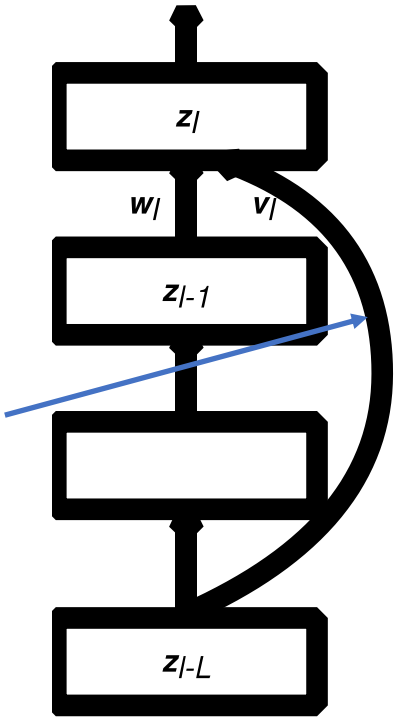
Intro to ML

December 13th, 2021

CHAPTER 12:

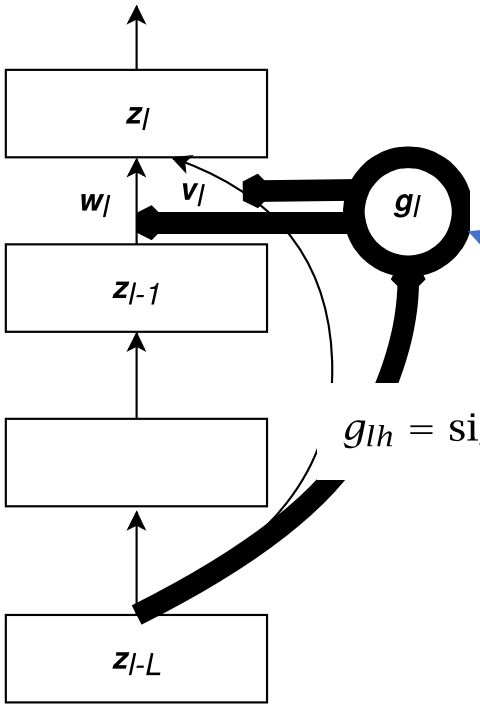
Deep Learning

More ways to handle vanishing gradient
Skip Connections



Skip connection

$$z_{lh} = f\left(\sum_i w_{l,h,i} z_{l-1,i} + \sum_j v_{l,h,j} z_{l-L,j}\right)$$



Gating unit

$$g_{lh} = \text{sigmoid}\left(\boldsymbol{a}_{lh}^T \boldsymbol{z}_{l-L} + a_{lh0}\right)$$

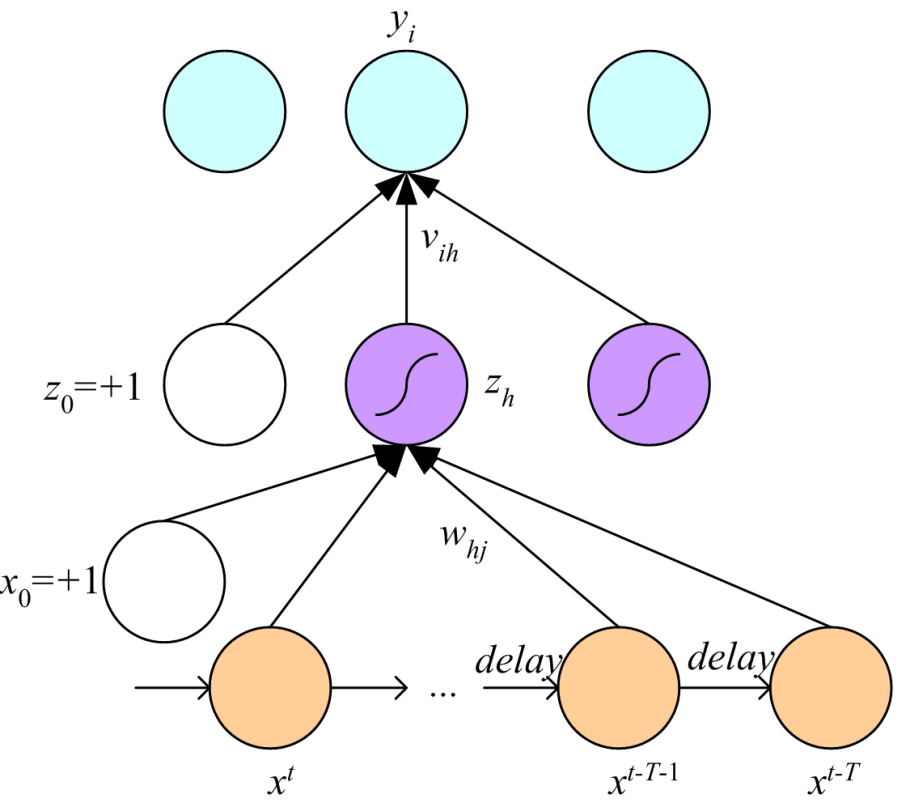
$$z_{lh} = f\left(g_{lh} \sum_i w_{lhi} z_{l-1,i} + (1 - g_{lh}) \sum_j v_{lhj} z_{l-L,j}\right)$$

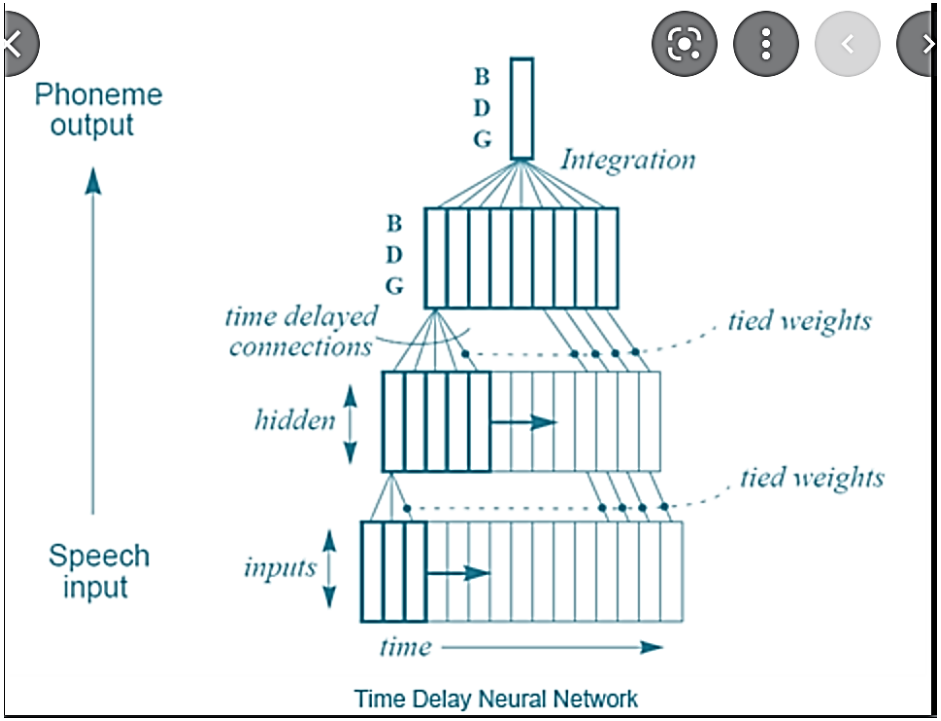
Learning Time (sequence modeling)

- Applications:
  - Sequence recognition: Speech recognition
  - Sequence reproduction: Time-series prediction

- Network architectures
  - Time-delay networks (Waibel et al., 1989)
  - Recurrent networks (Rumelhart et al., 1986)

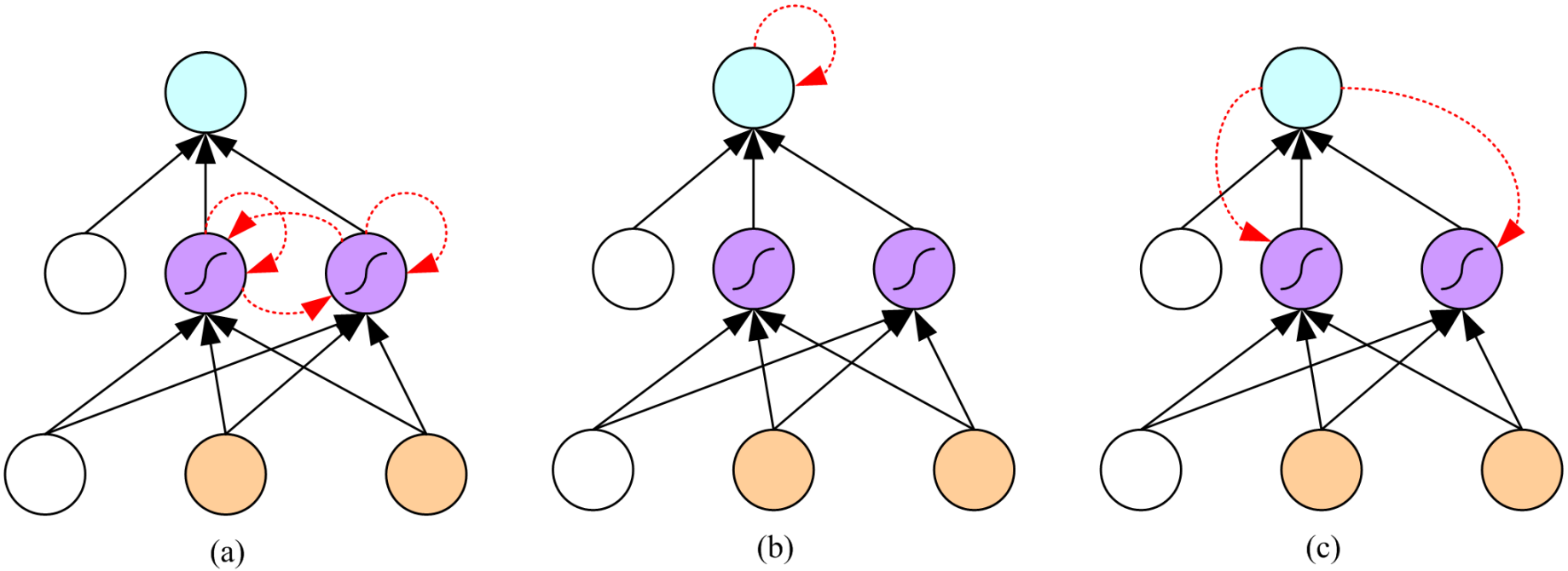
Time-Delay Neural Networks (TDNN)





It is also just like a 1-D convolution neural network

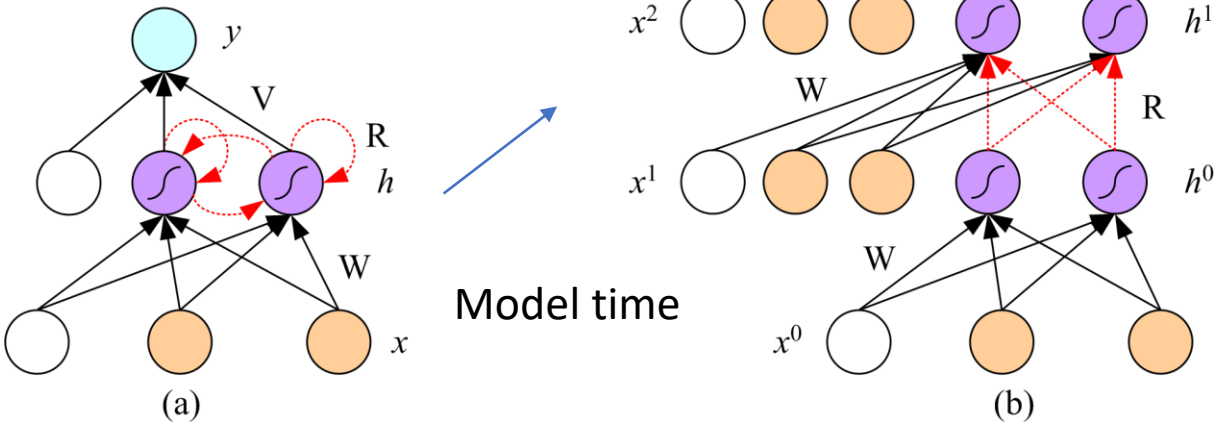
Recurrent Networks(RNN)



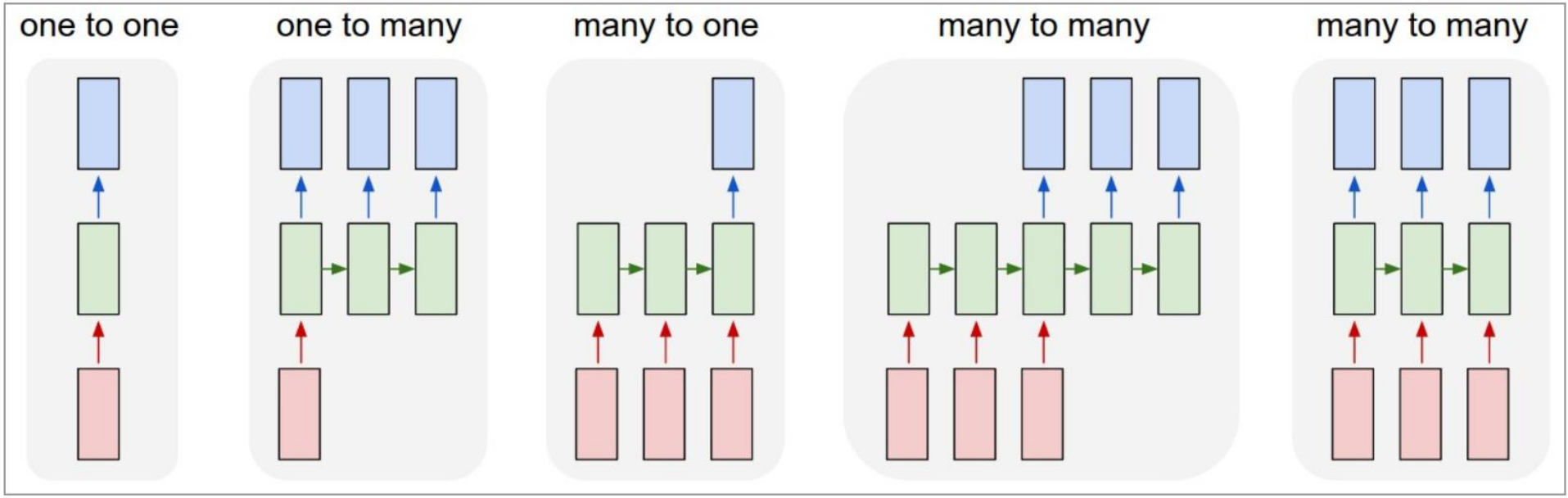
Unfolding in Time -> sequence model

$$z_h^t = f\left(\sum_{j=0}^{d} w_{hj} x_j^t + \sum_{l=1}^{H} r_{hl} z_l^{t-1}\right)$$

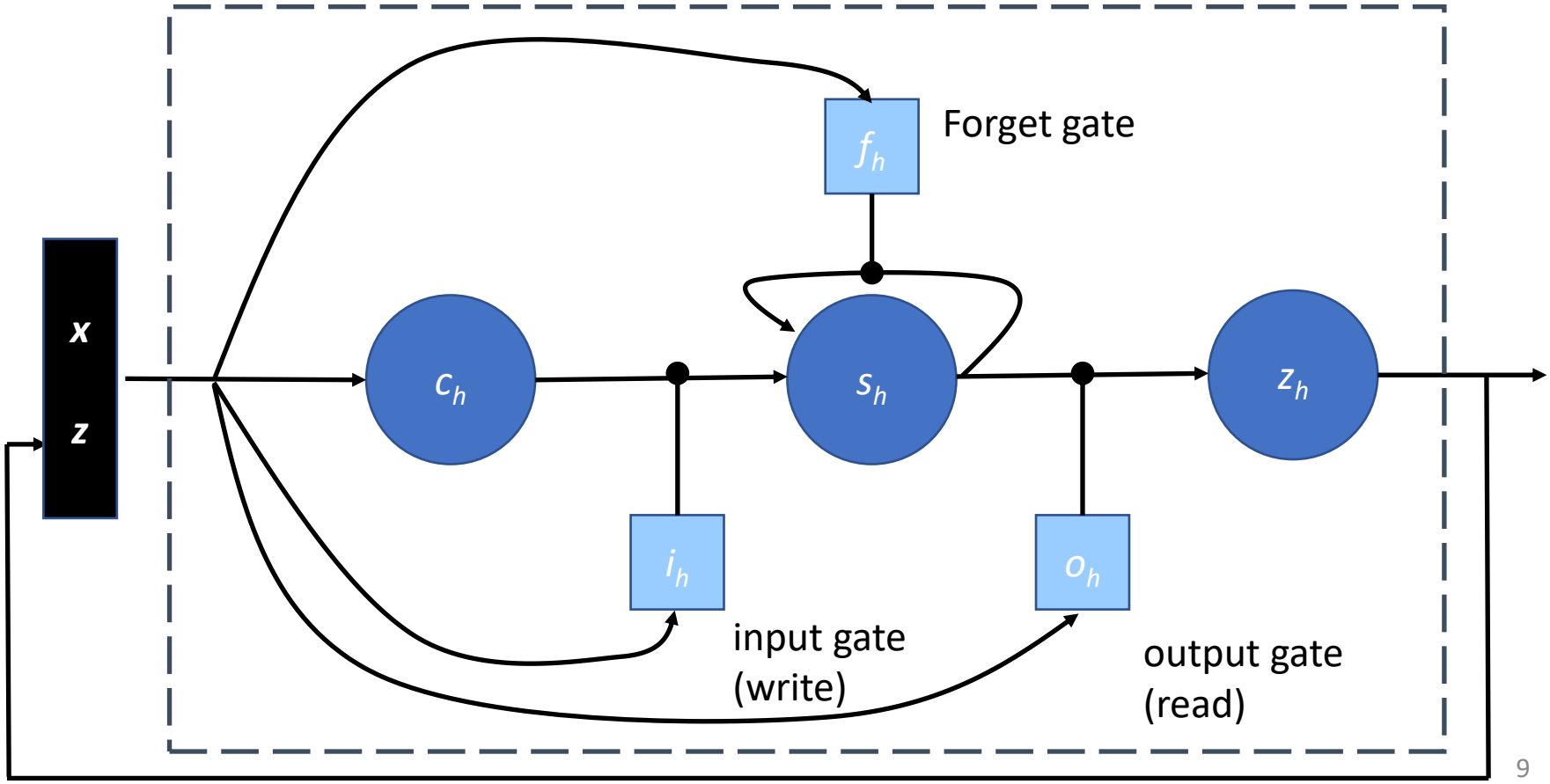
$$y^t = g\left(\sum_{h=0}^{H} v_h z_h^t\right)$$



Many variants of RNN for different applications

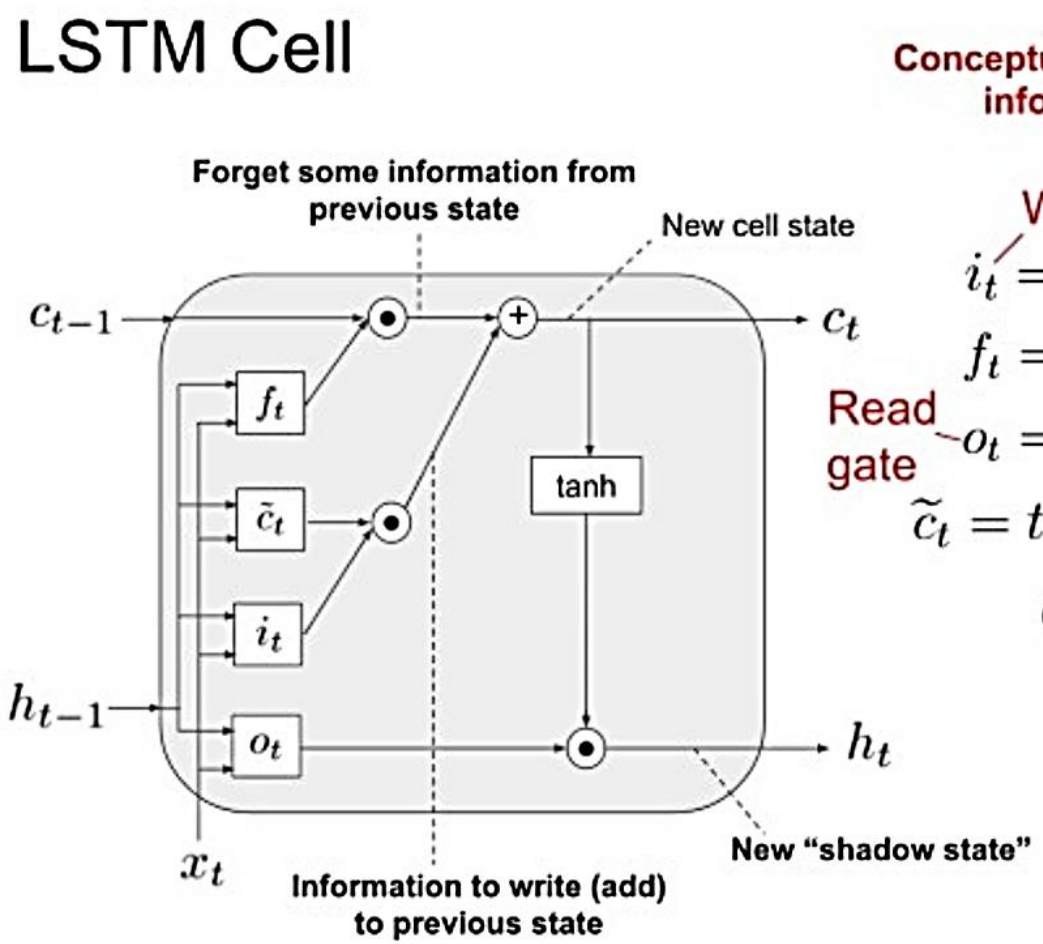


Long Short-Term Memory (LSTM)



Long short-term memory (LSTM)

Activation function



Conceptually, we lose information

We use shadow state to calculate gates

LSTM equations

Write gate

Read gate

$$i_t = \sigma(W_i h_{t-1} + U_i x_t + b_i)$$ Input gate

$$f_t = \sigma(W_f h_{t-1} + U_f x_t + b_f)$$ Forget gate

$$o_t = \sigma(W_o h_{t-1} + U_o x_t + b_o)$$ Output gate

$$\tilde{c}_t = tanh(W h_{t-1} + U x_t + b)$$ Memory cell candidate

$$c_t = f_t \circ c_{t-1} + i_t \circ \tilde{c}_t$$ Memory cell

$$h_t = o_t \circ tanh(c_t)$$ Shadow state

$$y_t = h_t$$ Cell Output

Read occurs after writing