Recurrent Neural Networks
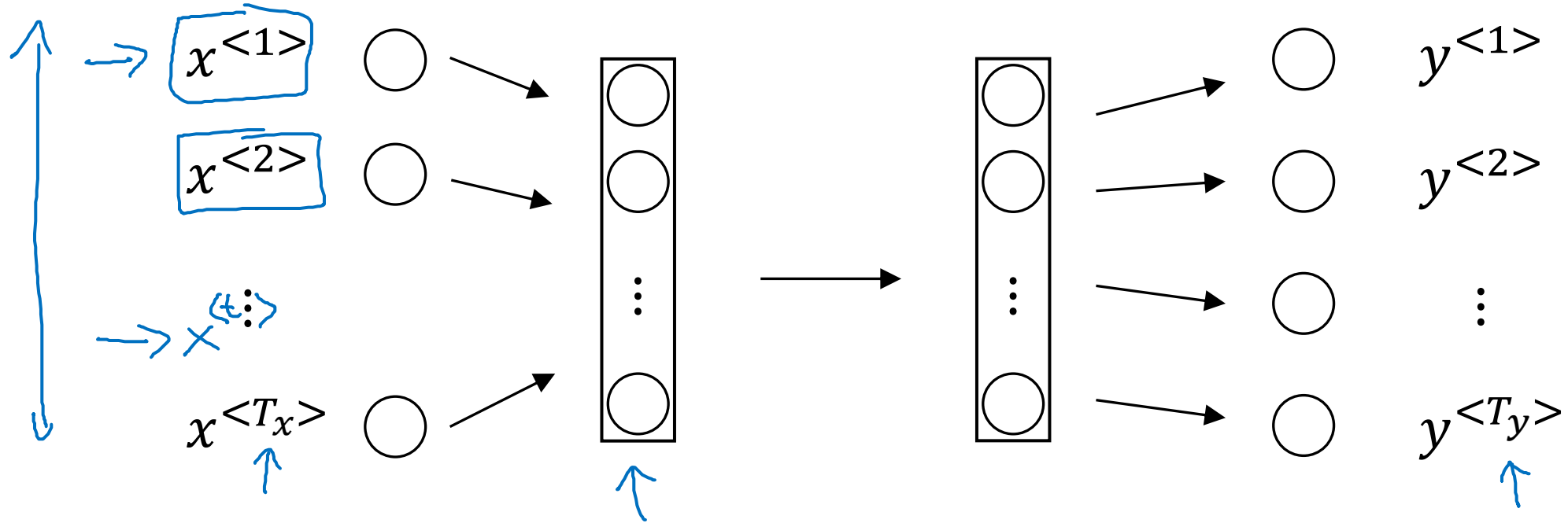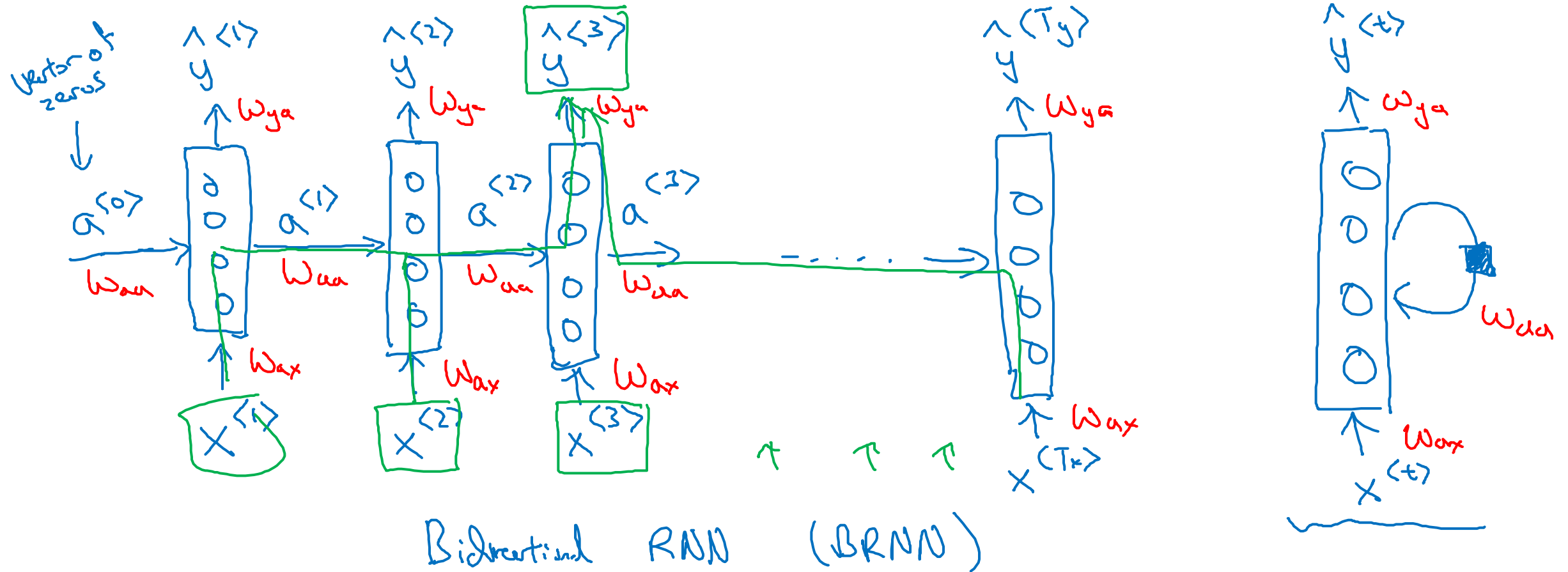
deeplearning.ai

Recurrent Neural Network Model

Why not a standard network?

Problems:

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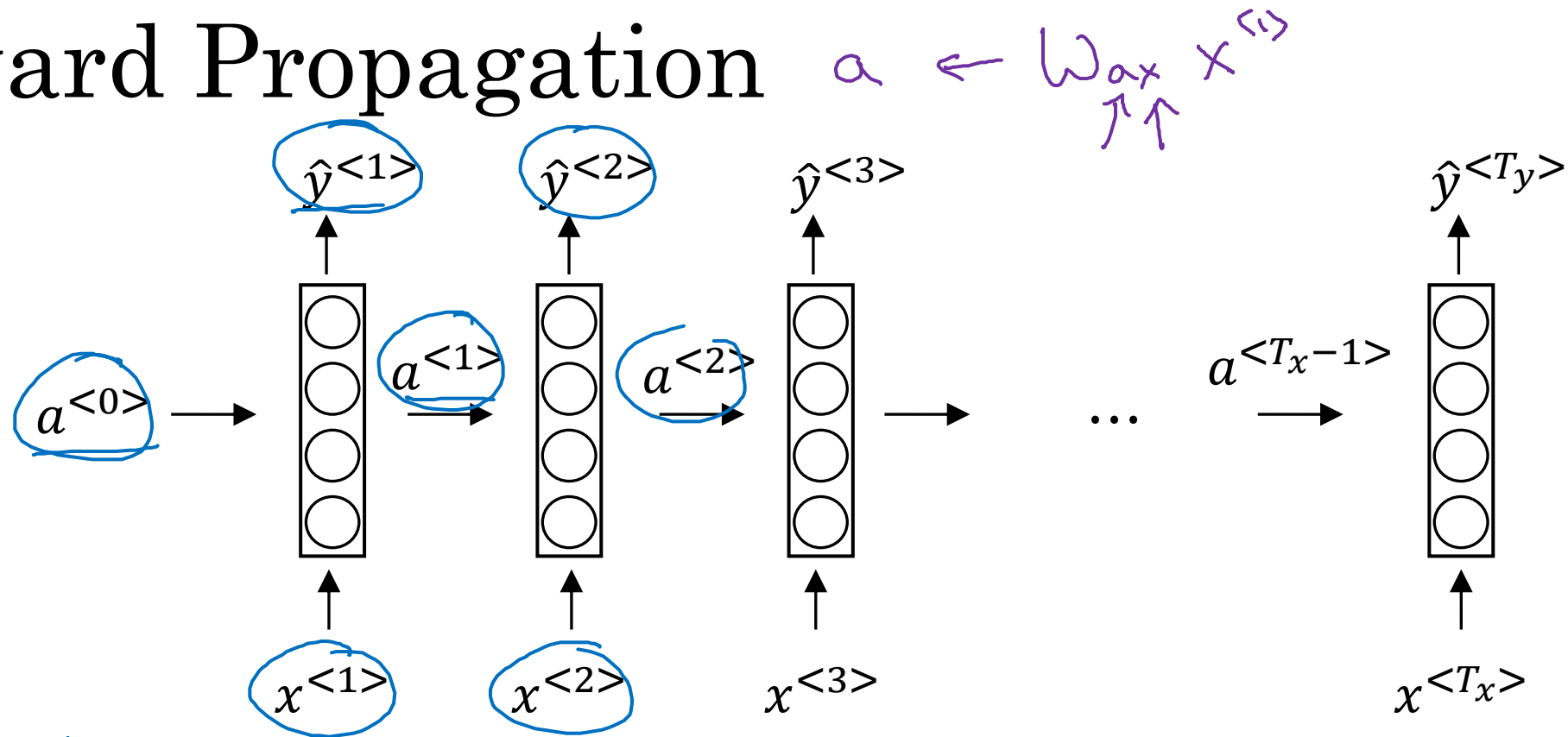- Doesn't share features learned across different positions of text.

Recurrent Neural Networks

Bidirectional RNN (BRNN)

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

He said, "Teddy bears are on sale!"

Forward Propagation

$$a \leftarrow W_{ax} x^{<1>}$$

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

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

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

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

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

Andrew Ng

Simplified RNN notation

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

100 $(100,100)$ $10,000$ $(100,10,000)$

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

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

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

$$100 \begin{bmatrix} W_{aa} & ; & W_{ax} \end{bmatrix} = W_a$$

100 $10\,000$

$(100, 10100)$

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

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

Andrew Ng