

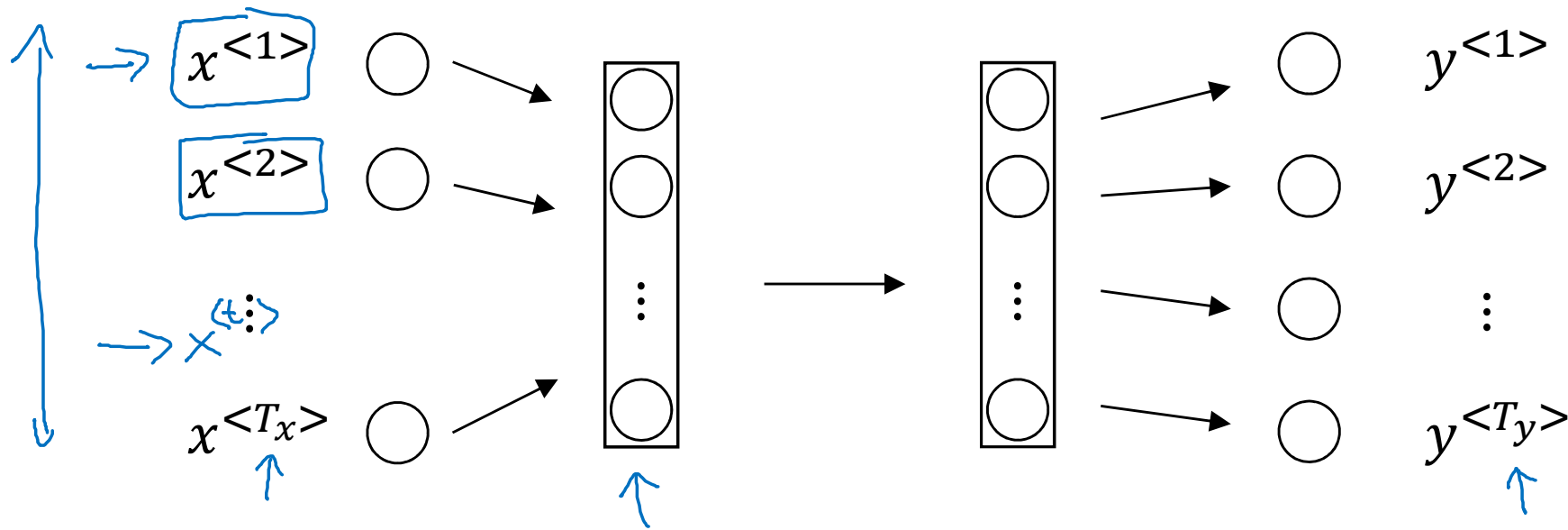


deeplearning.ai

Recurrent Neural Networks

Recurrent Neural Network Model

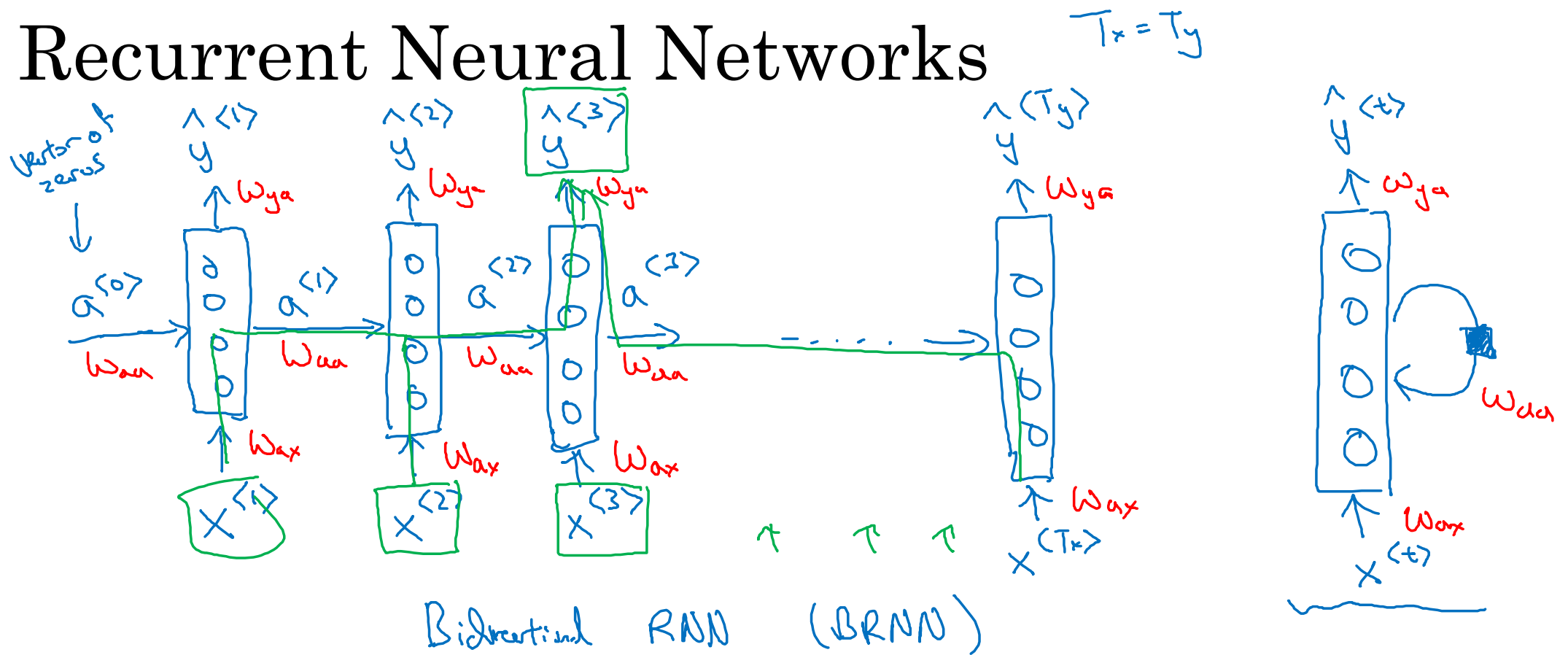
Why not a standard network?



Problems:

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

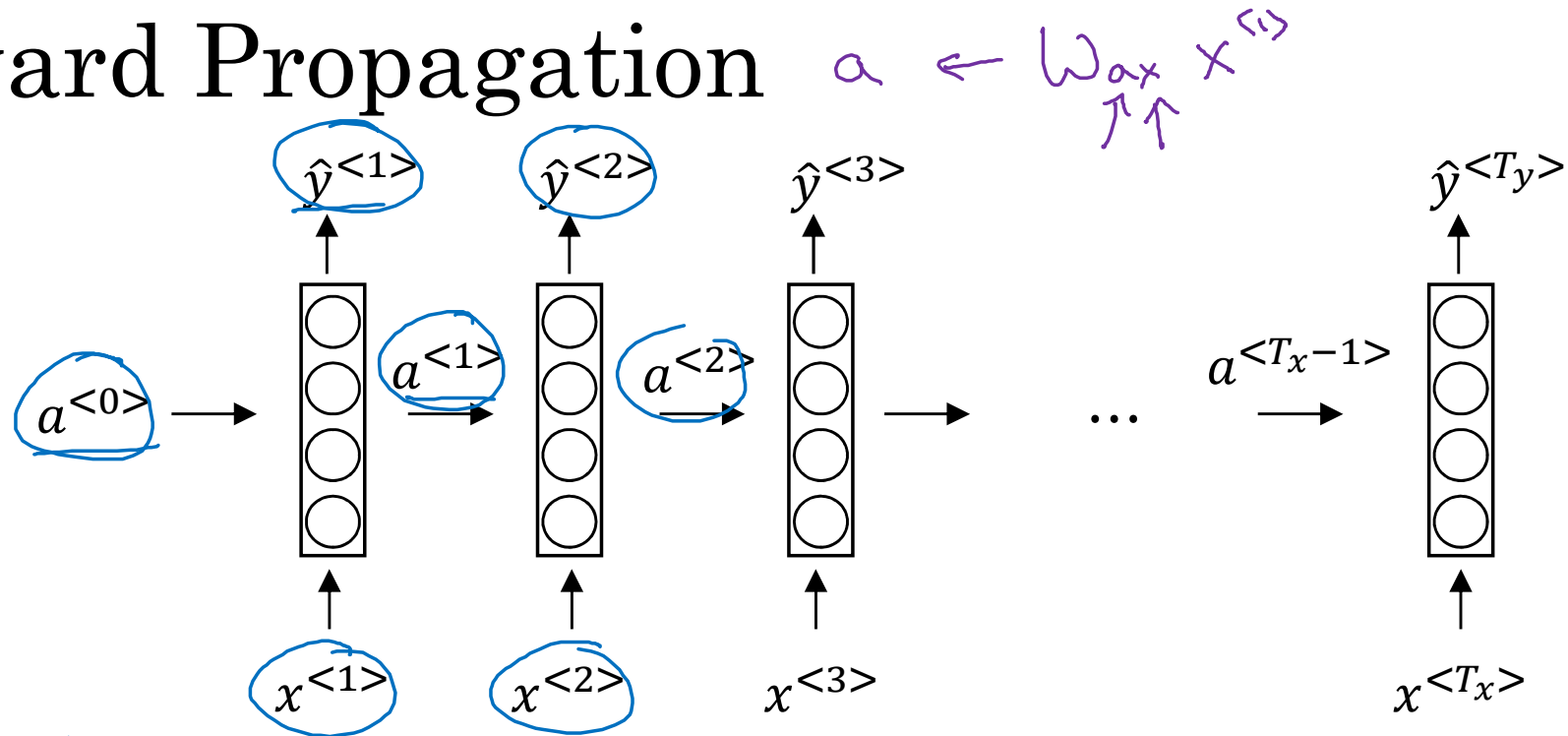
Recurrent Neural Networks



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

He said, "Teddy bears are on sale!"

Forward Propagation



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

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

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

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

Simplified RNN notation

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

Annotations: W_{aa} is a 100×100 matrix, W_{ax} is a $100 \times 10,000$ matrix. A green arrow indicates the sequence from $a^{<t-1>}$ to $a^{<t>}$.

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

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

Annotations: W_y is a 1×100 matrix. A blue arrow points from $\hat{y}^{<t>}$ to $y^{<t>}$.

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

Annotations: W_a is a 100×10100 matrix. The input vector $[a^{<t-1>}, x^{<t>}]$ is shown in a purple box.

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

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

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

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

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