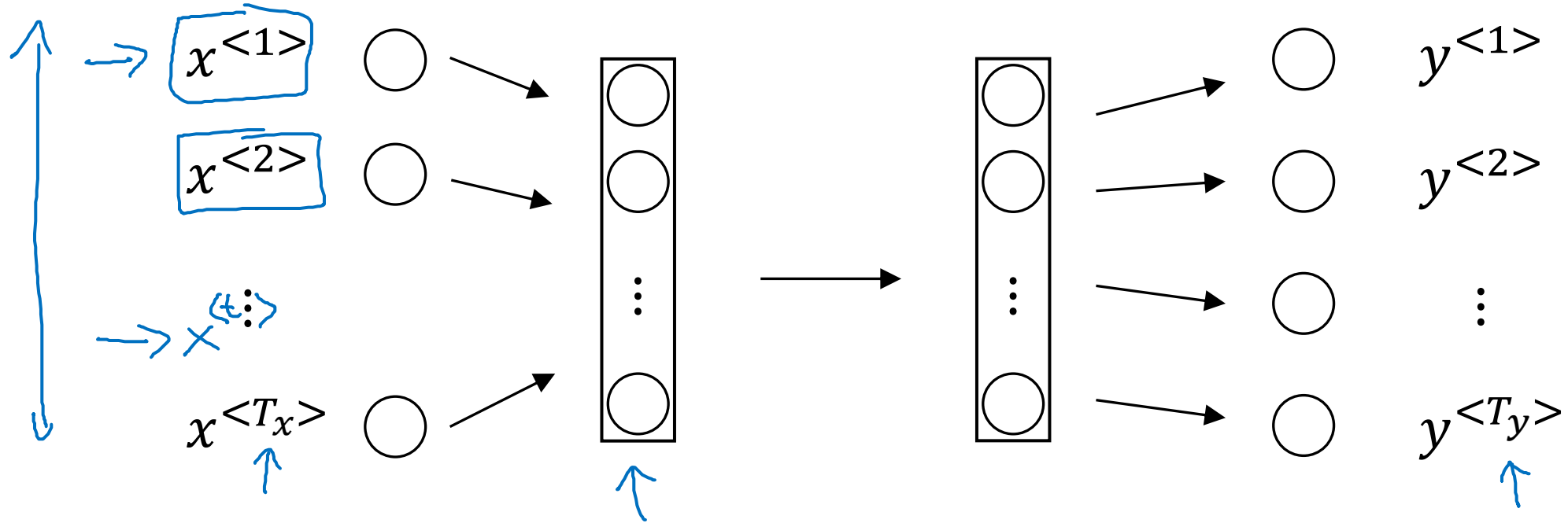deeplearning.ai

Recurrent Neural
Networks

Recurrent Neural
Network Model
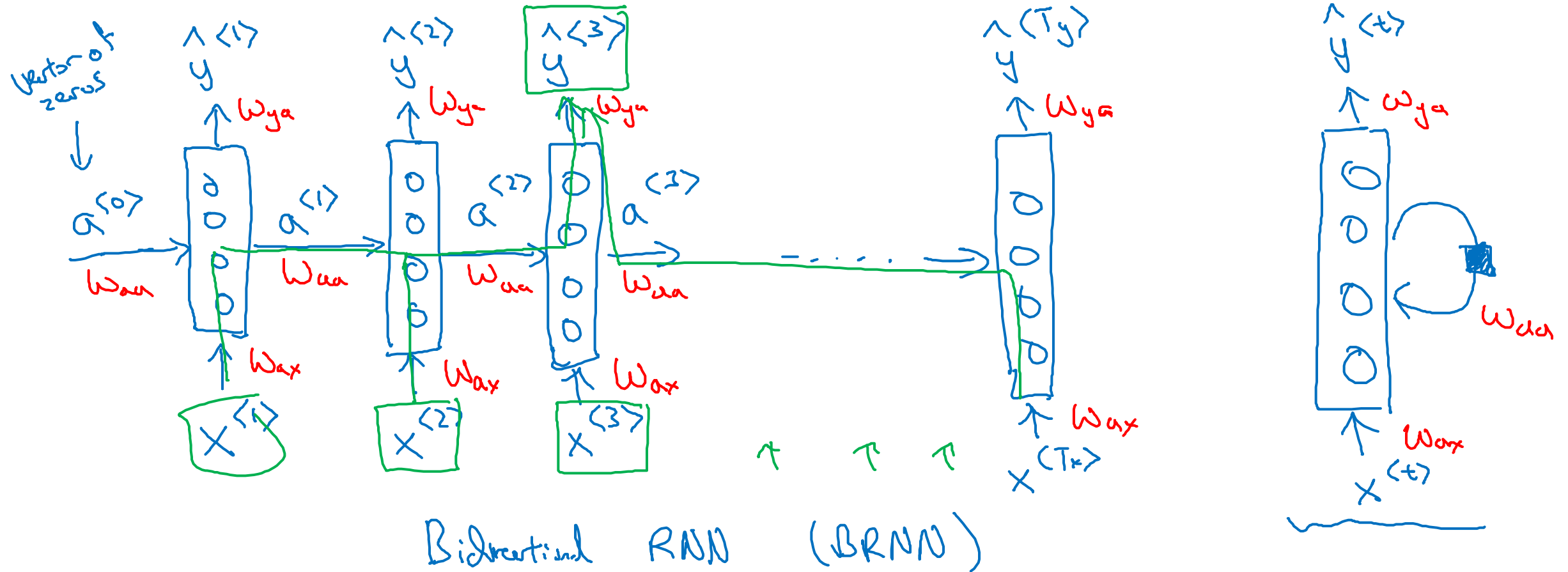
Why not a standard network?

Problems:

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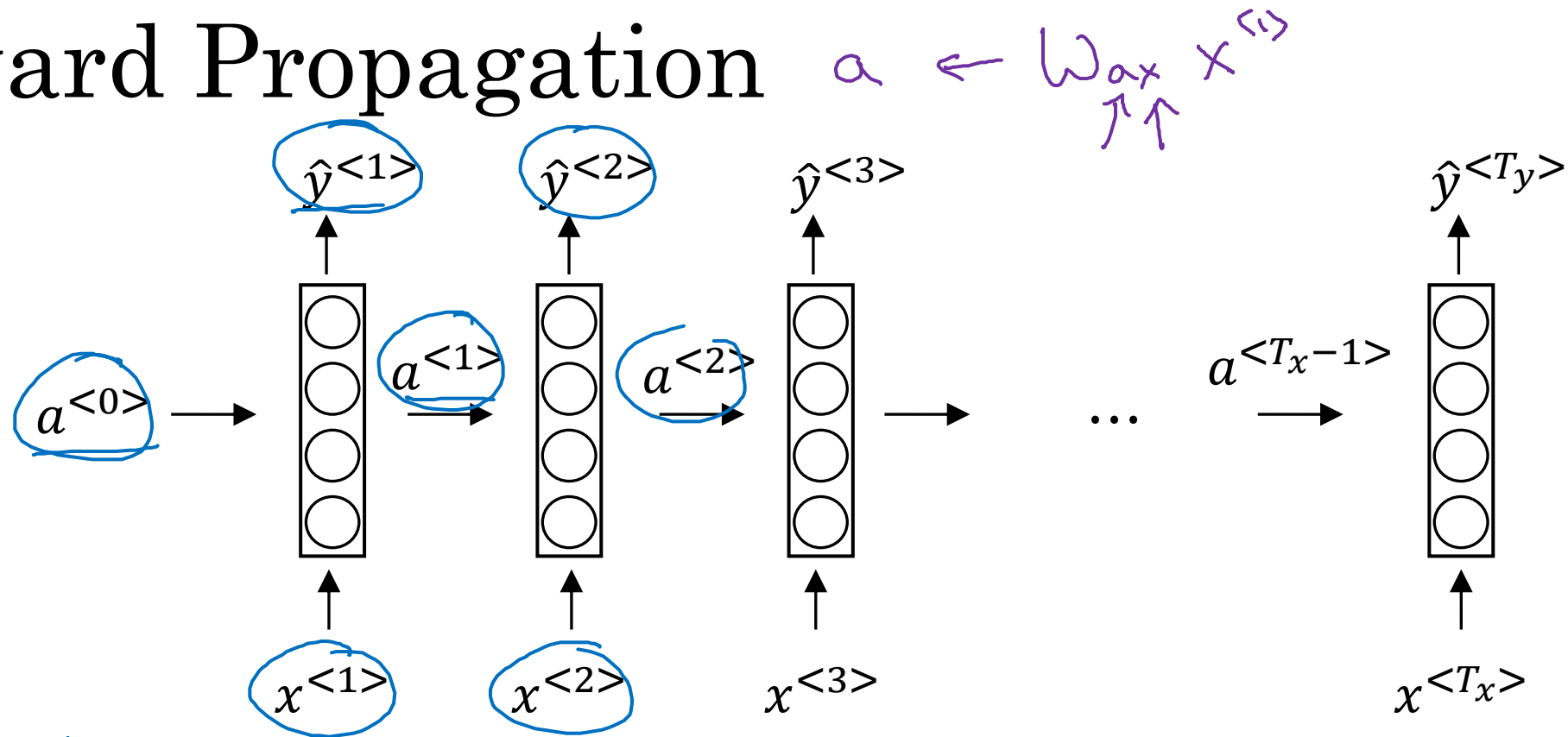- Doesn't share features learned across different positions of text.

Recurrent Neural Networks

$T_x = T_y$

Bidirectional RNN (BRNN)

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

He said, "Teddy bears are on sale!"

Andrew Ng

Forward Propagation

$$a \leftarrow W_{ax} x^{(1)}$$

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

$$a^{(1)} = g_1\left(W_{aa} a^{<0>} + W_{ax} x^{(1)} + b_a\right) \leftarrow \tanh / ReLu$$

$$\hat{y}^{(1)} = g_2\left(W_{ya} a^{(1)} + b_y\right) \leftarrow Sigmoid$$

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

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

Andrew Ng

Simplified RNN notation

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

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

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

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

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

$$100 \left[W_{aa} ; W_{ax} \right] = W_a$$

$$100 \qquad 10000$$

$$(100, 10100)$$

$$[a^{<t-1>}, x^{<t>}] = \begin{bmatrix} a^{<t-1>} \\ x^{<t>} \end{bmatrix} \begin{matrix} 100 \\ 10000 \end{matrix} \quad 10100$$

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