

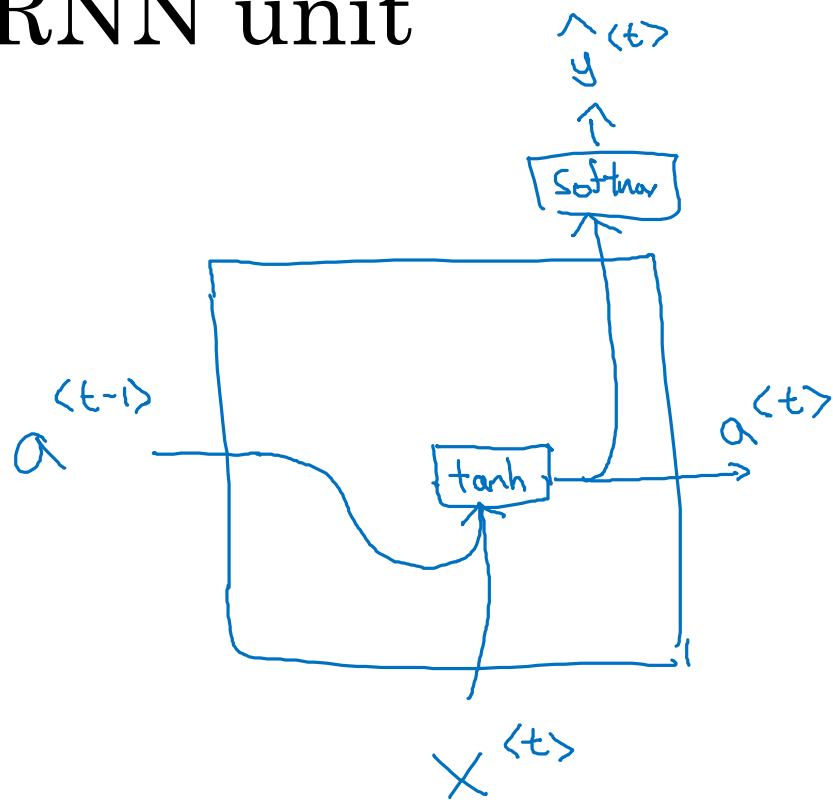


deeplearning.ai

Recurrent Neural Networks

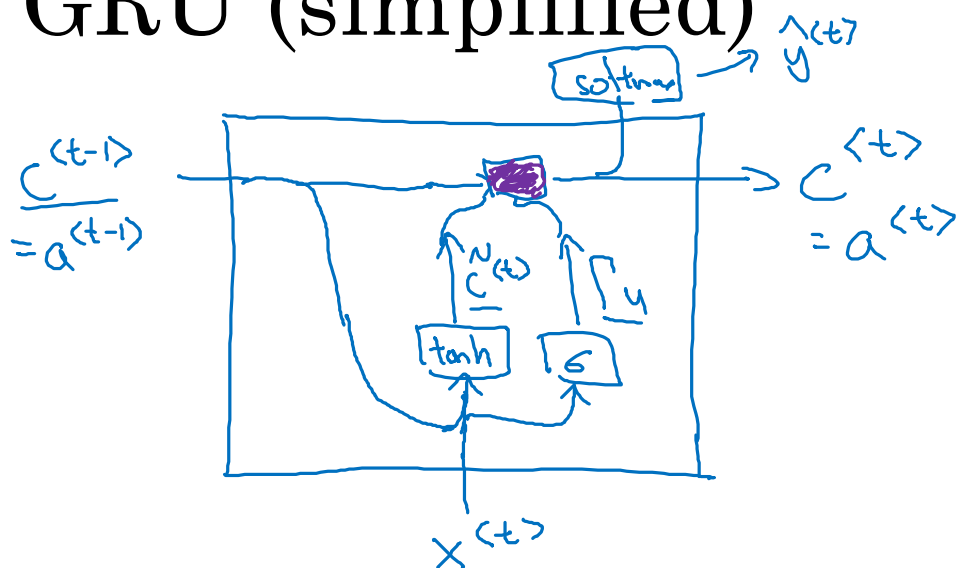
Gated Recurrent Unit (GRU)

RNN unit



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

GRU (simplified)



C = memory cell

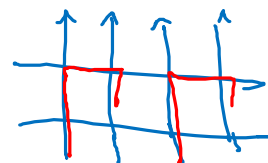
$$\Rightarrow \boxed{C^{(t)}} = \underline{a}^{(t)}$$

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

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

$$\boxed{C^{(t)}} = \underbrace{\Gamma_u}_{\uparrow=1} * \tilde{C}^{(t)} + \underbrace{(1 - \Gamma_u)}_1 * \boxed{C^{(t-1)}}$$

element-wise
Gate



$\Gamma_u = 1$ $\Gamma_u = 0$ $\Gamma_u = 0$ $\Gamma_u = 0$...
 \Rightarrow 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

$$\tilde{c}^{<t>} = \tanh(W_c[\underbrace{c^{<t-1>} * \Gamma_r}_{\text{LSTM}}, x^{<t>}] + b_c)$$

$$\Gamma_u = \sigma(W_u[c^{<t-1>}, x^{<t>}] + b_u)$$

$$\Gamma_r = \sigma(W_r[c^{<t-1>}, x^{<t>}] + b_r)$$

LSTM

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

The cat, which ate already, was full.