell update

$$c_t = f_t c_{t-1} + \tilde{c}_t I_t$$

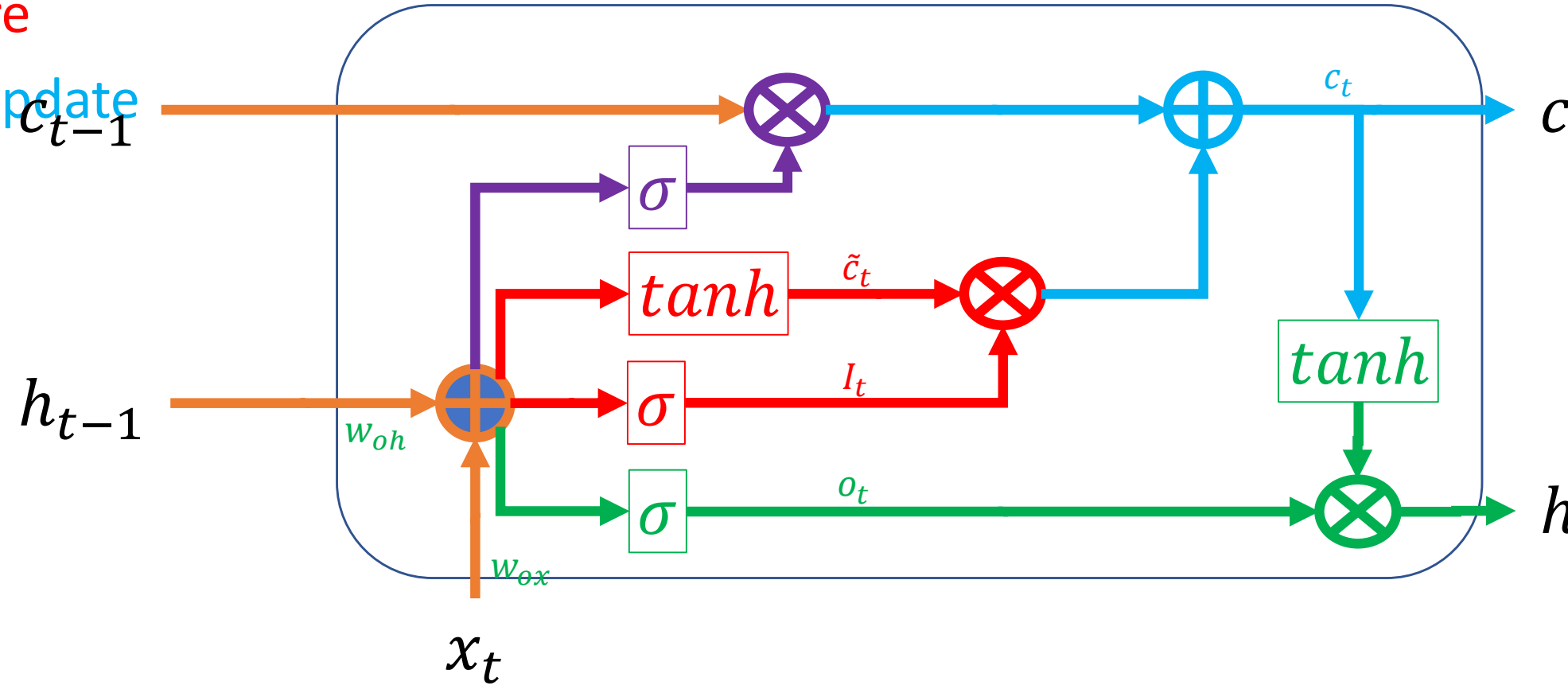


LSTM

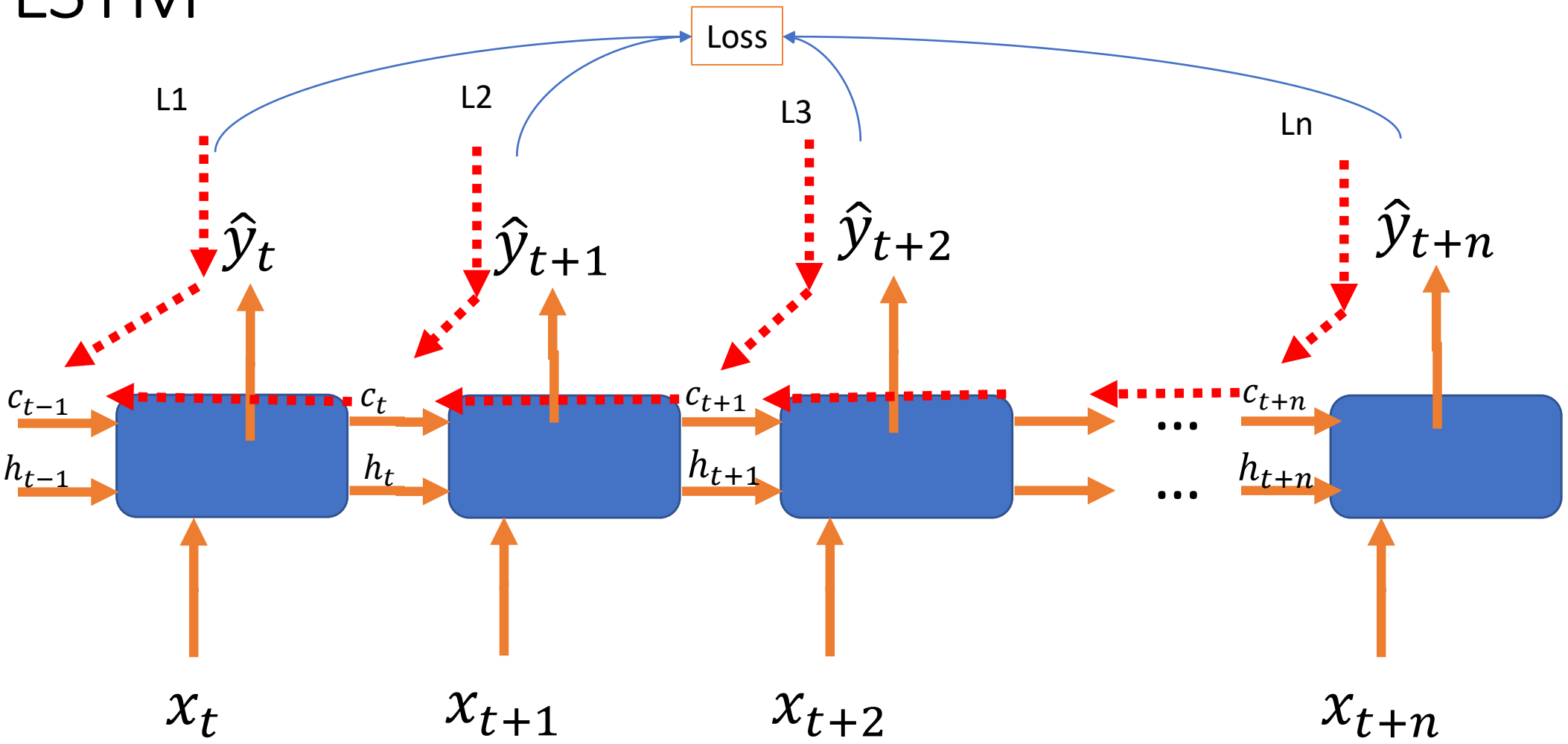
- Forget
- Input / Store
- Cell State Update
- Output

$$o_t = \sigma(w_{ox}x_t + w_{oh}h_{t-1})$$

$$h_t = tanh(c_t) \otimes o_t$$

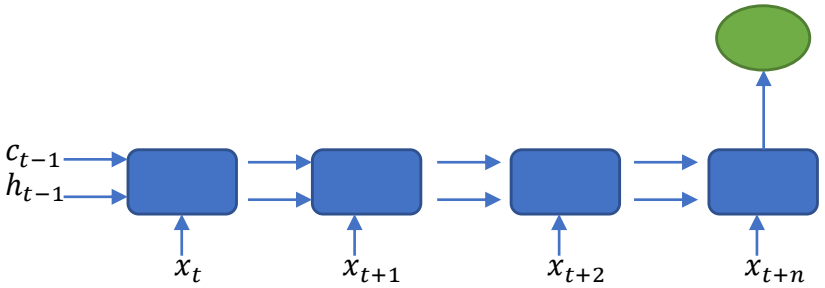


LSTM

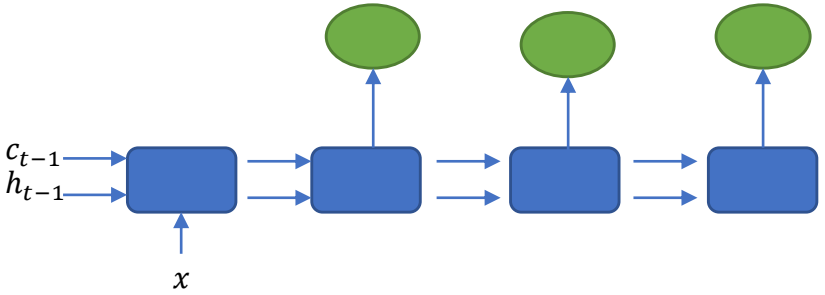

Loss
L1
L2
L3
Ln
$\hat{y}_t$
$\hat{y}_{t+1}$
$\hat{y}_{t+2}$
$\hat{y}_{t+n}$
c_{t-1}
c_t
c_{t+1}
c_{t+n}
h_{t-1}
h_t
h_{t+1}
h_{t+n}
x_t
x_{t+1}
x_{t+2}
x_{t+n}

LSTM Applications

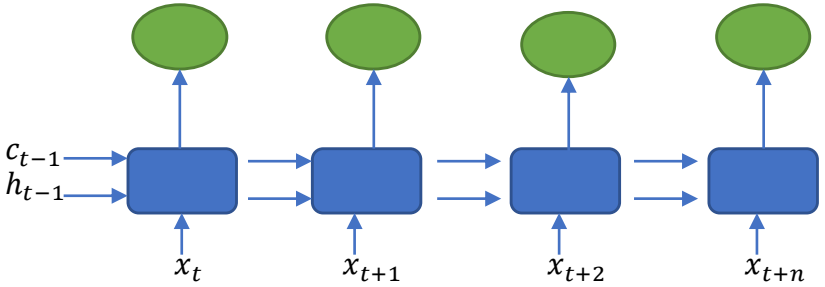
Many to one



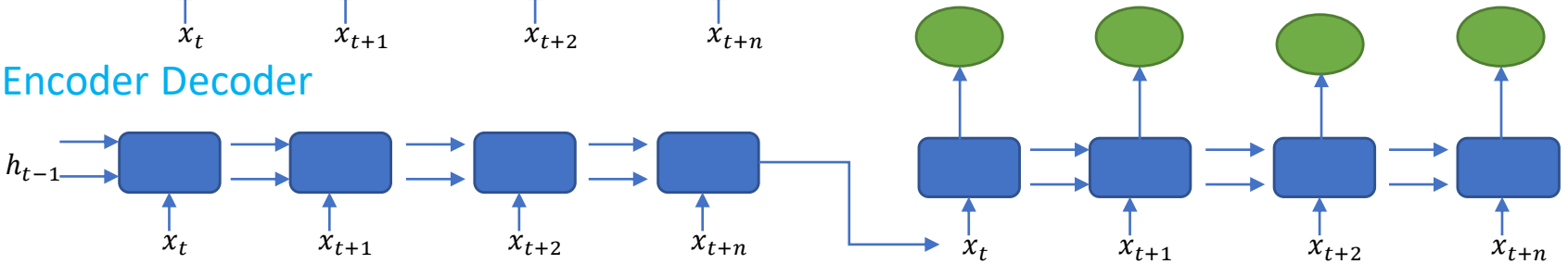
One to Many



Many to Many



Encoder Decoder



Evaluation of Language Models

- Extrinsic
- Intrinsic
  - Held-out data: $\ell(w) = \sum_{m=1}^{M} \log p(w_m | w_{m-1}, \dots, w_1)$
  - $Perplexity(w) = 2^{-\frac{\ell(w)}{M}}$

Summary

- Sequential Language Models
- RNN
  - Variable length computation graph
- LSTM
  - Separate Cell State independent of output
  - Forget, Store, Update, Output