Recurrent Neural Networks

LSTM (long short term memory) unit

GRU and LSTM

GRU

$$\tilde{c}^{<t>} = \tanh(W_c[\Gamma_r * c^{<t-1>}, x^{<t>}] + b_c)$$

$$\Gamma_u = \sigma(W_u[c^{<t-1>}, x^{<t>}] + b_u)$$

$$\Gamma_r = \sigma(W_r[c^{<t-1>}, x^{<t>}] + b_r)$$

$$c^{<t>} = \Gamma_u * \tilde{c}^{<t>} + (1 - \Gamma_u) * c^{<t-1>}$$

$$a^{<t>} = c^{<t>}$$

$$\Gamma_f$$

LSTM

$$\tilde{c}^{<t>} = \tanh(W_c[a^{<t-1>}, x^{<t>}] + b_c)$$

$$(\text{update}) \quad \Gamma_u = \sigma(W_u[a^{<t-1>}, x^{<t>}] + b_u)$$

$$(\text{forget}) \quad \Gamma_f = \sigma(W_f[a^{<t-1>}, x^{<t>}] + b_f)$$

$$(\text{output}) \quad \Gamma_o = \sigma(W_o[a^{<t-1>}, x^{<t>}] + b_o)$$

$$c^{<t>} = \Gamma_u * \tilde{c}^{<t>} + \Gamma_f * c^{<t-1>}$$

$$a^{<t>} = \Gamma_o * c^{<t>}$$

[Hochreiter & Schmidhuber 1997. Long short-term memory] ⟵

Andrew Ng

LSTM in pictures

$$\tilde{c}^{<t>} = \tanh(W_c[a^{<t-1>}, x^{<t>}] + b_c)$$

$$\Gamma_u = \sigma(W_u[\,a^{<t-1>}, x^{<t>}] + b_u)$$
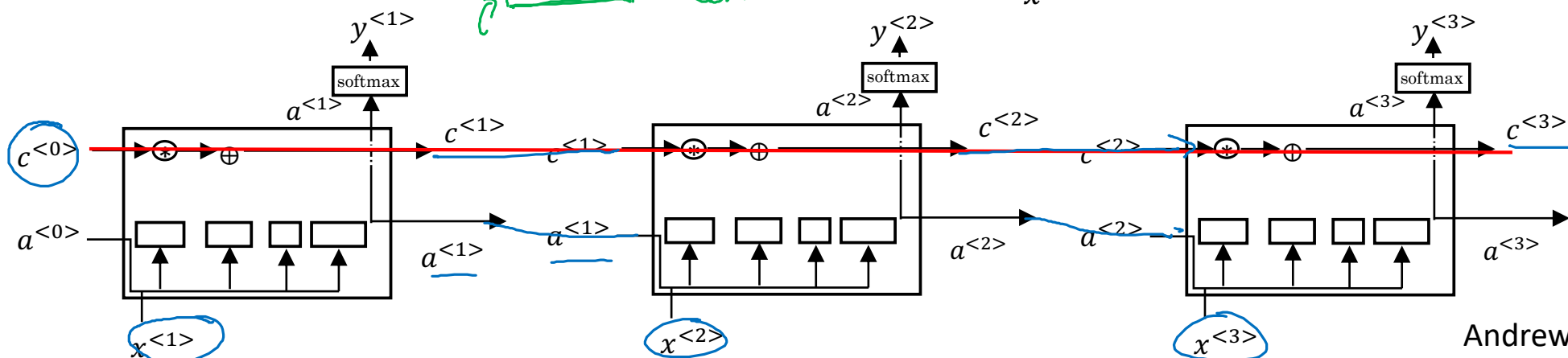
$$\Gamma_f = \sigma(W_f[\,a^{<t-1>}, x^{<t>}] + b_f)$$

$$\Gamma_o = \sigma(W_o[\,a^{<t-1>}, x^{<t>}] + b_o)$$

$$c^{<t>} = \Gamma_u * \tilde{c}^{<t>} + \Gamma_f * c^{<t-1>}$$

$$a^{<t>} = \Gamma_o * c^{<t>}$$

peephole connection

$c^{<t-1>}$

$y^{<t>}$

softmax

$a^{<t>}$

$c^{<t-1>}$ $c^{<t>}$ $c^{<t>}$

tanh

$a^{<t-1>}$

$f^{<t>}$ $i^{<t>}$ $\tilde{c}^{<t>}$ $o^{<t>}$ $a^{<t>}$

forget gate   update gate   tanh   output gate

$a^{<t>}$

$x^{<t>}$

$y^{<1>}$ $y^{<2>}$ $y^{<3>}$

softmax   softmax   softmax

$a^{<1>}$ $c^{<1>}$ $a^{<2>}$ $c^{<2>}$ $a^{<3>}$ $c^{<3>}$

$c^{<0>}$ $c^{<1>}$ $c^{<2>}$ $c^{<3>}$

$a^{<0>}$ $a^{<1>}$ $a^{<1>}$ $a^{<2>}$ $a^{<2>}$ $a^{<3>}$

$x^{<1>}$ $x^{<2>}$ $x^{<3>}$

Andrew Ng