deeplearning.ai

Recurrent Neural Networks
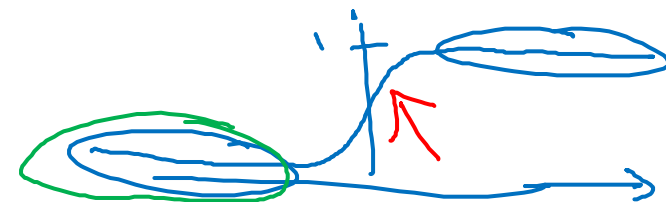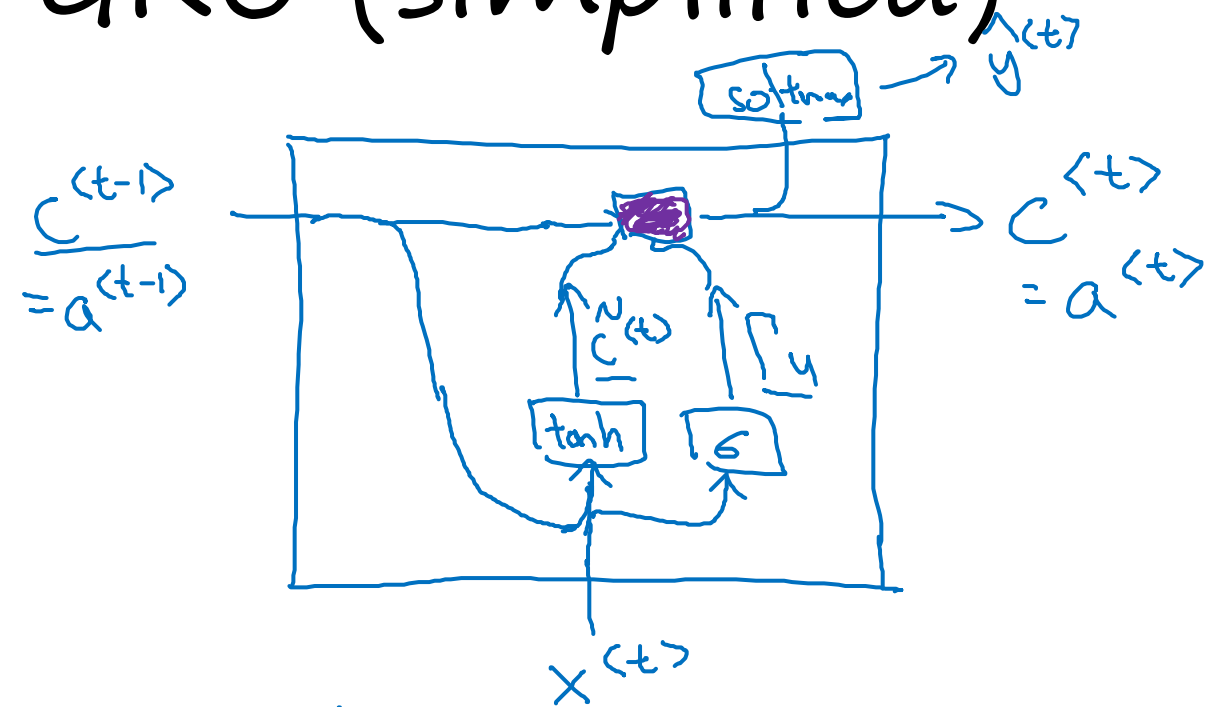

Gated Recurrent Unit (GRU)

RNN unit
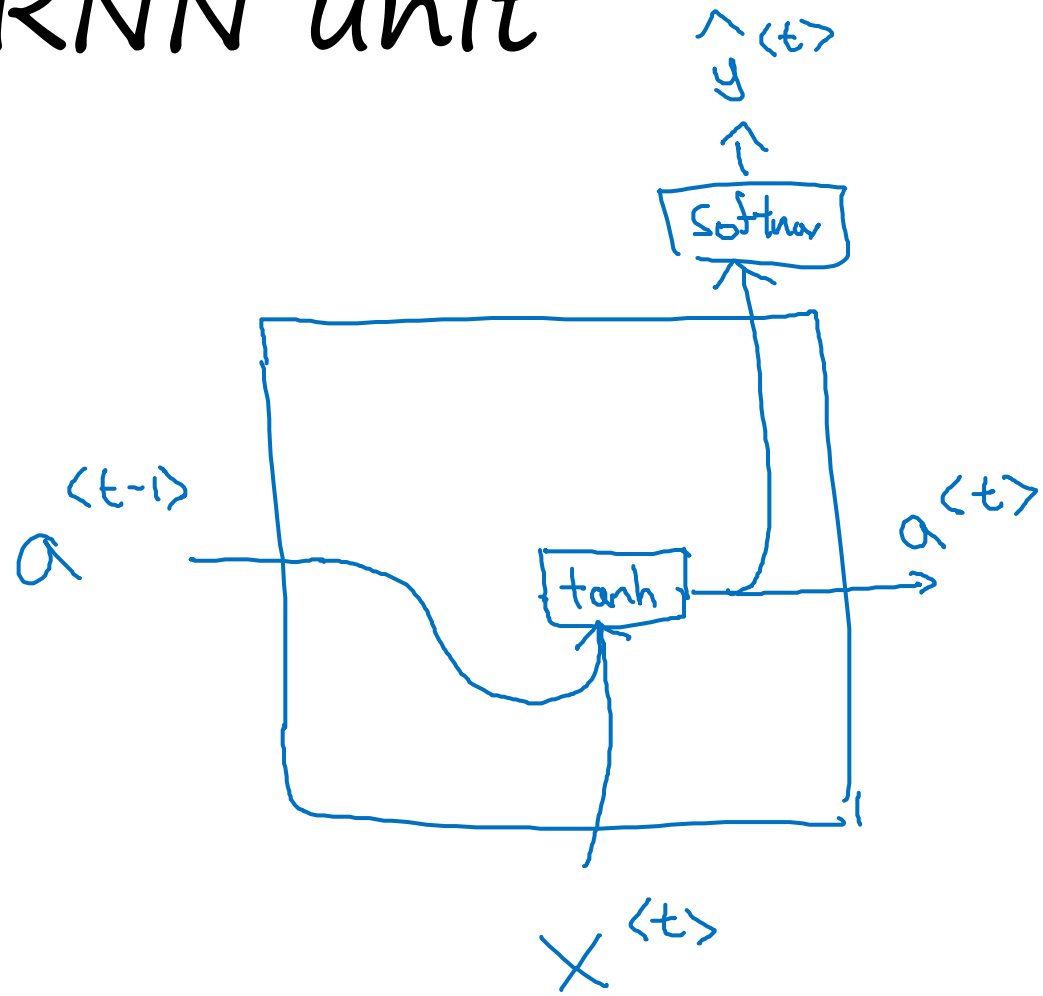
$$a^{<t>} = g(W_a[a^{<t-1>}, x^{<t>}] + b_a)$$

GRU (simplified)

$$C = \text{memory cell}$$

$$C^{\langle t \rangle} = a^{\langle t \rangle}$$

$$\tilde{C}^{\langle t \rangle} = \tanh\left(W_c\left[c^{\langle t-1 \rangle}, x^{\langle t \rangle}\right] + b_c\right)$$

$$\Gamma_u = \sigma\left(W_u\left[c^{\langle t-1 \rangle}, x^{\langle t \rangle}\right] + b_u\right)$$

"update"

$$C^{\langle t \rangle} = \Gamma_u * \tilde{C}^{\langle t \rangle} + (1 - \Gamma_u) * C^{\langle t-1 \rangle}$$

$$\Gamma_u = 1$$

element-wise

$$\Gamma_u = 0.00000l$$

Gate

$y^{\langle t \rangle}$

softmax

$C^{\langle t-1 \rangle} = a^{\langle t-1 \rangle}$

$\tilde{C}^{\langle t \rangle}$

Γ_u

tanh σ

$x^{\langle t \rangle}$

$C^{\langle t \rangle} = a^{\langle t \rangle}$

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

$c^{\langle t \rangle} = 1$ $= 1$
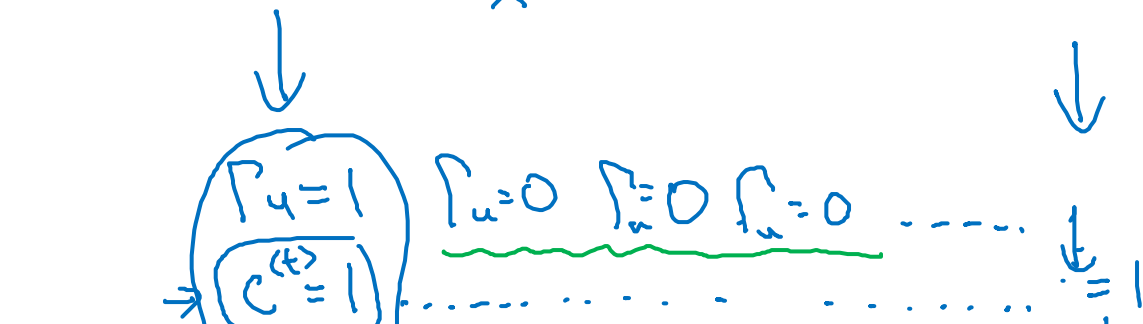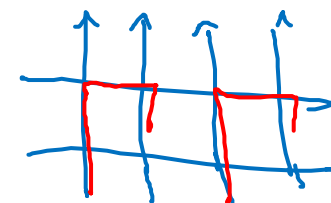
The cat, which already ate ..., was full.

[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

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

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

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

LSTM

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

The cat, which ate already, was full.