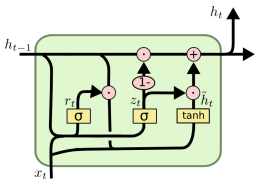
LSTM and GRU Cells for RNN

- LSTM, Long Short-Term Memory Cells (Hochreiter and Schmidhuber, 1997).
- GRU, Gated Recurrent Unit Cells (Chung et al., 2014):



$$z_t = \sigma\left(W_z[h_{t-1}; x_t] + b_z\right) \quad (4)$$

$$r_t = \sigma\left(W_r[h_{t-1}; x_t] + b_r\right) \quad (5)$$

$$\tilde{h}_t = \tanh\left(W[r_t \odot h_{t-1}; x_t]\right) \quad (6)$$

$$h_t = (1 - z_t) \odot h_{t-1} + z_t \odot \tilde{h}_t \quad (7)$$

- Gates control:
  - what to use from input x_t (GRU: everything),
  - what to use from hidden state h_{t-1} (reset gate r_t),
  - what to put into output (update gate z_t)
- Linear "information highway" preserved.

  $\Rightarrow$ All states h_t belong to the same vector space.