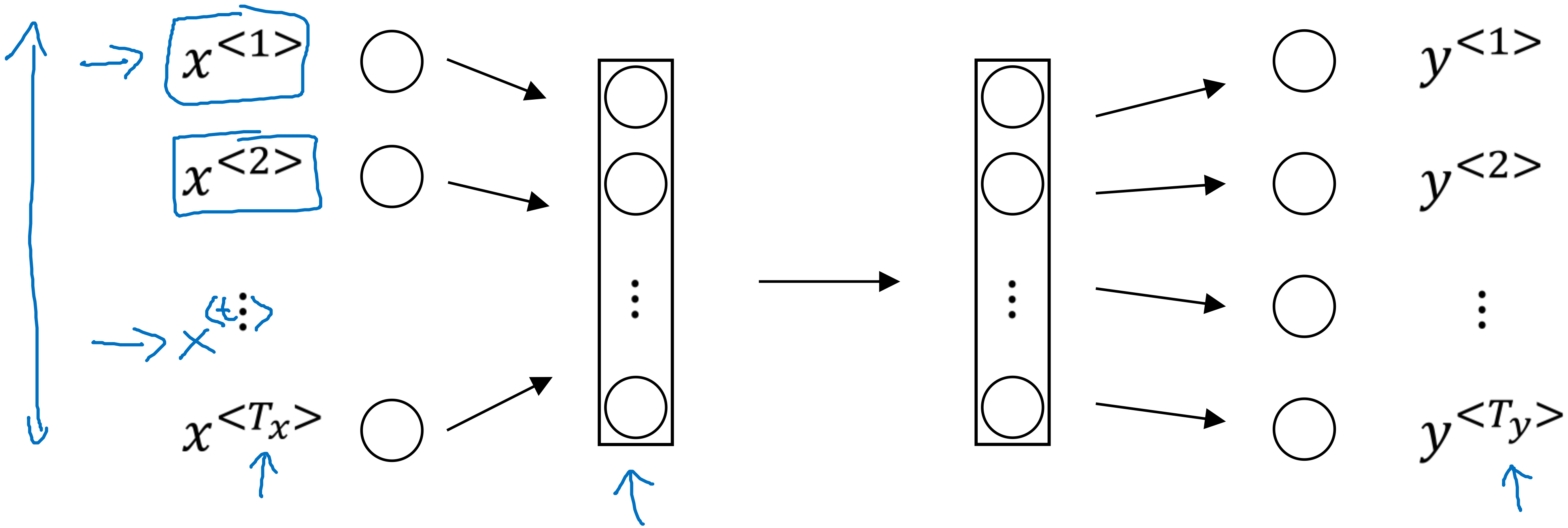
deeplearning.ai

Recurrent Neural Networks

Recurrent Neural Network Model

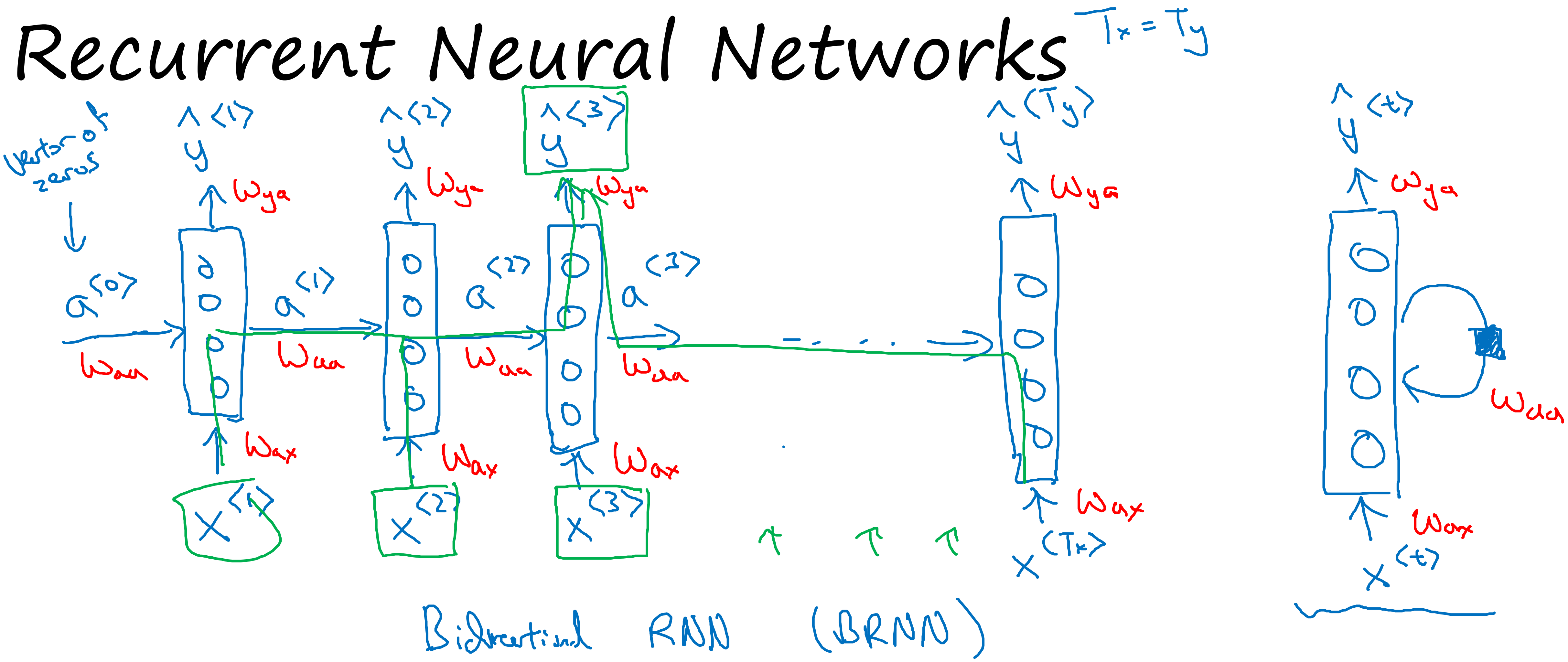
Why not a standard network?



Problems:

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

Recurrent Neural Networks



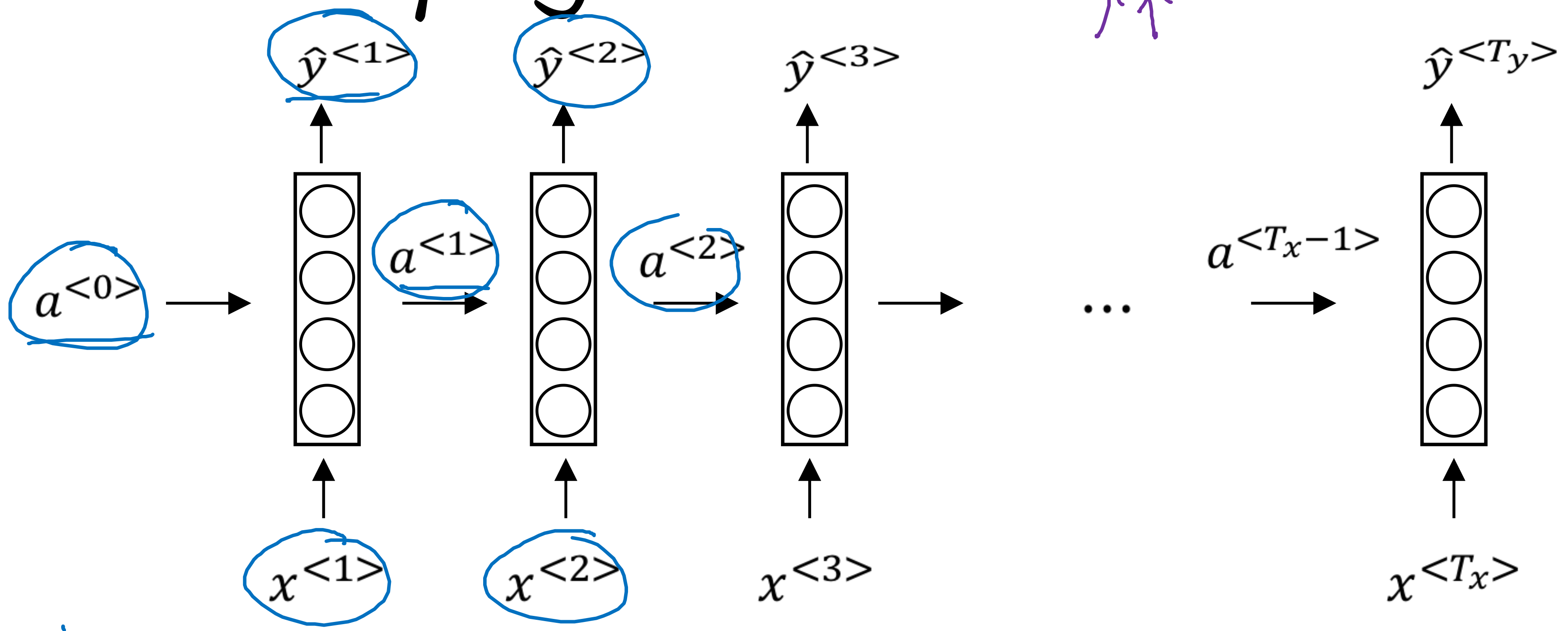
He said, "Teddy Roosevelt was a great President."

He said, "Teddy bears are on sale!"

Andrew Ng

Forward Propagation

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



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

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

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

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

Simplified RNN notation

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

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

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

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

$$a^{\langle t\rangle} = g(W_a[a^{\langle t-1\rangle}, x^{\langle t\rangle}] + b_a)$$

$$100 \left[\underbrace{W_{aa}}_{100} \vdots \underbrace{W_{ax}}_{10\,000} \right] = W_a$$

(100, 10100)

$$[a^{\langle t-1\rangle}, x^{\langle t\rangle}] = \begin{bmatrix} a^{\langle t-1\rangle} \\ \hline x^{\langle t\rangle} \end{bmatrix} \begin{array}{l} 100 \\ 10000 \end{array} \quad 10100$$

$$[W_{aa} \vdots W_{ax}] \begin{bmatrix} a^{\langle t-1\rangle} \\ x^{\langle t\rangle} \end{bmatrix} = W_{aa}a^{\langle t-1\rangle} + W_{ax}x^{\langle t\rangle}$$

Andrew Ng