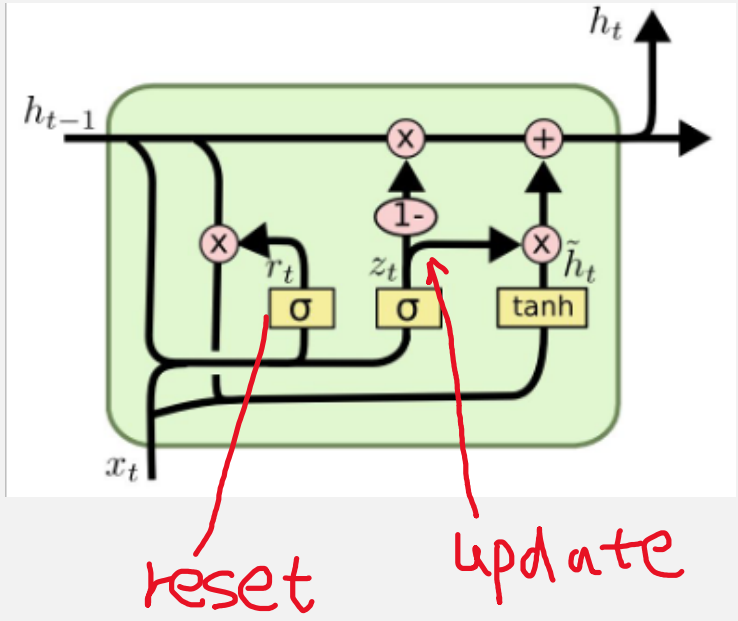
LSTM vs GRU

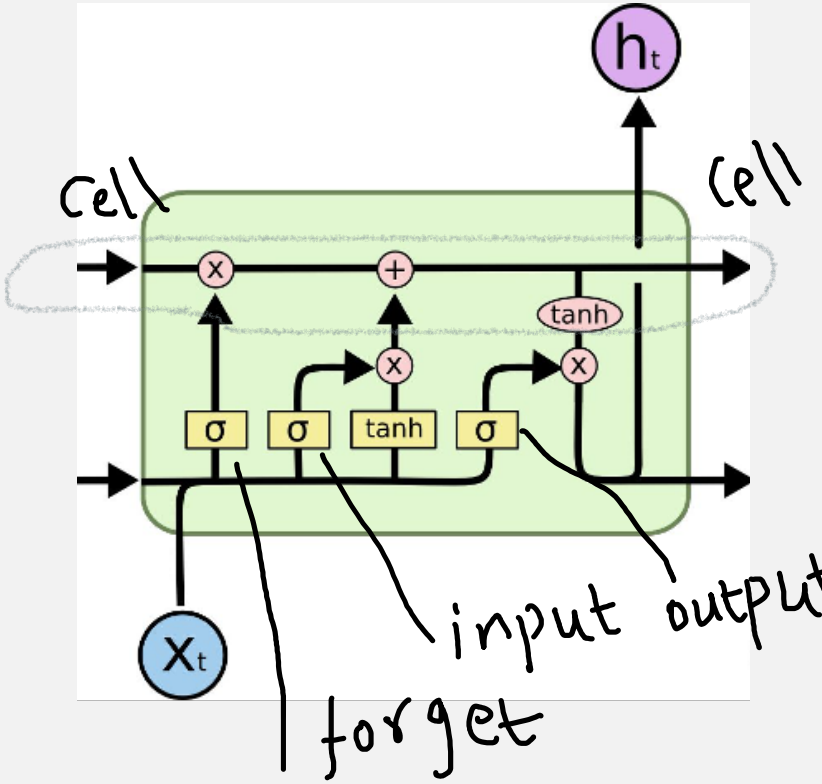
Dr. Min Chi
mchi@ncsu.edu

AI Academy

GRU

Similar with LSTM but with only two gates and less parameters. Fewer Gates:
-Combine forget and input gates into “updated” grate
-Eliminates cell state vector





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

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

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

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

The “update gate” determines how much of previous memory to be kept.
The “reset gate” determines how to combine the new input with the previous memory by determining what should be removed from the cell’s internal state before passing itself along to the next time step

GRU vs. LSTM

- **GRU has significantly fewer parameters and trains faster**
- Experimental results comparing the two are still inconclusive, many problems they perform the same, but each has problems on which they work better