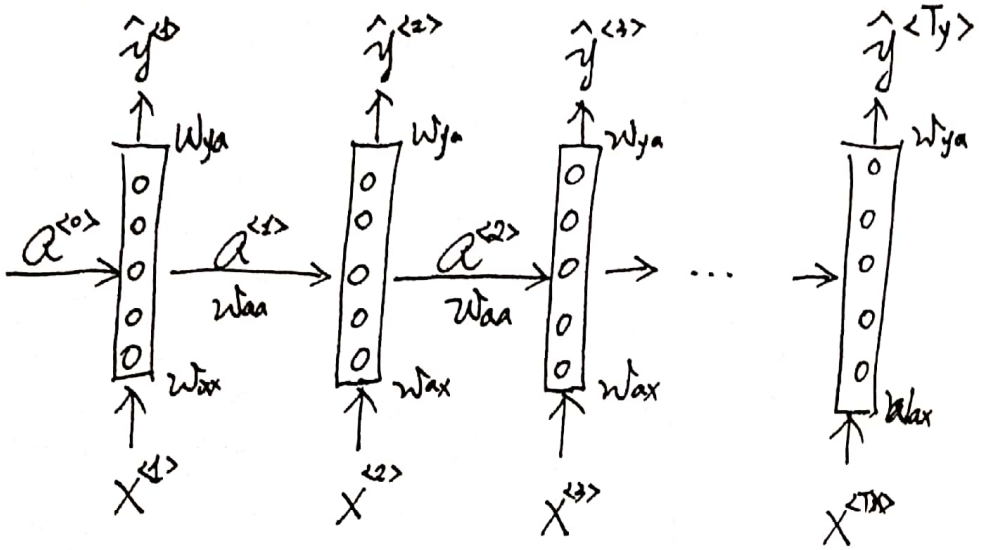
RNN



Ws are shared across from the network

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

$$a^{<1>} = g(W_{aa}\, a^{<0>} + W_{ax}\, x^{<1>} + b_a) \leftarrow g: \tanh/\text{Relu}$$

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

$$\boxed{\begin{aligned} a^{<t>} &= g(W_{aa}\, a^{<t-1>} + W_{ax}\, x^{<t>} + b_a) \\ \hat{y}^{<t>} &= g(W_{ya}\, a^{<t>} + b_y) \end{aligned}}$$

$$\boxed{\begin{aligned} a^{<t>} &= g(W_a [a^{<t-1>}, x^{<t>}] + b_a) \\ \hat{y}^{<t>} &= g(W_y\, a^{<t>} + b_y) \end{aligned}} \qquad W_a = [W_{aa} \vdots W_{ax}]$$

$$[a^{<t-1>}, x^{<t>}] = \begin{bmatrix} a^{<t-1>} \\ \cdots \\ x^{<t>} \end{bmatrix}$$

GRU Gated Recurrent Unit

C : memory cell

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

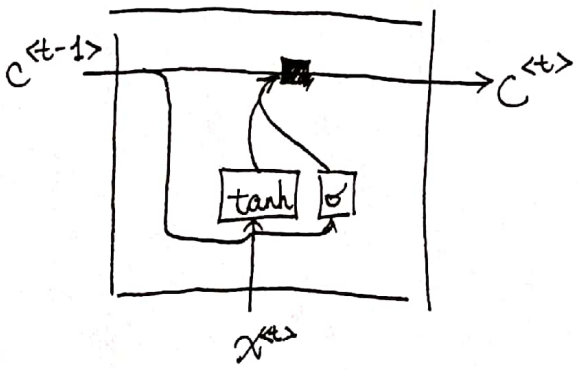
$$\tilde{C}^{<t>} = \tanh\left(W_c\left[\Gamma_r * C^{<t-1>}, x^{<t>}\right] + b_c\right)$$

the candidate value to update $C^{<t-1>}$

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

Γ_u decides how much to update $C^{<t>}$ to the candidate value

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



LSTM Long Short Term Memory

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

$$\Gamma_u = \sigma\left(w_u\left[a^{\langle t-1 \rangle}, x^{\langle t \rangle}\right] + b_u\right) \quad \text{update}$$

$$\Gamma_f = \sigma\left(w_f\left[a^{\langle t-1 \rangle}, x^{\langle t \rangle}\right] + b_f\right) \quad \text{forget}$$

$$\Gamma_o = \sigma\left(w_o\left[a^{\langle t-1 \rangle}, x^{\langle t \rangle}\right] + b_o\right) \quad \text{output}$$

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

$$a^{\langle t \rangle} = \Gamma_o * c^{\langle t \rangle}\left(\Gamma_o * \tanh c^{\langle t \rangle}\right)$$