



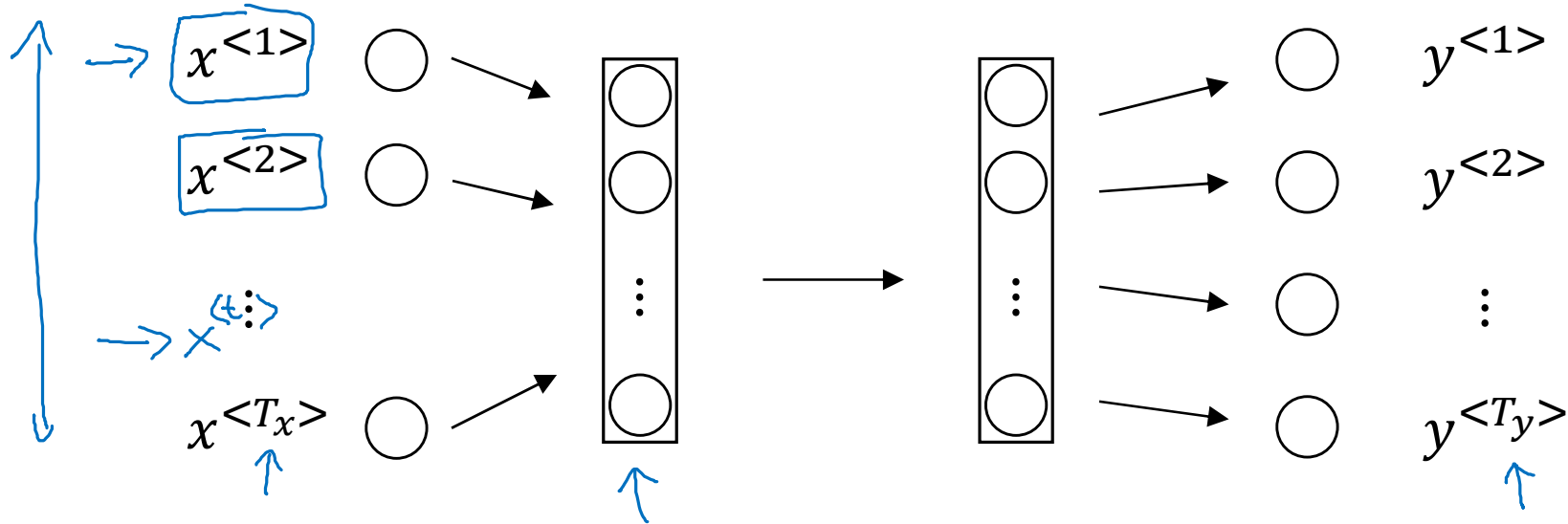
deeplearning.ai

# Recurrent Neural Networks

---

## Recurrent Neural Network Model

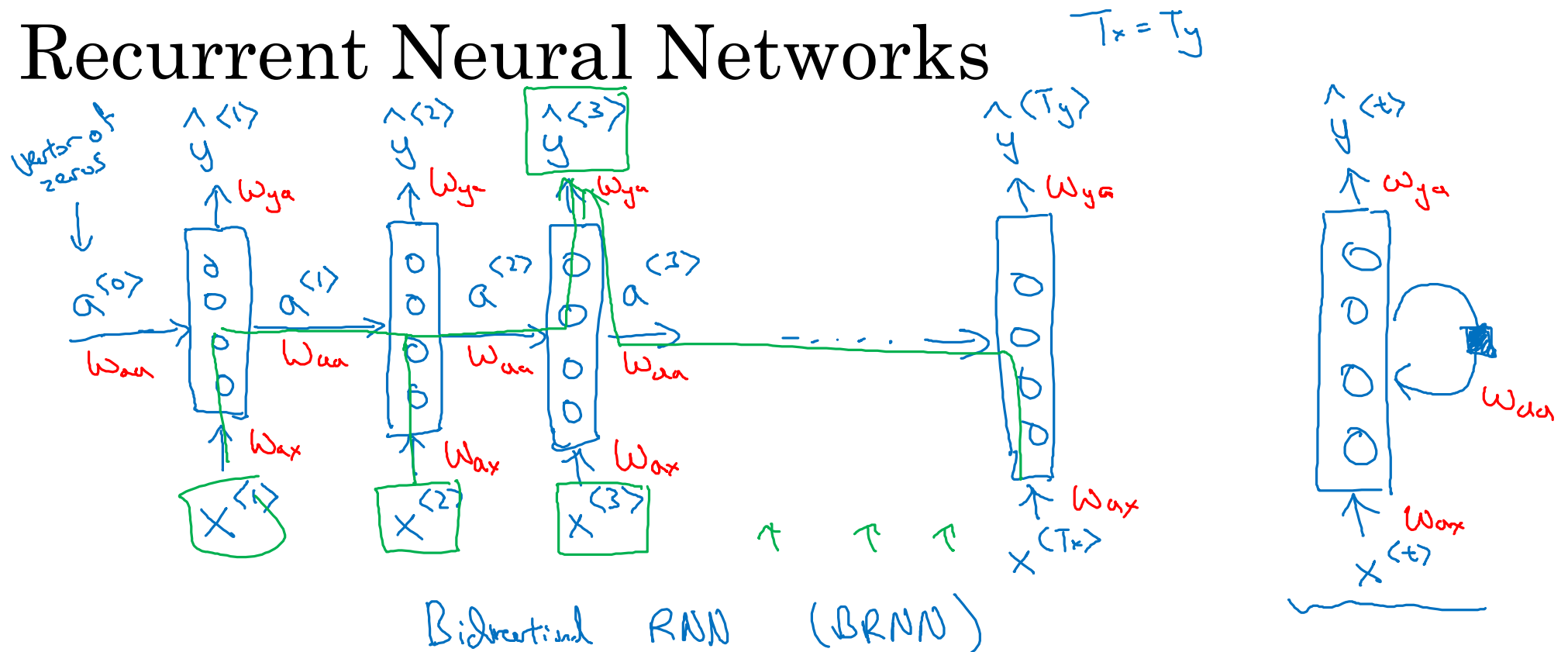
# Why not a standard network?



## Problems:

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

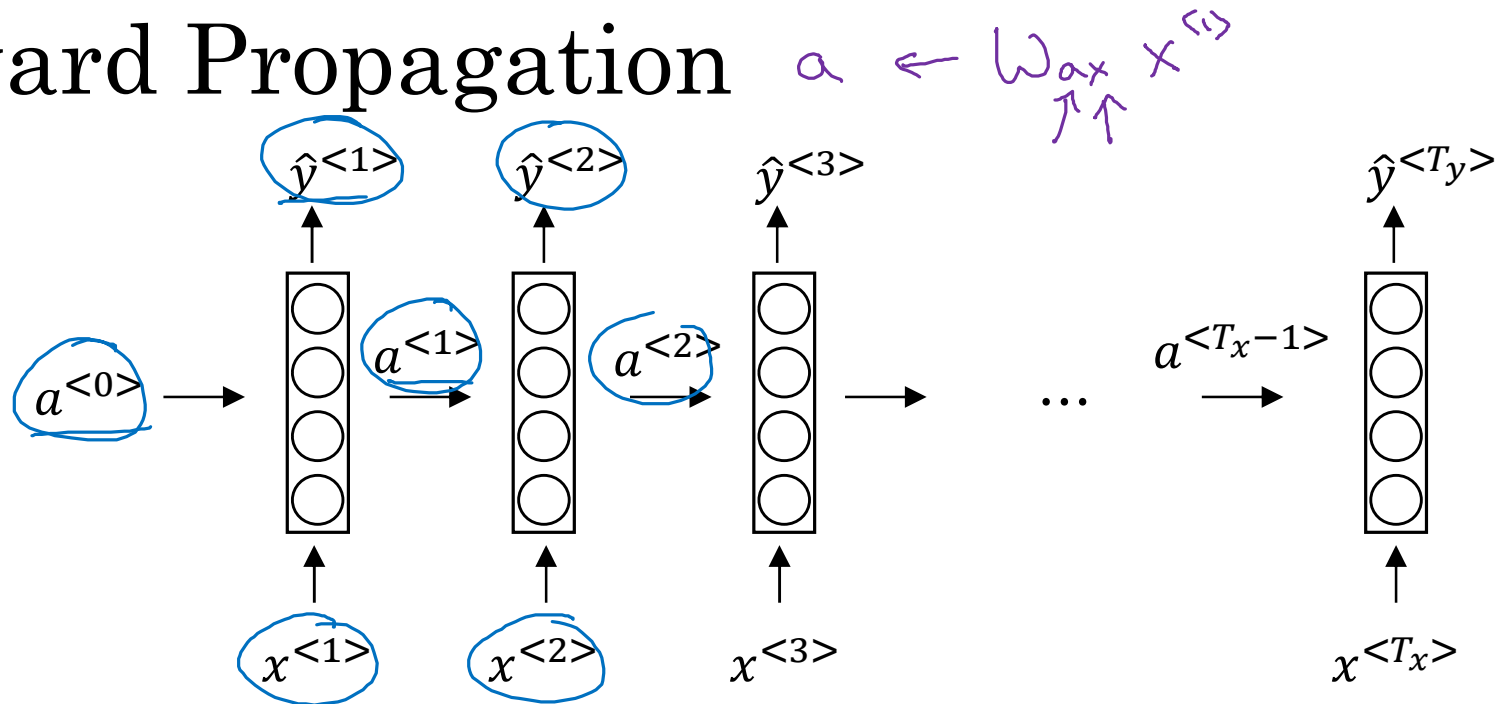
# Recurrent Neural Networks



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

He said, "Teddy bears are on sale!"

# Forward Propagation



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

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

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

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

# Simplified RNN notation

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

Annotations:  $W_{aa}$  is a  $100 \times 100$  matrix,  $W_{ax}$  is a  $100 \times 10,000$  matrix.  $a^{<t-1>}$  has dimension 100,  $x^{<t>}$  has dimension 10,000.

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

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

Annotations:  $W_y$  is a  $100 \times 100$  matrix.  $\hat{y}^{<t>}$  has dimension 100.

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

Annotations:  $W_a$  is a  $100 \times 10100$  matrix. The input vector  $[a^{<t-1>}, x^{<t>}]$  has dimension 10100.

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

Annotations:  $W_{aa}$  is  $100 \times 100$ ,  $W_{ax}$  is  $100 \times 10,000$ . The combined matrix  $W_a$  is  $100 \times 10100$ .

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

Annotations:  $a^{<t-1>}$  has dimension 100,  $x^{<t>}$  has dimension 10,000. The combined vector has dimension 10100.

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