



deeplearning.ai

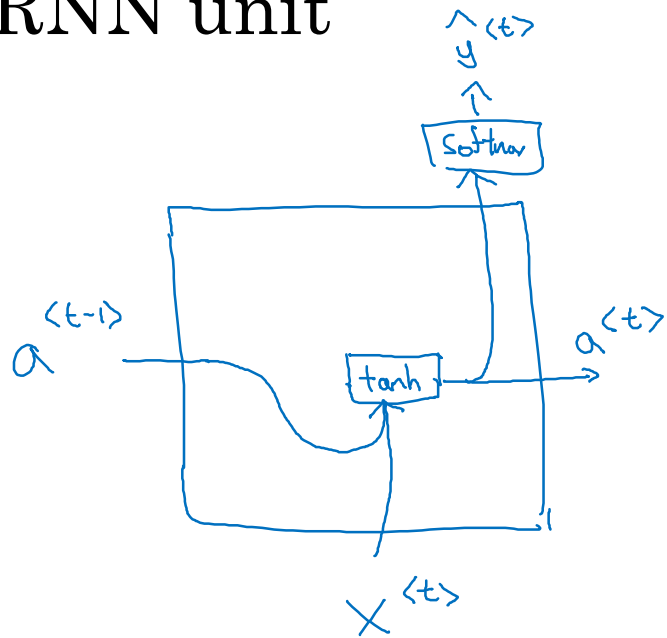
更好地捕捉深层连接
改善了梯度消失问题

Recurrent Neural Networks

门控循环单元

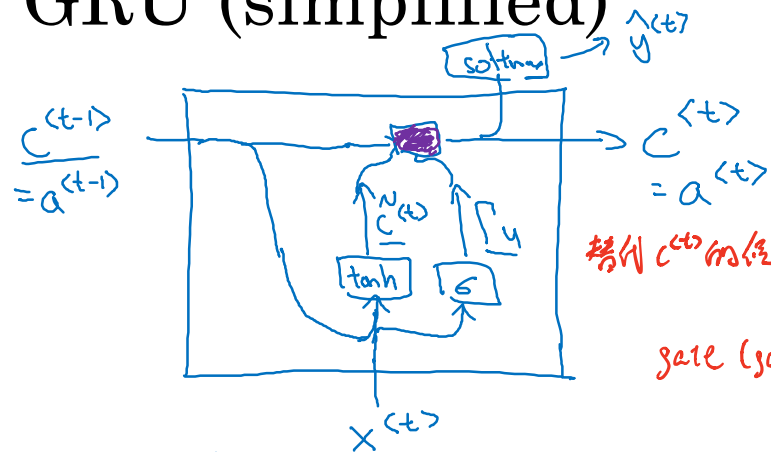
Gated Recurrent Unit (GRU)

RNN unit



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

GRU (simplified)



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

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

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

"update" $\rightarrow \boxed{C^{(t)}} = \Gamma_u * \tilde{C}^{(t)} + (1 - \Gamma_u) * \boxed{C^{(t-1)}}$
 $!!$

element-wise
 Gate

$\Gamma_u = 0.000001$

key !!
 $\Gamma_u = 1$
 $\Gamma_u = 0$ $\Gamma_u = 0$ $\Gamma_u = 0$...
 \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

r: relevance

*$\Gamma_r * c^{<t-1>}$*

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

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

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

LSTM

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

The cat, which ate already, was full.