

Gated Recurrent Units

Lecture 22-23



Gated Recurrent Unit (GRU)



Paying attention to a sequence

- Not all observations are equally relevant



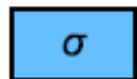
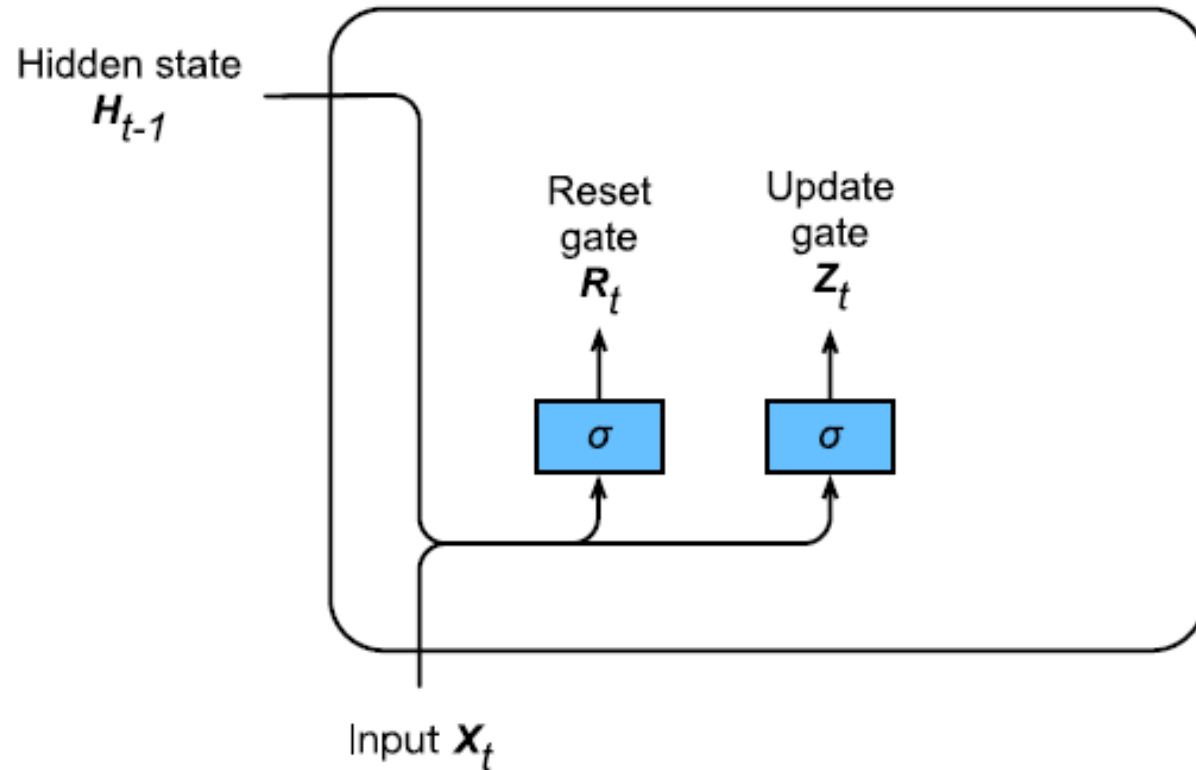
- Only remember the relevant ones
 - Need mechanism to **pay attention (update gate)**
 - Need mechanism to **forget (reset gate)**



Gating

$$R_t = \sigma(X_t W_{xr} + H_{t-1} W_{hr} + b_r),$$

$$Z_t = \sigma(X_t W_{xz} + H_{t-1} W_{hz} + b_z)$$



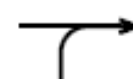
FC layer with
activation function



Element-wise
Operator



Copy

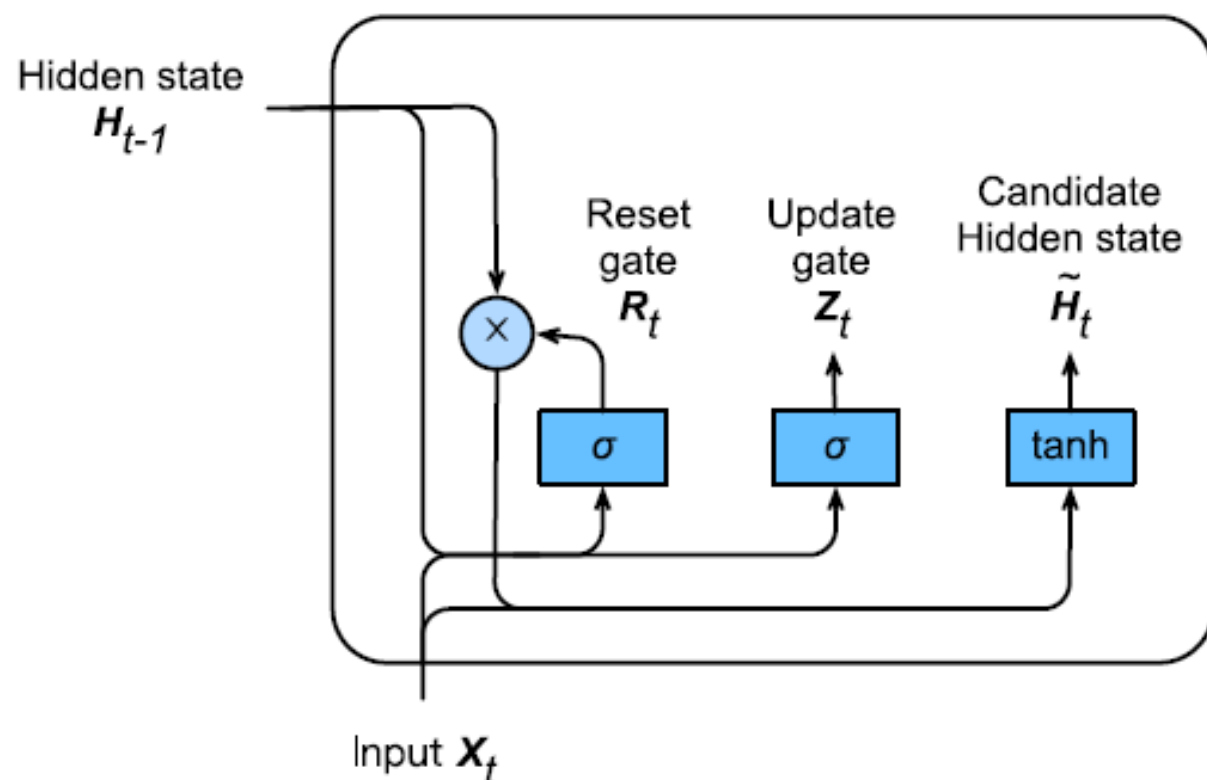


Concatenate



Candidate Hidden State

$$\tilde{H}_t = \tanh(X_t W_{xh} + (R_t \odot H_{t-1}) W_{hh} + b_h)$$



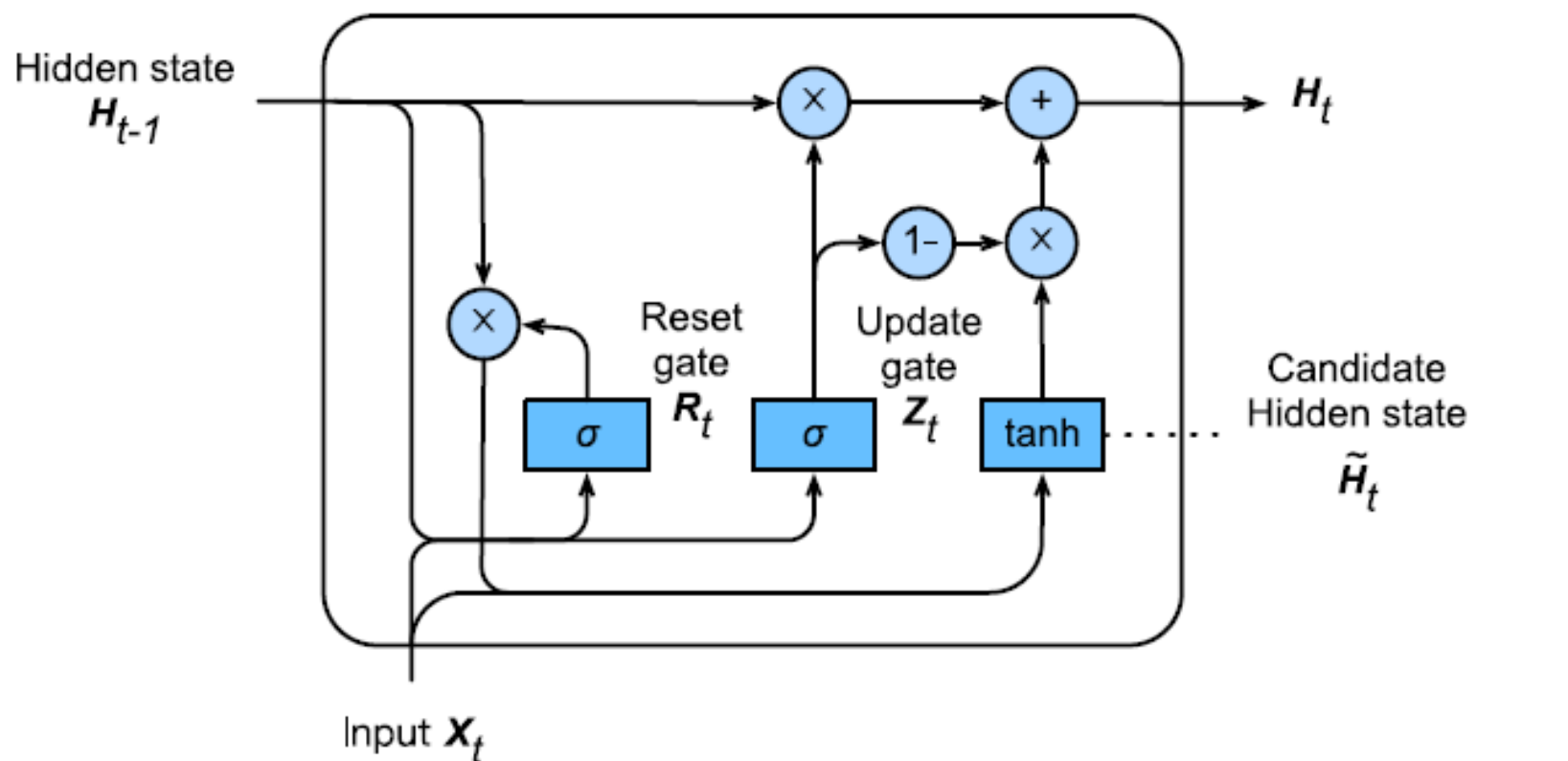
Gate

$$\begin{matrix} R_t & \times & H_{t-1} \\ \begin{bmatrix} 0.3 \\ 0 \\ 0.6 \\ 0.4 \\ 0.1 \\ 1 \\ 0 \\ 0.5 \end{bmatrix} & \times & \begin{bmatrix} 2 \\ 3 \\ 1 \\ 5 \\ 3 \\ 0.4 \\ 4 \\ 1 \end{bmatrix} & = & \begin{bmatrix} 0.6 \\ 0 \\ 0.6 \\ 2 \\ 0.3 \\ 0.4 \\ 0 \\ 0.5 \end{bmatrix} \end{matrix}$$



Hidden State

$$H_t = Z_t \odot H_{t-1} + (1 - Z_t) \odot \tilde{H}_t$$



Summary

$$R_t = \sigma(X_t W_{xr} + H_{t-1} W_{hr} + b_r),$$

$$Z_t = \sigma(X_t W_{xz} + H_{t-1} W_{hz} + b_z)$$

$$\tilde{H}_t = \tanh(X_t W_{xh} + (R_t \odot H_{t-1}) W_{hh} + b_h)$$

$$H_t = Z_t \odot H_{t-1} + (1 - Z_t) \odot \tilde{H}_t$$

