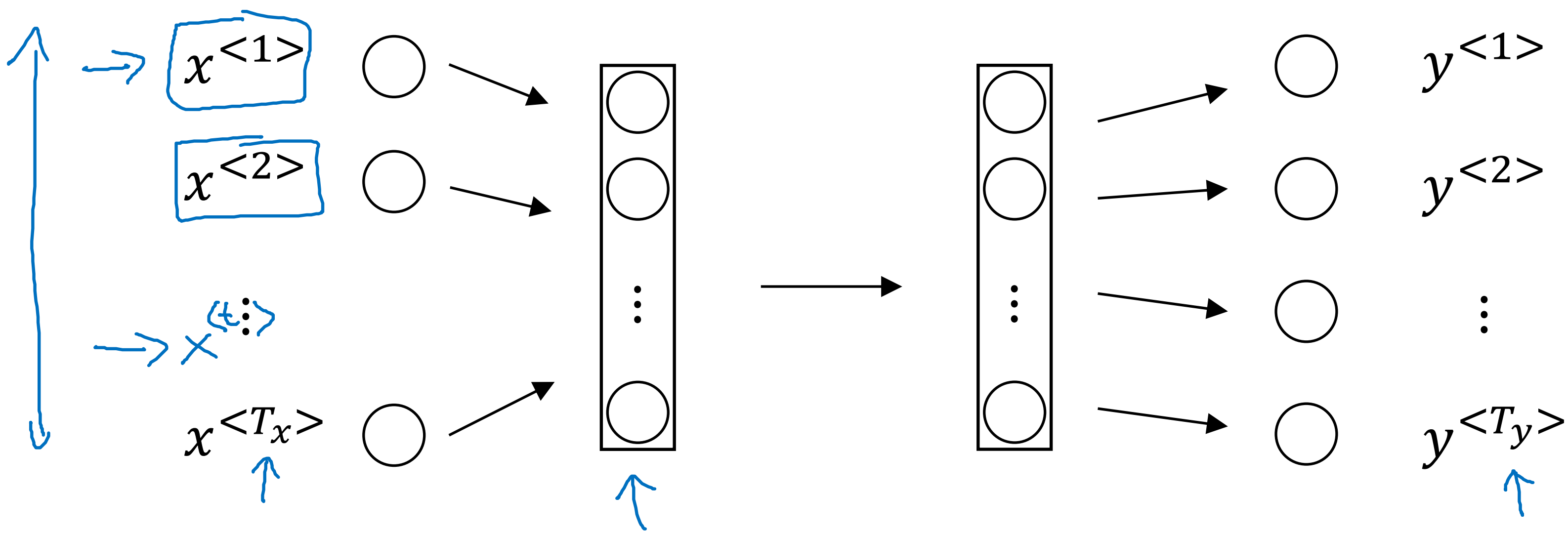
deeplearning.ai

Recurrent Neural Networks

Recurrent Neural Network Model

Why not a standard network?



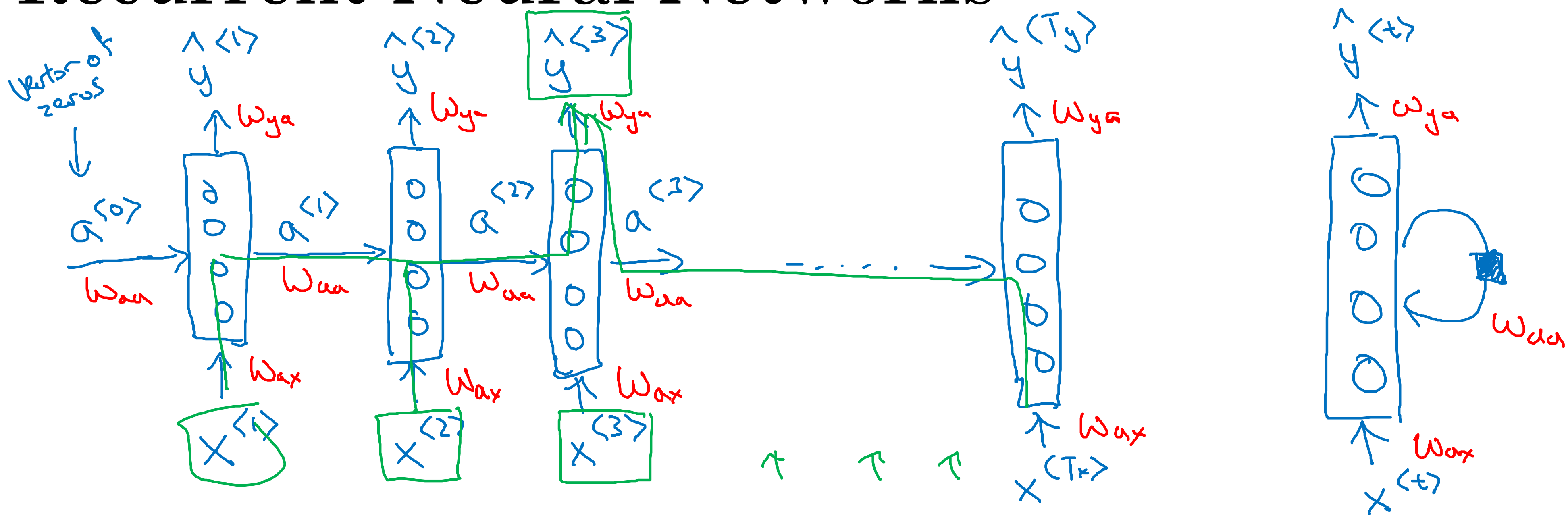
Problems:

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

Andrew Ng

Recurrent Neural Networks

$T_x = T_y$



Bidirectional RNN (BRNN)

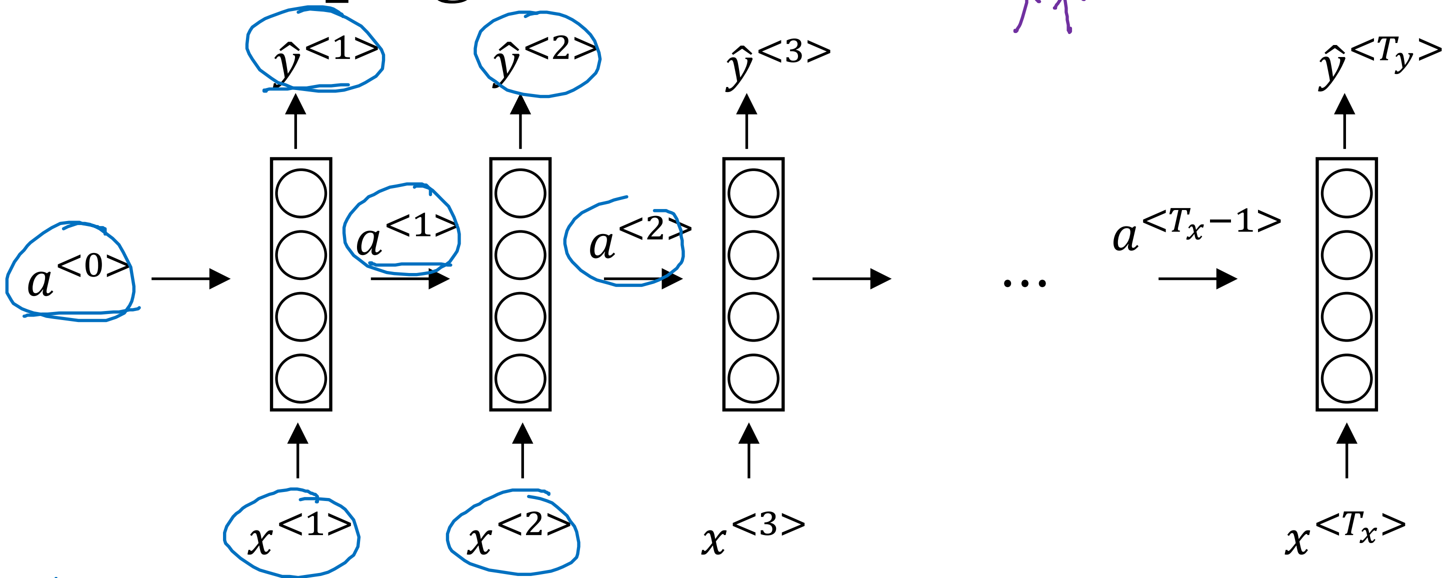
He said, "Teddy Roosevelt was a great President."

He said, "Teddy bears are on sale!"

Andrew Ng

Forward Propagation

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



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

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

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

$$a^{<t>} = g(\underbrace{\boxed{W_{aa}}}_{(100,100)} \underbrace{a^{<t-1>}}_{100} + \underbrace{\boxed{W_{ax}}}_{(100,10{,}000)} \underbrace{x^{<t>}}_{10{,}000} + b_a)$$

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

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

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

$$\underbrace{\left[W_{aa} \,\vdots\, W_{ax} \right]}_{100 \quad\quad 10\,000} \; (100) = W_a \qquad (100, 10100)$$

$$[a^{<t-1>}, x^{<t>}] = \begin{bmatrix} a^{<t-1>} \\ \hline x^{<t>} \end{bmatrix} \begin{array}{l} \updownarrow 100 \\ \updownarrow 10000 \end{array} \; \updownarrow 10100$$

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

Andrew Ng