Gated Recurrent Unit (GRU)

GRU

Gated Recurrent Unit:

- Proposed in 2014

GRU

Gated Recurrent Unit:

- Proposed in 2014
- Simpler alternative to LSTM (fewer gates)

GRU

Gated Recurrent Unit:

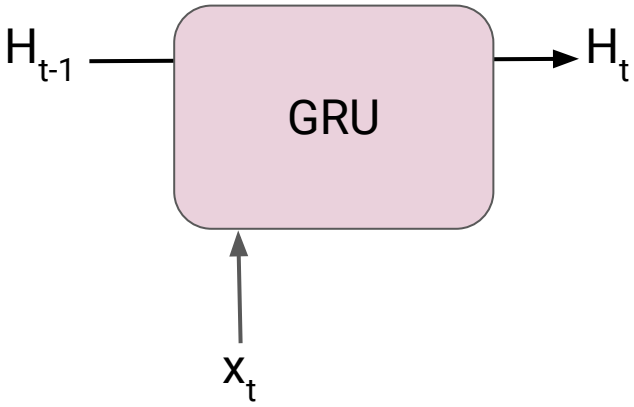
- Proposed in 2014
- Simpler alternative to LSTM (fewer gates)
- No separate cell state (C_t)

GRU

Gated Recurrent Unit:

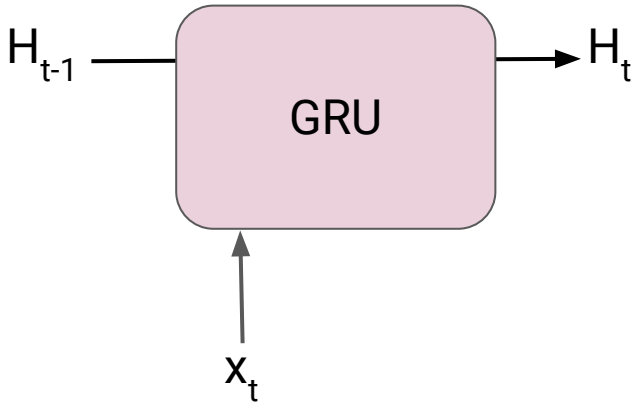
- Proposed in 2014
- Simpler alternative to LSTM (fewer gates)
- No separate cell state (C_t)
- Faster training

GRU



GRU

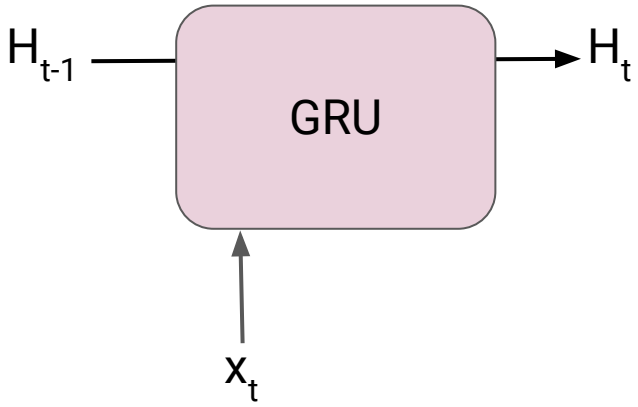
- Reset Gate (short-term memory)



GRU

- Reset Gate (short-term memory)

$$r_t = \sigma\left(x_t * U_r + H_{t-1} * W_r\right)$$

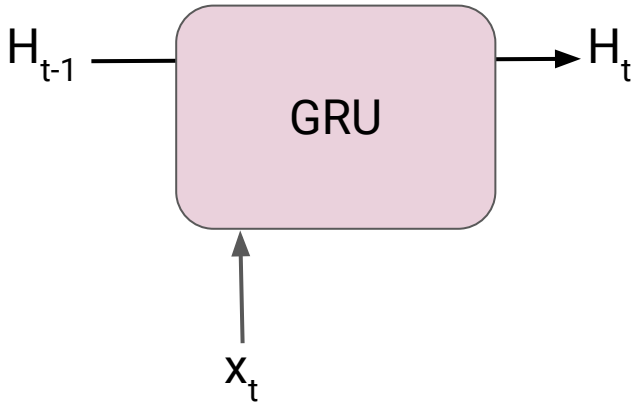


GRU

- Reset Gate (short-term memory)

  $r_t = \sigma (x_t * U_r + H_{t-1} * W_r)$

- Update Gate (long-term memory)



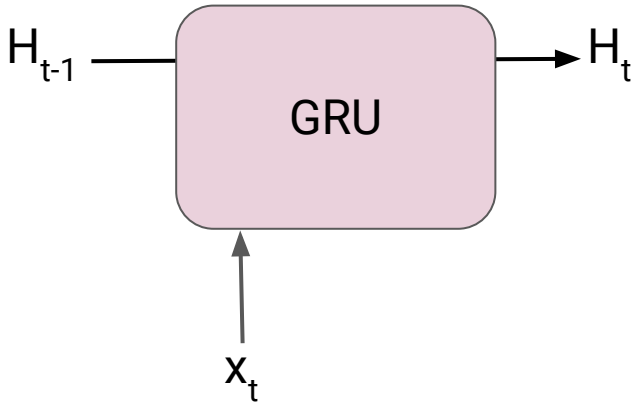
GRU

- Reset Gate (short-term memory)

  $r_t = \sigma\,(x_t * U_r + H_{t-1} * W_r)$

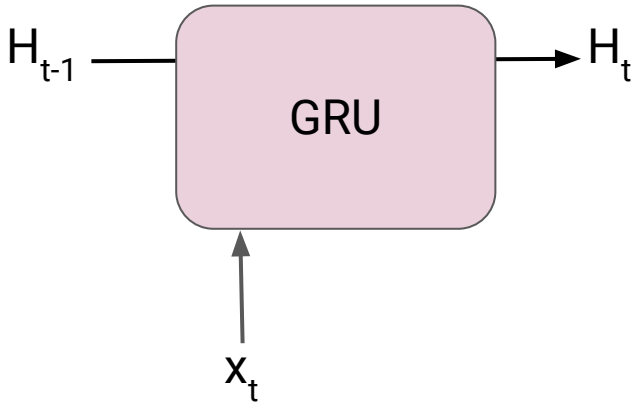
- Update Gate (long-term memory)

  $u_t = \sigma\,(x_t * U_u + H_{t-1} * W_u)$



GRU

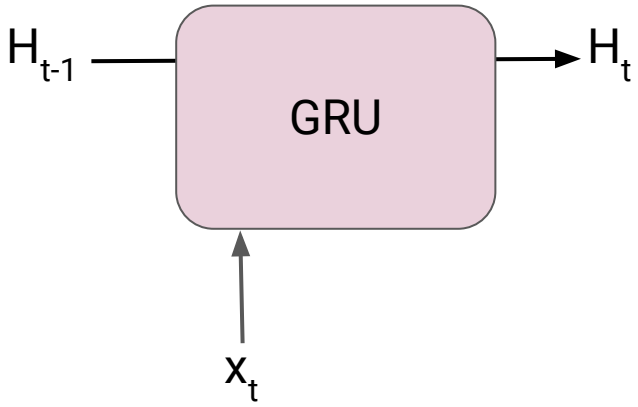
- Find candidate hidden state using new info



GRU

- Find candidate hidden state using new info

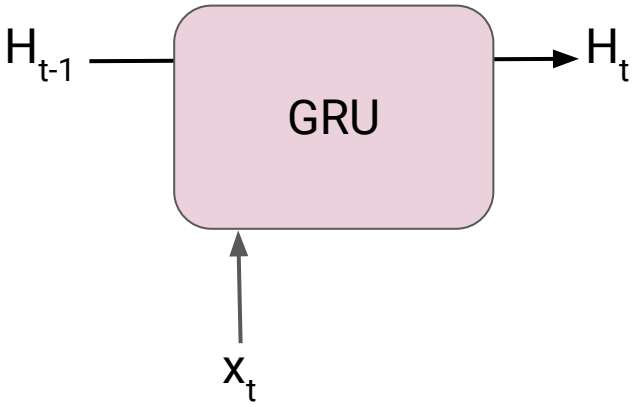
$$\hat{H}_t = \tanh(x_t * U_g + (r_t \circ H_{t-1}) * W_g)$$



GRU

- Find candidate hidden state using new info

$$\hat{H}_t = \tanh(x_t * U_g + (r_t \circ H_{t-1}) * W_g)$$

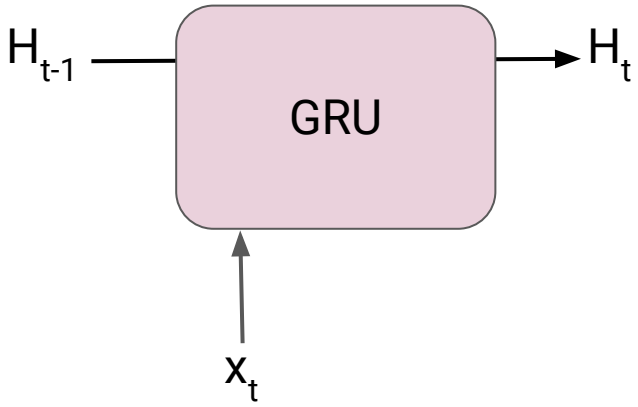


GRU

- Find candidate hidden state using new info

  $\hat{H}_t = \tanh(x_t * U_g + (r_t \circ H_{t-1}) * W_g)$

- Update the hidden state using the update gate



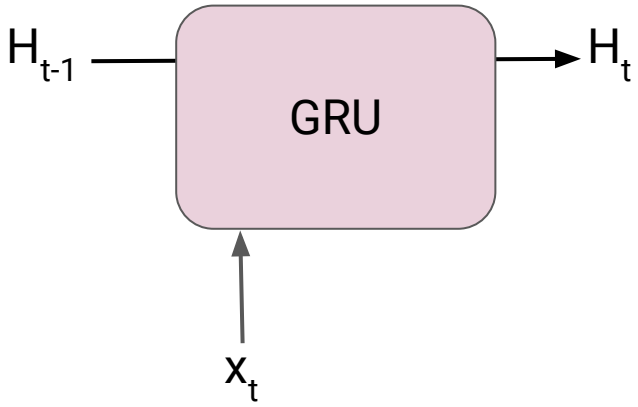
GRU

- Find candidate hidden state using new info

  $\hat{H}_t = \tanh(x_t * U_g + (r_t \circ H_{t-1}) * W_g)$

- Update the hidden state using the update gate

  $H_t = u_t \circ H_{t-1} + (1-u_t) \circ \hat{H}_t$



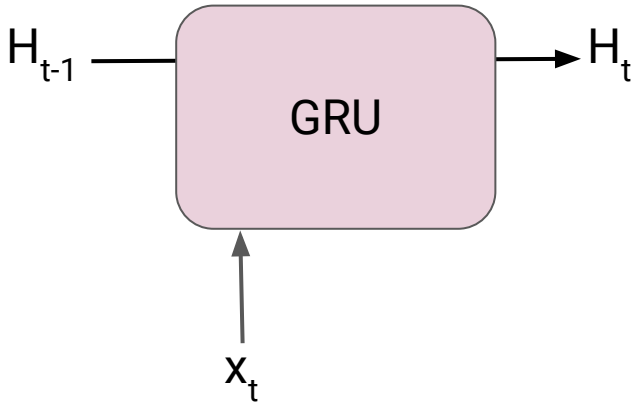
GRU

- Find candidate hidden state using new info

  $\hat{H}_t = \tanh(x_t * U_g + (r_t \circ H_{t-1}) * W_g)$

- Update the hidden state using the update gate

  $H_t = u_t \circ H_{t-1} + (1-u_t) \circ \hat{H}_t$



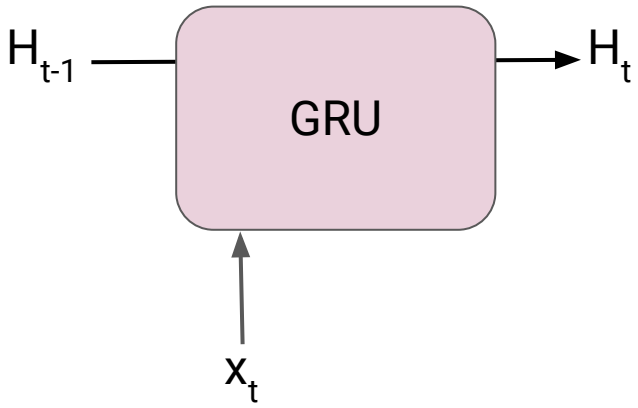
GRU

- Find candidate hidden state using new info

  $\hat{H}_t = \tanh(x_t * U_g + (r_t \circ H_{t-1}) * W_g)$

- Update the hidden state using the update gate

  $H_t = u_t \circ H_{t-1} + (1-u_t) \circ \hat{H}_t$



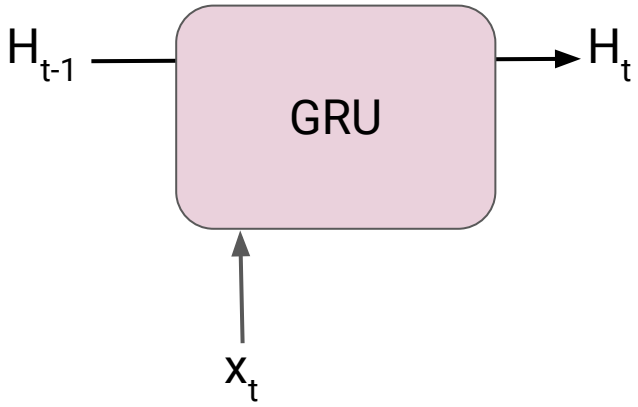
GRU

- Find candidate hidden state using new info

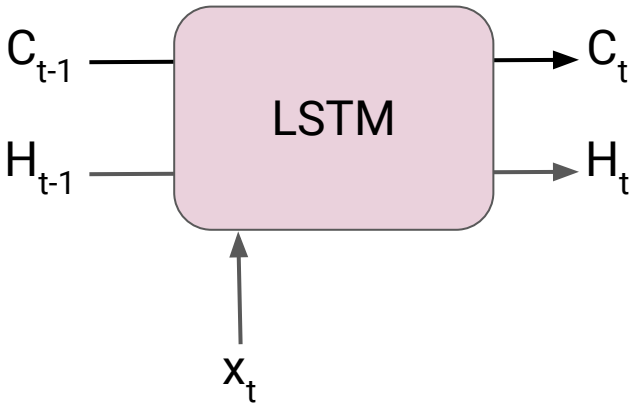
  $\hat{H}_t = \tanh(x_t * U_g + (r_t \circ H_{t-1}) * W_g)$

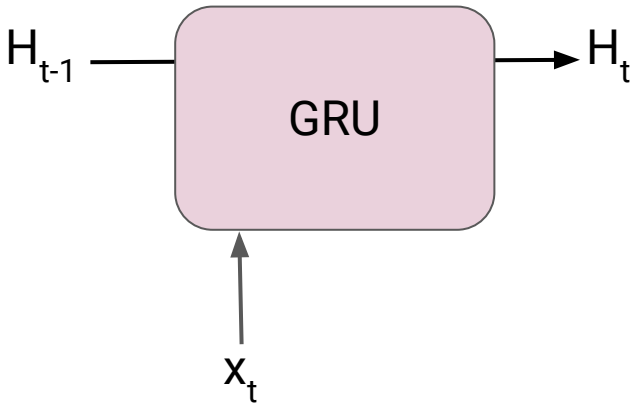
- Update the hidden state using the update gate

  $H_t = u_t \circ H_{t-1} + (1-u_t) \circ \hat{H}_t$

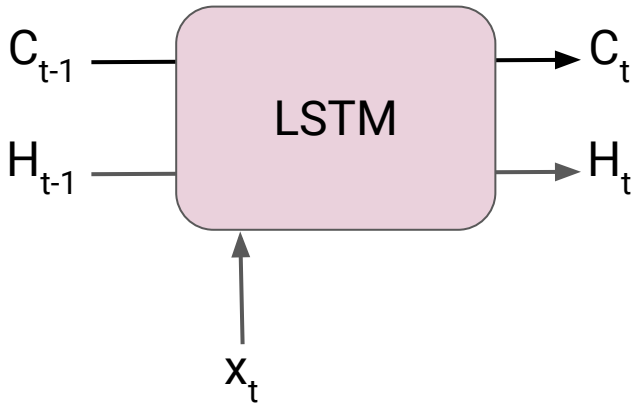


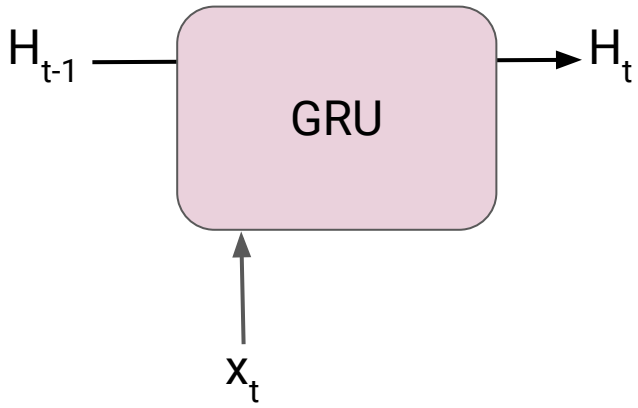
GRU





GRU

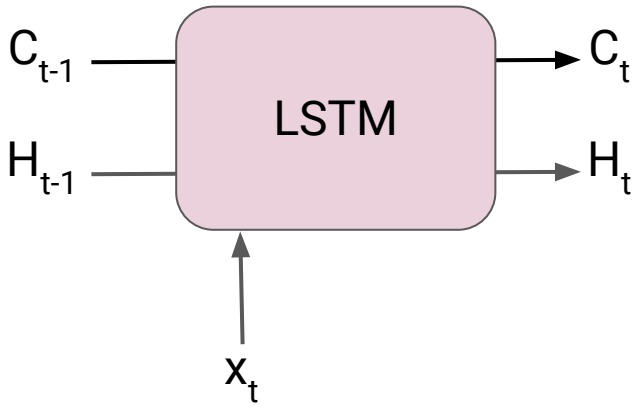


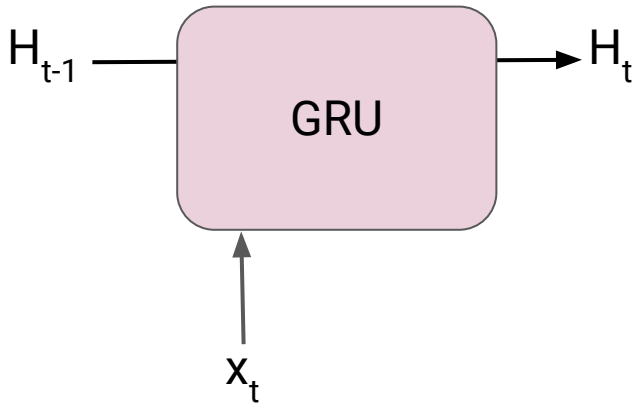


- 3 Gates

- 2 Gates

GRU

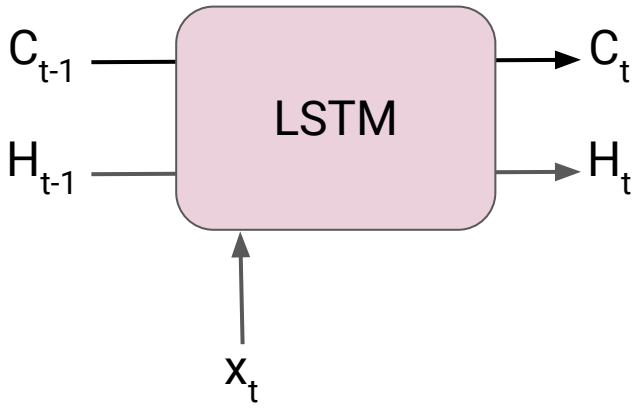


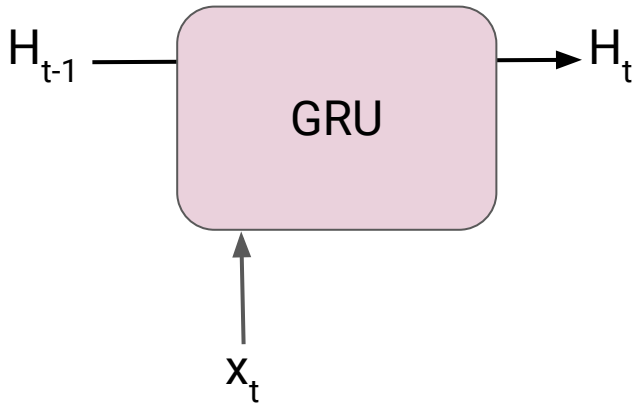


- 3 Gates
- Input, Output, and Forget

- 2 Gates
- Update and Reset

GRU





- 3 Gates
- Input, Output, and Forget
- Separate cell state C_t and Hidden state H_t

- 2 Gates
- Update and Reset
- Single Hidden state H_t

Thank You