



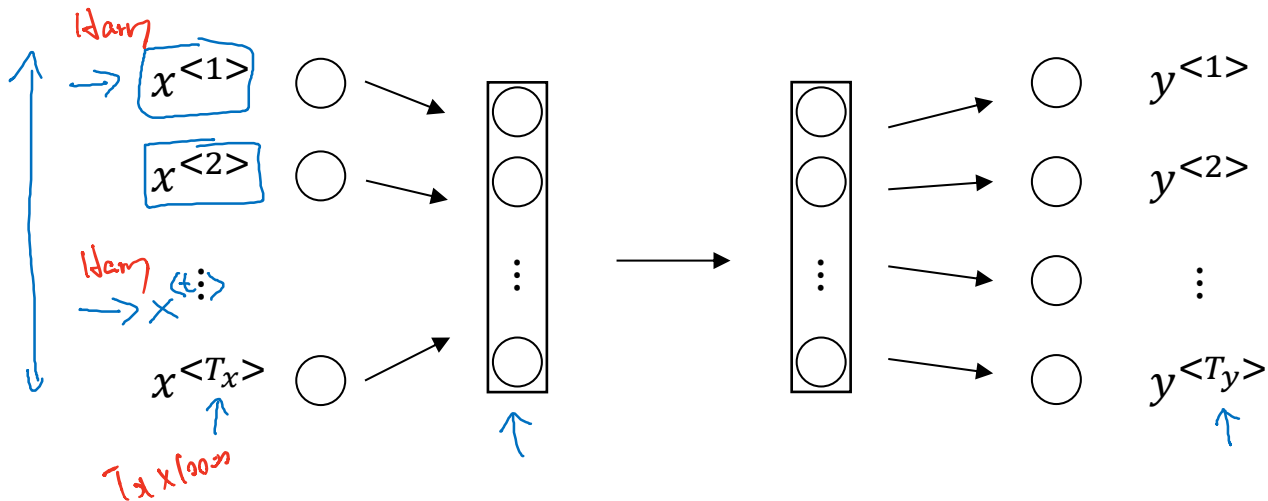
deeplearning.ai

# Recurrent Neural Networks

---

## Recurrent Neural Network Model

# Why not a standard network?

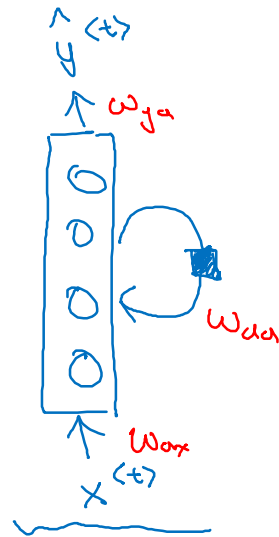
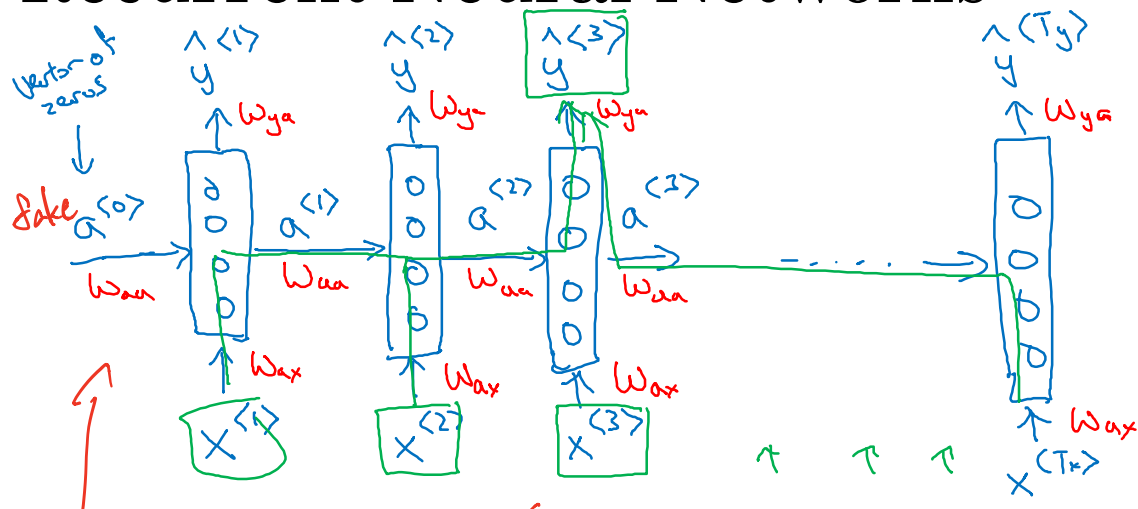


## Problems:

- - Inputs, outputs can be different lengths in different examples. (maybe pad)
- - Doesn't share features learned across different positions of text.

# Recurrent Neural Networks

$$T_x = T_y$$



226  
Bidirectional RNN (BRNN)

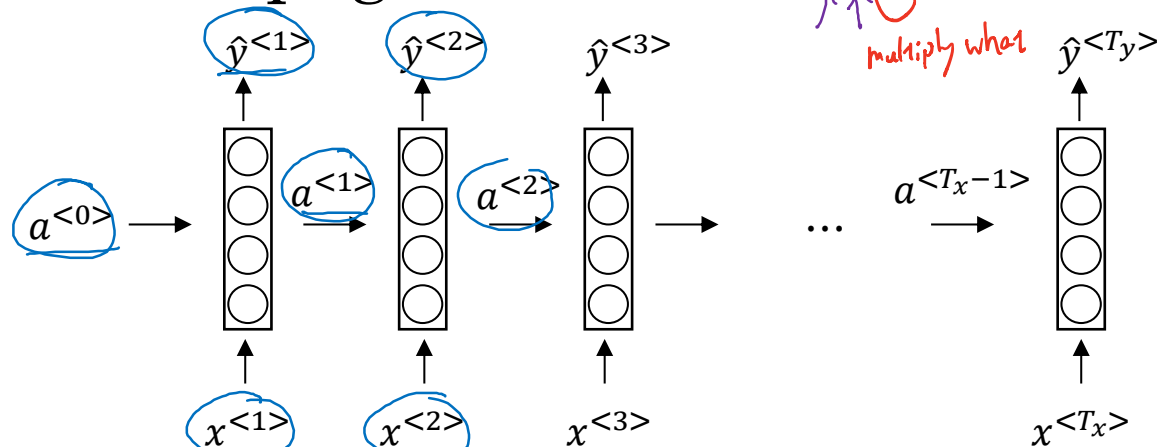
some papers' representation

He said, "Teddy Roosevelt was a great President."

He said, "Teddy bears are on sale!"

complete wheel

# Forward Propagation



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

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

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

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

# Simplified RNN notation

$$a^{<t>} = g(W_{aa}a^{<t-1>} + W_{ax}x^{<t>} + b_a)$$

$\uparrow$  (100, 100)      100       $\uparrow$  (100, 10,000)      10,000

simplify

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

$$\begin{bmatrix} W_{aa} & W_{ax} \end{bmatrix} = W_a$$

$\uparrow$  100       $\leftarrow$  100       $\leftarrow$  10,000       $\downarrow$  (100, 10,000)

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

$\uparrow$  100       $\uparrow$  10,000       $\uparrow$  10100

$$\begin{bmatrix} W_{aa} & W_{ax} \end{bmatrix} \begin{bmatrix} a^{<t-1>} \\ x^{<t>} \end{bmatrix} = W_{aa}a^{<t-1>} + W_{ax}x^{<t>}$$

$$\hat{y}^{<t>} = g(W_{ya}a^{<t>} + b_y)$$

simplify

$$\hat{y}^{<t>} = g(W_y a^{<t>} + b_y)$$

$\uparrow$        $\uparrow$        $\uparrow$