

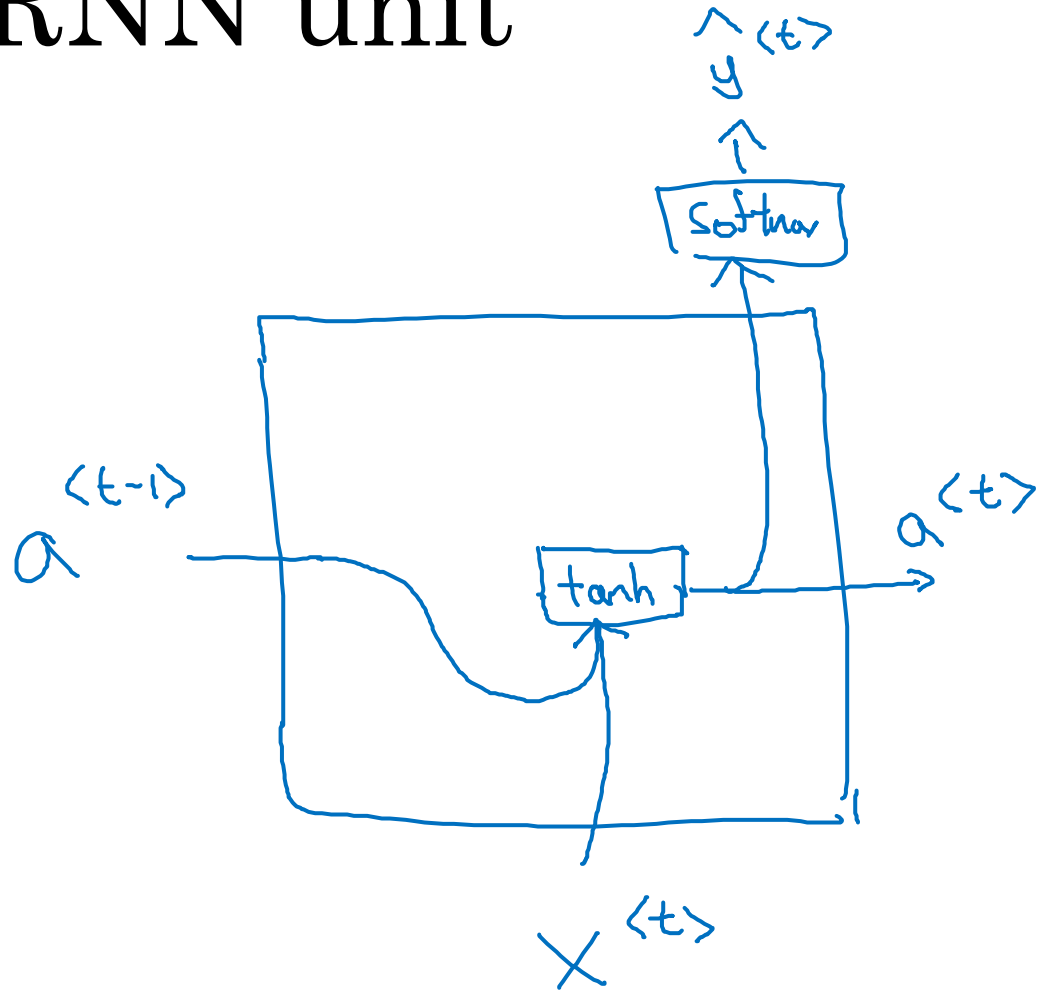


deeplearning.ai

Recurrent Neural Networks

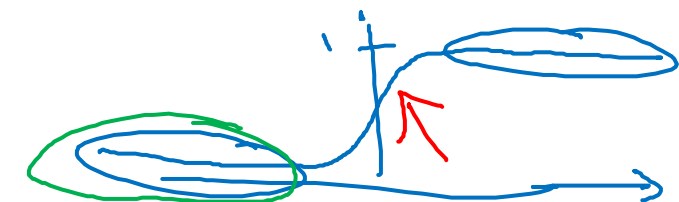
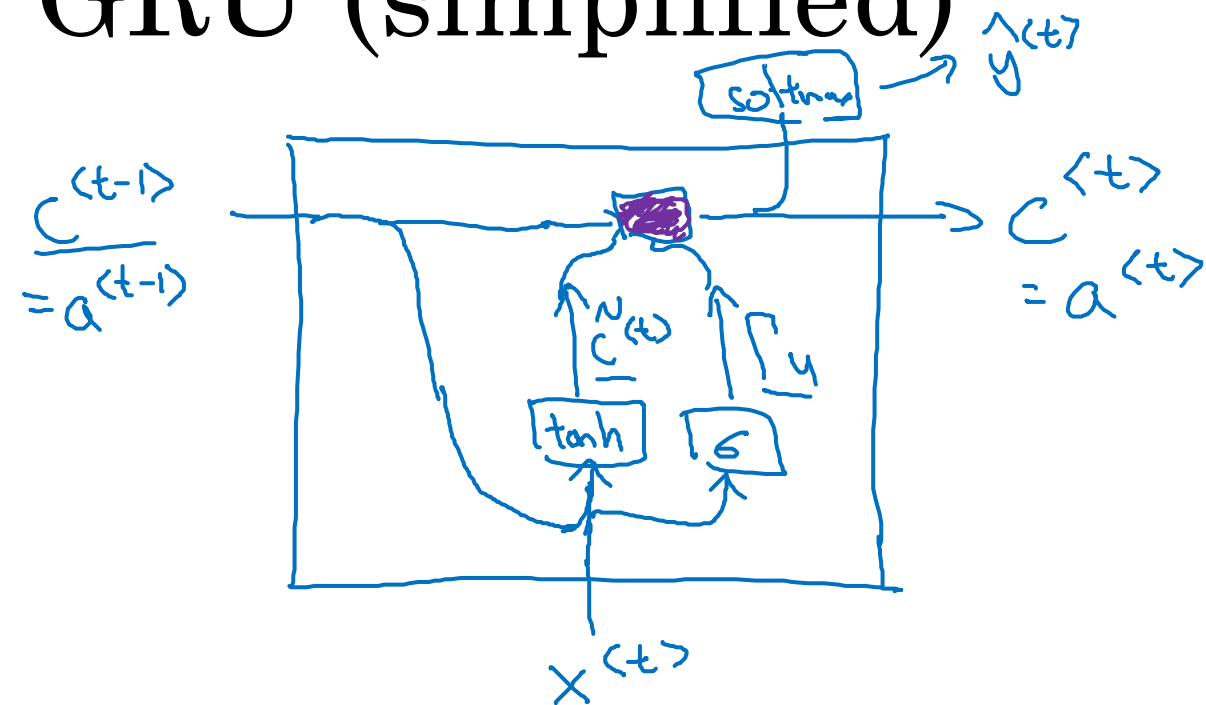
Gated Recurrent Unit (GRU)

RNN unit



$$\underline{a^{<t>}} = \overset{\substack{\text{tanh} \\ \downarrow}}{g}(\underbrace{W_a[a^{<t-1>}, x^{<t>}]}_{\uparrow} + b_a)$$

GRU (simplified)



C = memory cell

$$\rightarrow \boxed{C^{(t)}} = \underline{a^{(t)}}$$

$$\rightarrow \boxed{\tilde{C}^{(t)}} = \tanh(W_c [C^{(t-1)}, x^{(t)}] + b_c)$$

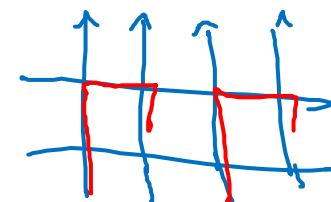
$$\rightarrow \boxed{\Gamma_u^{(t)}} = \sigma(W_u [C^{(t-1)}, x^{(t)}] + b_u)$$

$$\boxed{C^{(t)}} = \underbrace{\Gamma_u^{(t)}}_{\text{"update"}} * \tilde{C}^{(t)} + (1 - \Gamma_u^{(t)}) * \boxed{C^{(t-1)}}$$

element-wise

$$\Gamma_u = 0.000001$$

Gate



$\Gamma_u = 1$
 $C^{(t)} = 1$
 $\Gamma_u = 0 \quad \Gamma_u = 0 \quad \Gamma_u = 0 \quad \dots$
 The cat, which already ate ..., was full.

[Cho et al., 2014. On the properties of neural machine translation: Encoder-decoder approaches]

[Chung et al., 2014. Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling]

Andrew Ng

Full GRU

$$\tilde{c}^{<t>} = \tanh(W_c [\tilde{c}^{<t-1>}, x^{<t>}] + b_c)$$

$$\begin{cases} \Gamma_u = \sigma(W_u [c^{<t-1>}, x^{<t>}] + b_u) \\ \Gamma_r = \sigma(W_r [c^{<t-1>}, x^{<t>}] + b_r) \end{cases}$$

LSTM

$$c^{<t>} = \Gamma_u * \tilde{c}^{<t>} + (1 - \Gamma_u) * c^{<t-1>}$$

The cat, which ate already, was full.