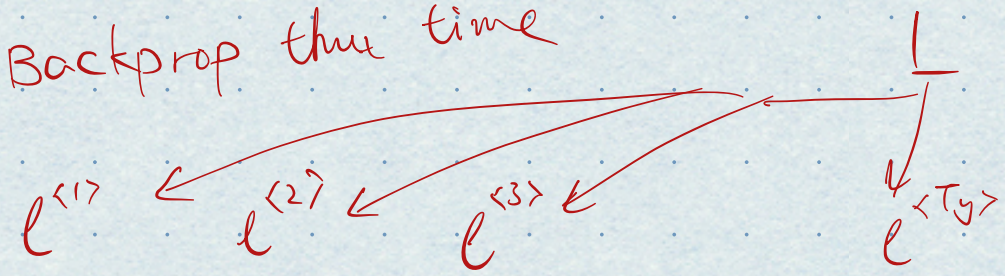
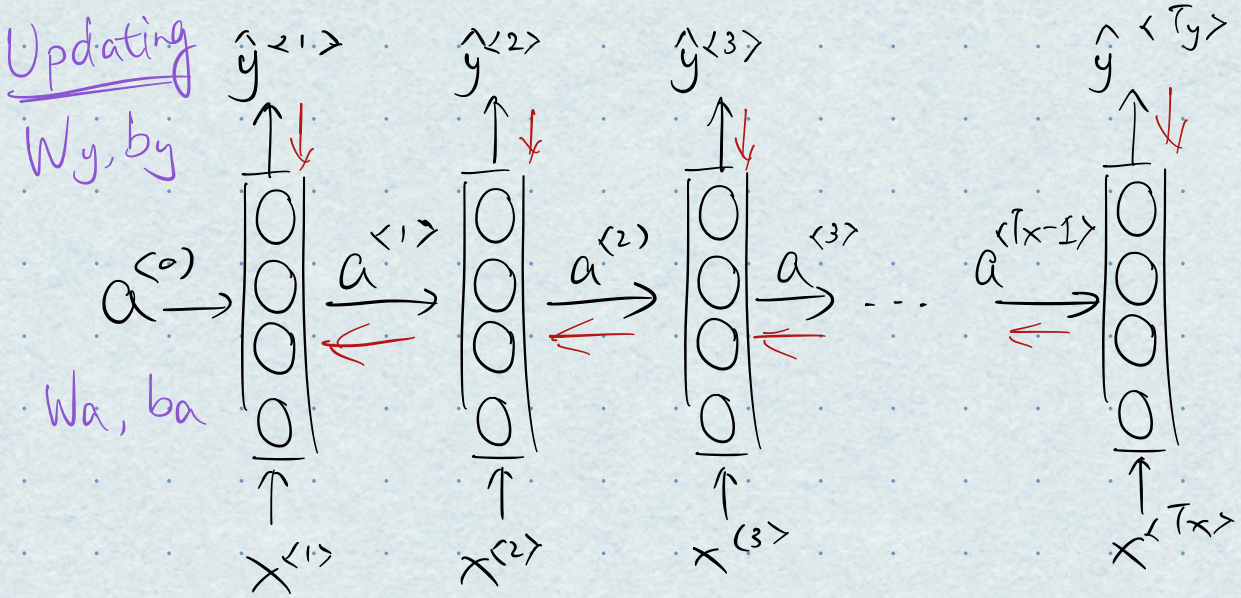
Recurrent neural nets



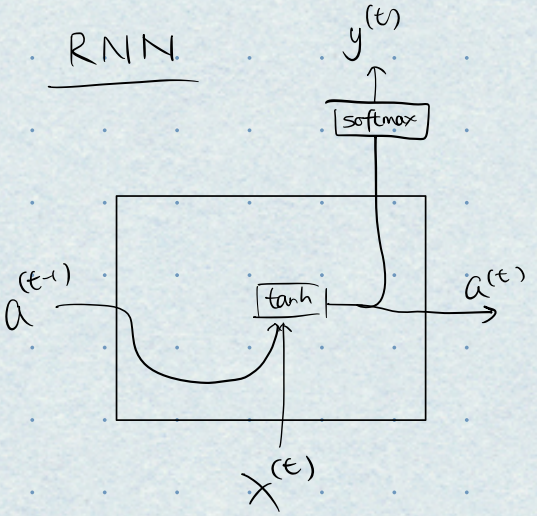


$$a^{\langle 0 \rangle} = \vec{0}$$

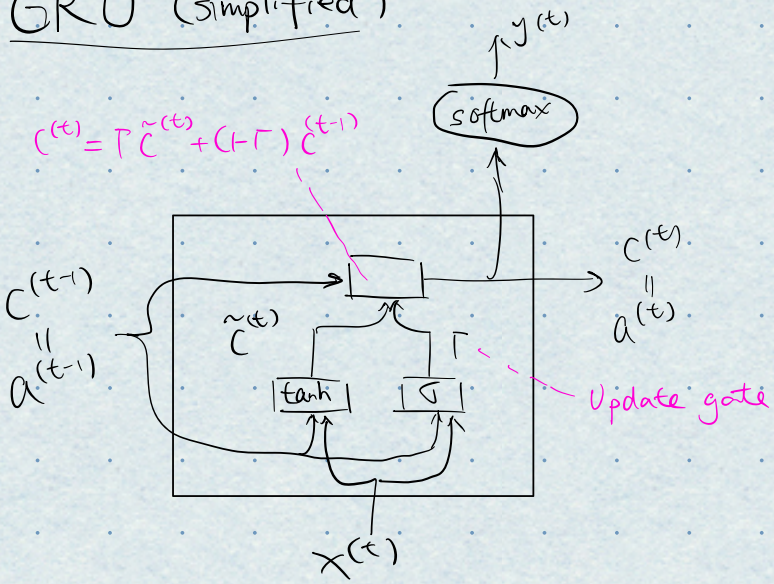
$$a^{\langle 1 \rangle} = g_1(W_{aa} a^{\langle 0 \rangle} + W_{ax} x^{\langle 1 \rangle} + b_a) \leftarrow \tanh$$

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

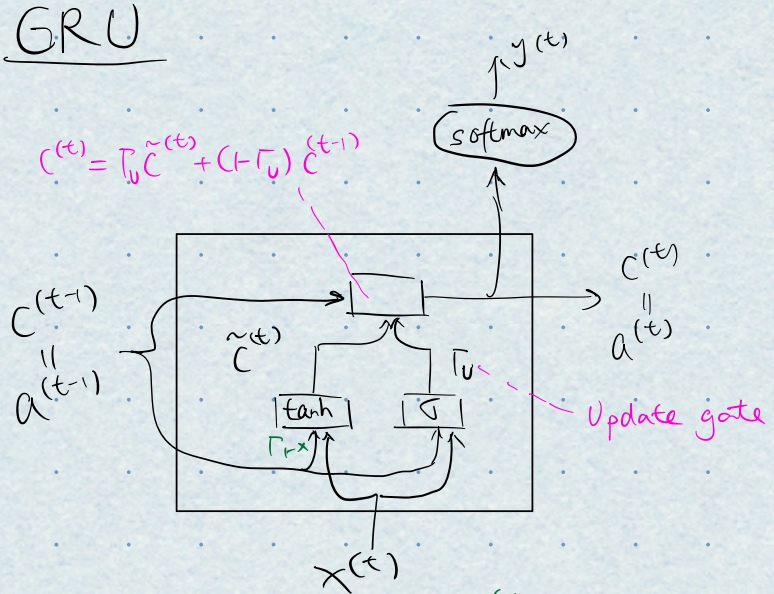
RNN



GRU (simplified)



GRU



$$\Gamma_r = \sigma(W_r [c^{\langle t-1 \rangle}, x^{\langle t \rangle}])$$

$$\tilde{c}^{\langle t \rangle} = \tanh(W_c [\Gamma_r c^{\langle t-1 \rangle}, x^{\langle t \rangle}])$$

GRU

$$\tilde{c}^{\langle t \rangle} = \tanh(W_c[\Gamma_r c^{\langle t-1 \rangle}, x^{\langle t \rangle}])$$

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

$$\Gamma_r = \sigma(W_r[c^{\langle t-1 \rangle}, x^{\langle t \rangle}])$$

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

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

LSTM

$$\tilde{c}^{\langle t \rangle} = \tanh(W_c[a^{\langle t-1 \rangle}, x^{\langle t \rangle}])$$

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

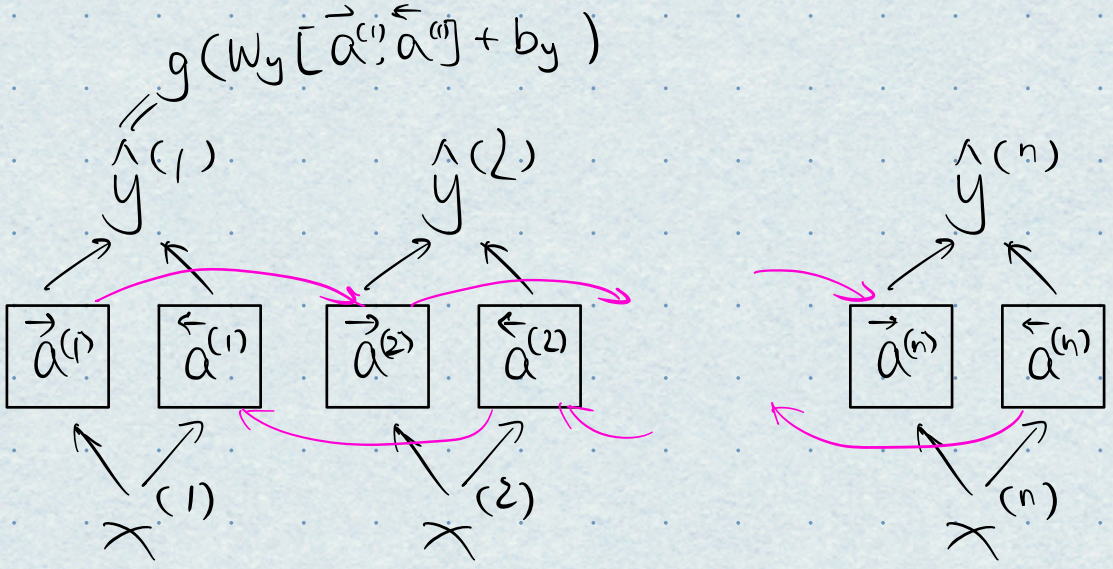
$$\Gamma_f = \sigma(W_f[a^{\langle t-1 \rangle}, x^{\langle t \rangle}])$$

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

$$c^{\langle t \rangle} = \Gamma_u \tilde{c}^{\langle t \rangle} + \Gamma_f c^{\langle t-1 \rangle}$$

$$a^{\langle t \rangle} = \Gamma_o c^{\langle t \rangle}$$

Bidirectional RNN



Attention

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

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

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