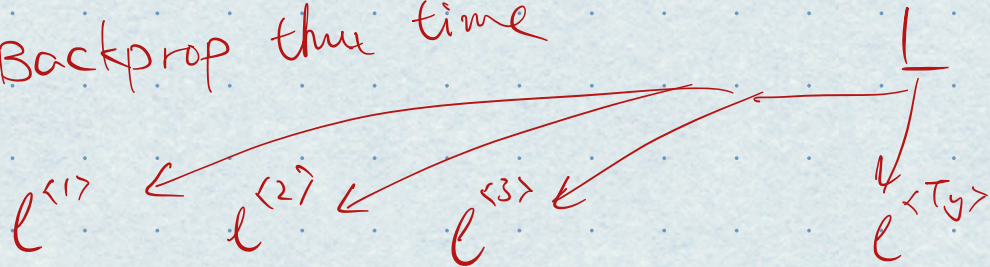


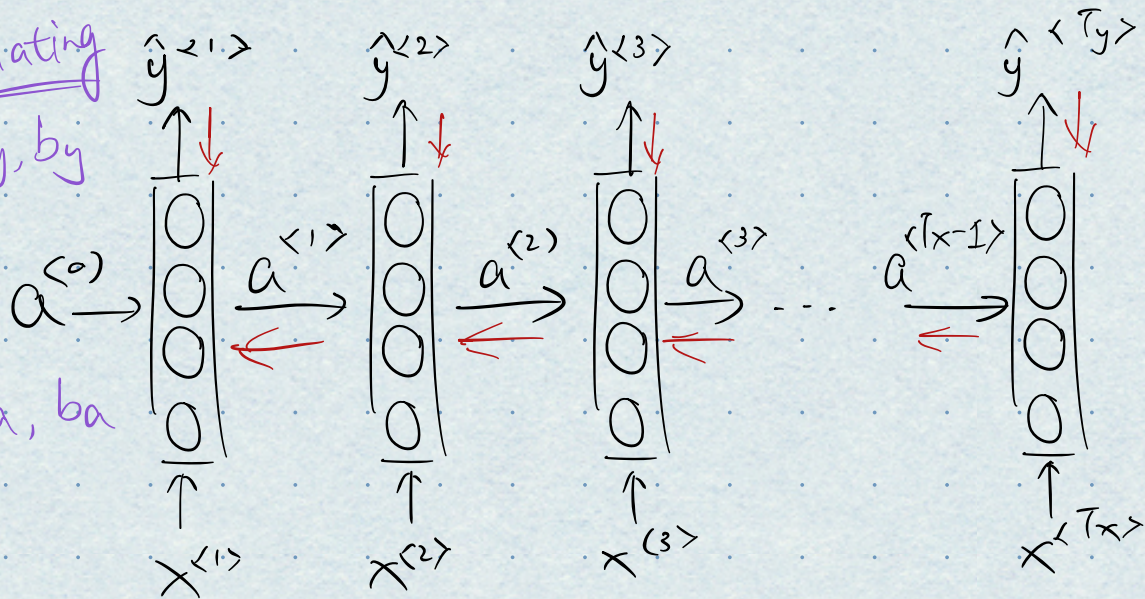
Recurrent neural nets

Backprop thru time



Updating
 W_y, b_y

W_a, b_a

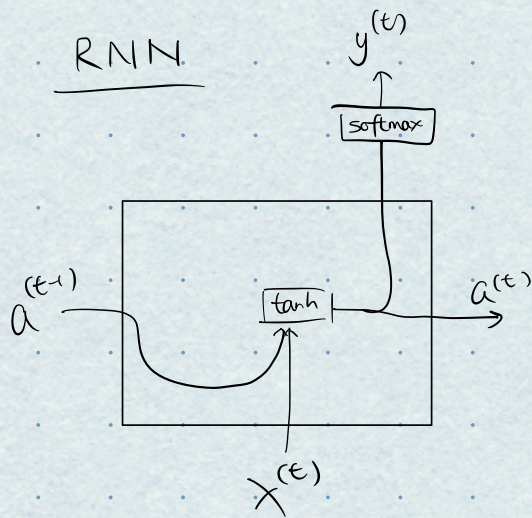


$$a^{<0>} = \vec{0}$$

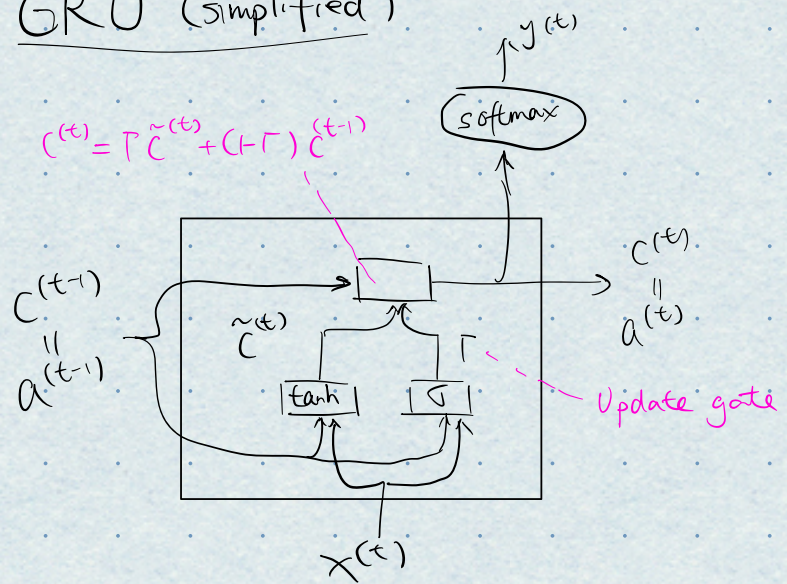
$$a^{<1>} = g_1(W_{aa}a^{<0>} + W_{ax}x^{<1>} + b_a) \leftarrow \tanh$$

$$\hat{y}^{<1>} = g_2(W_{ya}a^{<1>} + b_y) \leftarrow \text{sigmoid}$$

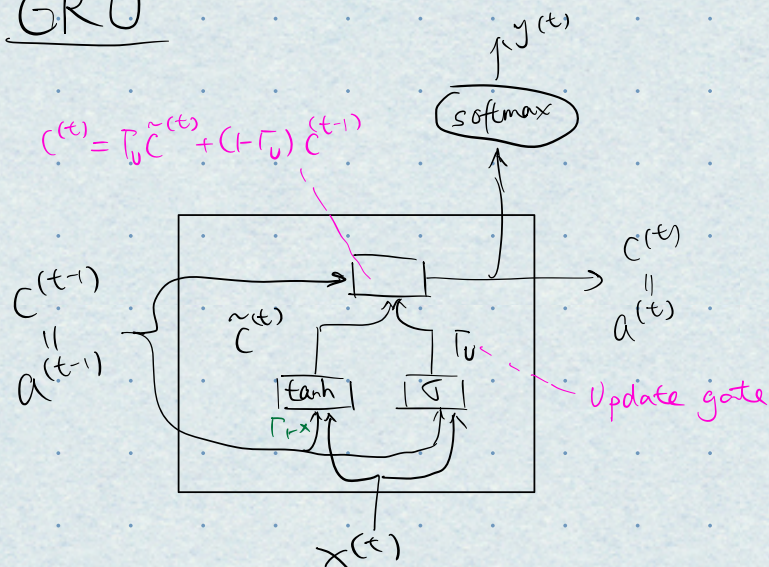
RNN



GRU (simplified)



GRU



$$\Gamma_r = \sigma(W_r [C^{(t-1)}; x^{(t)}])$$

$$\tilde{C}^{(t)} = \tanh(W_c [\Gamma_r C^{(t-1)}; x^{(t)}])$$

GRU

$$\tilde{C}^{(t)} = \tanh(W_c [\Gamma_r C^{(t-1)}; x^{(t)}])$$

$$\Gamma_u = \sigma(W_u [C^{(t-1)}; x^{(t)}])$$

$$\Gamma_r = \sigma(W_r [C^{(t-1)}; x^{(t)}])$$

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

$$a^{(t)} = C^{(t)}$$

LSTM

$$\tilde{C}^{(t)} = \tanh(W_c [a^{(t-1)}; x^{(t)}])$$

$$\Gamma_u = \sigma(W_u [a^{(t-1)}; x^{(t)}])$$

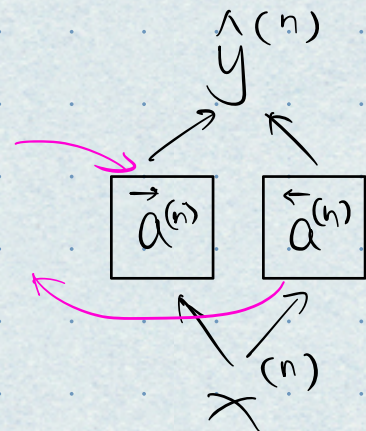
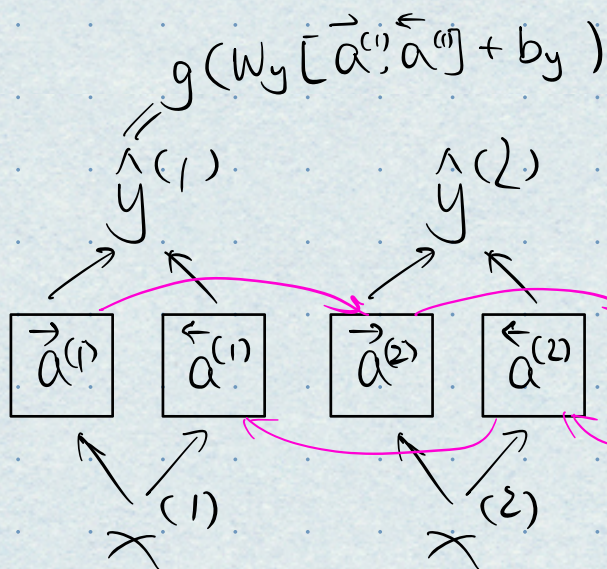
$$\Gamma_f = \sigma(W_f [a^{(t-1)}; x^{(t)}])$$

$$\Gamma_o = \sigma(W_o [a^{(t-1)}; x^{(t)}])$$

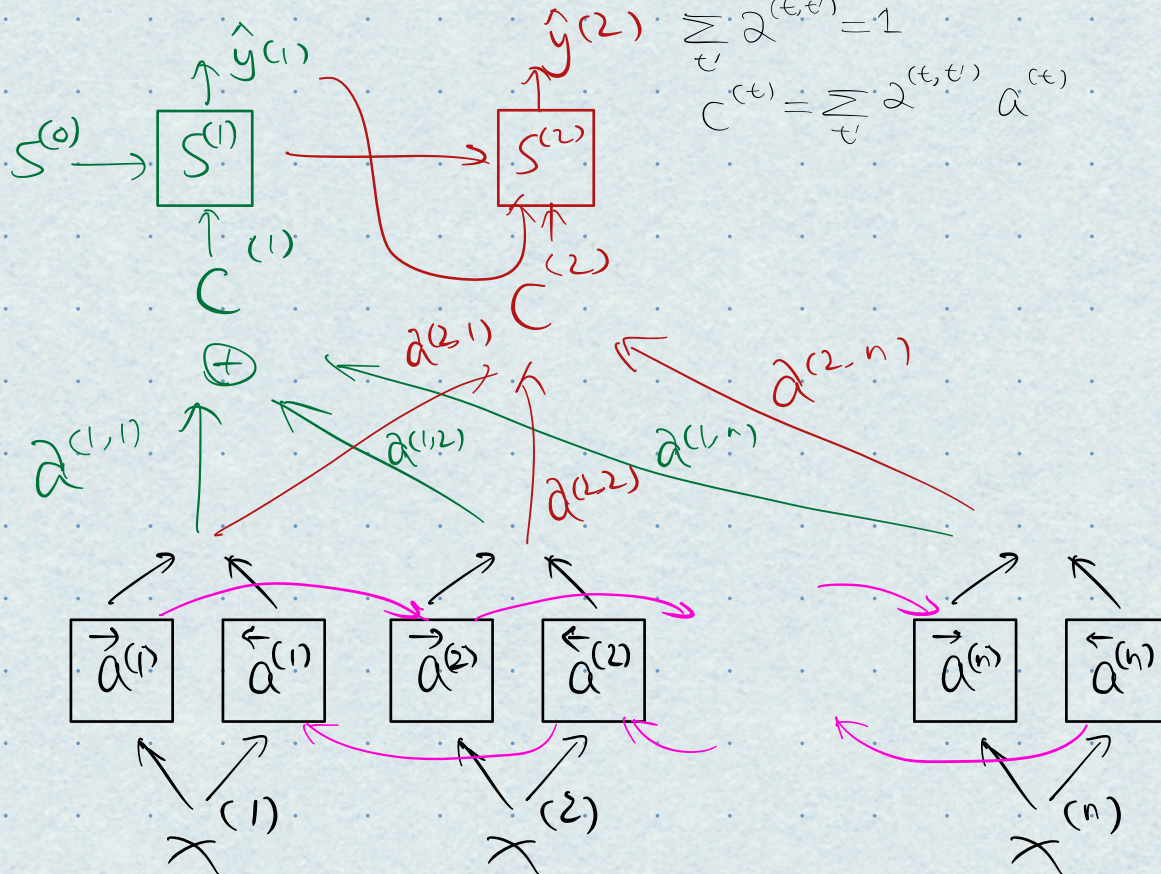
$$C^{(t)} = \Gamma_u \tilde{C}^{(t)} + \Gamma_f C^{(t-1)}$$

$$a^{(t)} = \Gamma_o C^{(t)}$$

Bidirectional RNN



Attention



$\alpha^{(t,t')} = \text{amt. of attention } y^{(t)} \text{ should pay to } a^{(t')}$

$$\sum_{t'} \alpha^{(t,t')} = 1$$

$$C^{(t)} = \sum_{t'} \alpha^{(t,t')} a^{(t')}$$

$$S^{(t-1)}, a^{(t')} \rightarrow \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} \rightarrow e^{(t,t')} \rightarrow \alpha^{(t,t')} = \frac{\exp(e^{(t,t')})}{\sum_{t'} \exp(e^{(t,t')})}$$