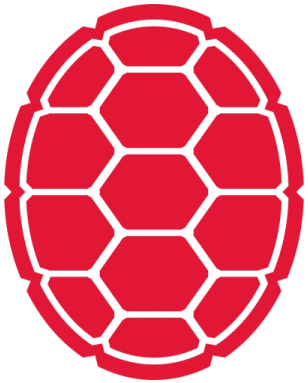
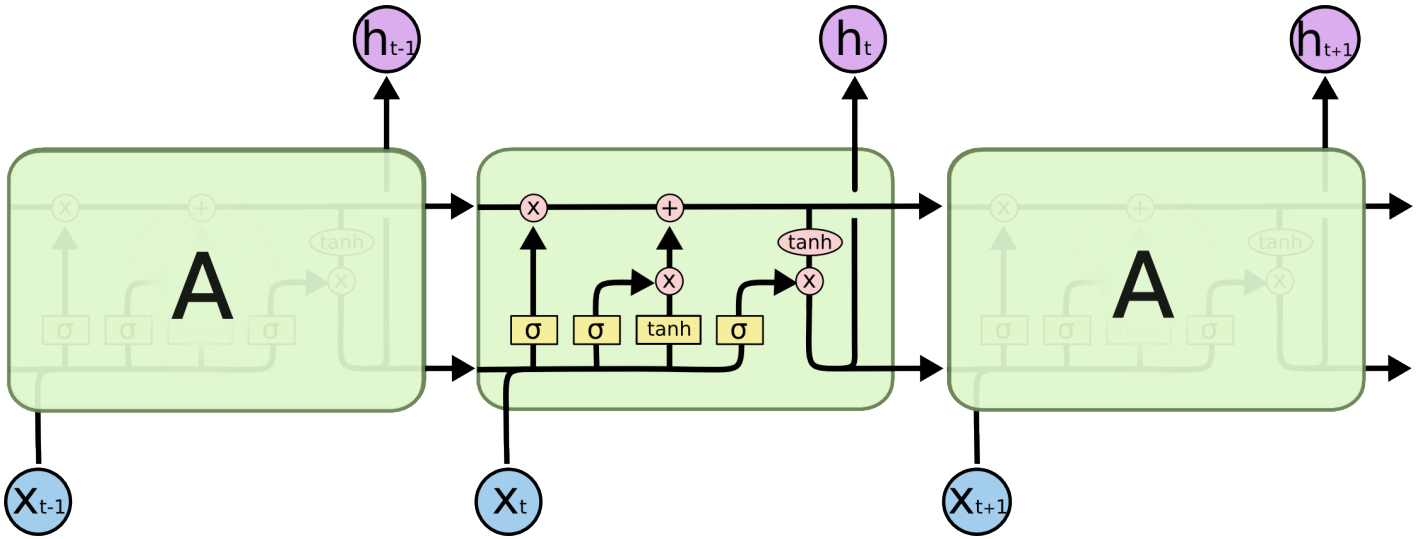


Long Short Term Memory Networks

Fenfei Guo and Jordan Boyd-Graber
University of Maryland

LSTM EXAMPLE

Recap of LSTM



Three gates: input (i_t), forget (f_t), out (o_t)

$$i_t = \sigma(W_{ii}x_t + b_{ii} + W_{hi}h_{t-1} + b_{hi})$$
$$f_t = \sigma(W_{if}x_t + b_{if} + W_{hf}h_{t-1} + b_{hf})$$
$$o_t = \sigma(W_{io}x_t + b_{io} + W_{ho}h_{t-1} + b_{ho})$$

New memory input: $\tilde{c}_t$

$$\tilde{c}_t = \tanh(W_{ic}x_t + b_{ic} + W_{hc}h_{t-1} + b_{hc})$$

Memorize and forget:

$$c_t = f_t * c_{t-1} + i_t * \tilde{c}_t$$
$$h_t = o_t * \tanh(c_t)$$

Figuring out this LSTM

A	B
1.0 0.0	0.0 1.0

- input sequence: A, A, B

$$x_1 = [1.0, 0.0] \quad x_2 = [1.0, 0.0] \quad x_3 = [0.0, 1.0]$$

Figuring out this LSTM

A	B
1.0 0.0	0.0 1.0

- input: A, A, B

$$x_1 = [1.0, 0.0] \quad x_2 = [1.0, 0.0] \quad x_3 = [0.0, 1.0]$$

- prediction output:

$$y_t = \text{softmax}(h_t) \quad [\text{number of hidden nodes} = 2]$$

Model parameters for x_t

Input's input gate

$$W_{ii} = \begin{bmatrix} 4 & 4 \\ 2 & 2 \end{bmatrix} \quad (1)$$

forget gate

$$W_{if} = \begin{bmatrix} -2 & 3 \\ 2 & 3 \end{bmatrix} \quad (2)$$

cell params

$$W_{ic} = \begin{bmatrix} 1 & 3 \\ 0 & -3 \end{bmatrix} \quad (3)$$

output gate

$$W_{io} = \begin{bmatrix} 5 & 5 \\ 3 & 5 \end{bmatrix} \quad (4)$$

Set all $b = 0$ for simplicity

Model parameters for h_t

input gate

$$W_{hi} = \begin{bmatrix} 1 & 0 \\ 4 & -2 \end{bmatrix} \tag{5}$$

cell params

$$W_{hc} = \begin{bmatrix} -4 & -8 \\ 4 & 3 \end{bmatrix} \tag{7}$$

forget gate

$$W_{hf} = \begin{bmatrix} -1 & -2 \\ 0 & 0 \end{bmatrix} \tag{6}$$

output gate

$$W_{ho} = \begin{bmatrix} 1 & 0 \\ 2 & 1 \end{bmatrix} \tag{8}$$

Set all $b = 0$ for simplicity

Inputs

- Initial hidden states:

$$h_0 = [0.0, 0.0]^\top$$

- Initial memory input:

$$c_0 = [0.0, 0.0]^\top$$

- Input sequences in time:

$$x_1 = \begin{bmatrix} 1.0 \\ 1.0 \end{bmatrix} \quad x_2 = \begin{bmatrix} 1.0 \\ 0.0 \end{bmatrix} \quad x_3 = \begin{bmatrix} 0.0 \\ 1.0 \end{bmatrix}$$

Forwards at time step 1: i_1

Input's input gate

$$W_{ii} = \begin{bmatrix} 4 & 4 \\ 2 & 2 \end{bmatrix} \tag{9}$$

input gate

$$W_{hi} = \begin{bmatrix} 1 & 0 \\ 4 & -2 \end{bmatrix} \tag{10}$$

Compute

$$i_1 = \sigma\left(W_{ii} x_1 + W_{hi} h_0\right) \tag{11}$$

(12)

Forwards at time step 1: i_1

Input's input gate

$$W_{ii} = \begin{bmatrix} 4 & 4 \\ 2 & 2 \end{bmatrix} \tag{9}$$

input gate

$$W_{hi} = \begin{bmatrix} 1 & 0 \\ 4 & -2 \end{bmatrix} \tag{10}$$

Compute

$$i_1 = \sigma(W_{ii} x_1 + W_{hi} h_0) \tag{11}$$

$$= \sigma\left(\begin{bmatrix} 4 & 4 \\ 2 & 2 \end{bmatrix} \times \begin{bmatrix} 1.0 \\ 0.0 \end{bmatrix}\right) \tag{12}$$

$$\tag{13}$$

Forwards at time step 1: i_1

Input's input gate

$$W_{ii} = \begin{bmatrix} 4 & 4 \\ 2 & 2 \end{bmatrix} \tag{9}$$

input gate

$$W_{hi} = \begin{bmatrix} 1 & 0 \\ 4 & -2 \end{bmatrix} \tag{10}$$

Compute

$$\begin{aligned} i_1 &= \sigma(W_{ii} x_1 + W_{hi} h_0) && (11) \\ &= \sigma\left(\begin{bmatrix} 4 & 4 \\ 2 & 2 \end{bmatrix} \times \begin{bmatrix} 1.0 \\ 0.0 \end{bmatrix}\right) && (12) \\ &= \sigma([4.0, 2.0]^\top) && (13) \\ & && (14) \end{aligned}$$

Forwards at time step 1: i_1

Input's input gate

$$W_{ii} = \begin{bmatrix} 4 & 4 \\ 2 & 2 \end{bmatrix} \tag{9}$$

input gate

$$W_{hi} = \begin{bmatrix} 1 & 0 \\ 4 & -2 \end{bmatrix} \tag{10}$$

Compute

$$\begin{aligned} i_1 &= \sigma(W_{ii} x_1 + W_{hi} h_0) && (11) \\ &= \sigma\left(\begin{bmatrix} 4 & 4 \\ 2 & 2 \end{bmatrix} \times \begin{bmatrix} 1.0 \\ 0.0 \end{bmatrix}\right) && (12) \\ &= \sigma([4.0, 2.0]^\top) && (13) \\ &= [1.0, 0.9]^\top && (14) \end{aligned}$$

Forwards at time step 1: f_1

forget gate

$$W_{if} = \begin{bmatrix} -2 & 3 \\ 2 & 3 \end{bmatrix} \tag{15}$$

forget gate

$$W_{hf} = \begin{bmatrix} -1 & -2 \\ 0 & 0 \end{bmatrix} \tag{16}$$

Compute

$$f_1 = \sigma\left(W_{if} x_1 + W_{hf} h_0\right) \tag{17}$$

$$\tag{18}$$

Forwards at time step 1: f_1

forget gate

$$W_{if} = \begin{bmatrix} -2 & 3 \\ 2 & 3 \end{bmatrix} \tag{15}$$

forget gate

$$W_{hf} = \begin{bmatrix} -1 & -2 \\ 0 & 0 \end{bmatrix} \tag{16}$$

Compute

$$f_1 = \sigma\left(W_{if} x_1 + W_{hf} h_0\right) \tag{17}$$

$$= \sigma\left(\begin{bmatrix} -2 & 3 \\ 2 & 3 \end{bmatrix} \times \begin{bmatrix} 1.0 \\ 0.0 \end{bmatrix}\right) \tag{18}$$

$$\tag{19}$$

Forwards at time step 1: f_1

forget gate

$$W_{if} = \begin{bmatrix} -2 & 3 \\ 2 & 3 \end{bmatrix} \tag{15}$$

forget gate

$$W_{hf} = \begin{bmatrix} -1 & -2 \\ 0 & 0 \end{bmatrix} \tag{16}$$

Compute

$$f_1 = \sigma(W_{if} x_1 + W_{hf} h_0) \tag{17}$$

$$= \sigma\left(\begin{bmatrix} -2 & 3 \\ 2 & 3 \end{bmatrix} \times \begin{bmatrix} 1.0 \\ 0.0 \end{bmatrix}\right) \tag{18}$$

$$= \sigma([-2.0, 2.0]^\top) \tag{19}$$

$$\tag{20}$$

Forwards at time step 1: f_1

forget gate

$$W_{if} = \begin{bmatrix} -2 & 3 \\ 2 & 3 \end{bmatrix} \tag{15}$$

forget gate

$$W_{hf} = \begin{bmatrix} -1 & -2 \\ 0 & 0 \end{bmatrix} \tag{16}$$

Compute

$$f_1 = \sigma(W_{if} x_1 + W_{hf} h_0) \tag{17}$$

$$= \sigma\left(\begin{bmatrix} -2 & 3 \\ 2 & 3 \end{bmatrix} \times \begin{bmatrix} 1.0 \\ 0.0 \end{bmatrix}\right) \tag{18}$$

$$= \sigma([-2.0, 2.0]^\top) \tag{19}$$

$$= [0.1, 0.9]^\top \tag{20}$$

Forwards at time step 1: o_1

output gate

$$W_{io} = \begin{bmatrix} 5 & 5 \\ 3 & 5 \end{bmatrix} \tag{21}$$

output gate

$$W_{ho} = \begin{bmatrix} 1 & 0 \\ 2 & 1 \end{bmatrix} \tag{22}$$

- $o_1 = \sigma(W_{io}x_1 + W_{ho}h_0)$

Forwards at time step 1: o_1

output gate

$$W_{io} = \begin{bmatrix} 5 & 5 \\ 3 & 5 \end{bmatrix} \tag{21}$$

output gate

$$W_{ho} = \begin{bmatrix} 1 & 0 \\ 2 & 1 \end{bmatrix} \tag{22}$$

- $o_1 = \sigma(W_{io}x_1 + W_{ho}h_0)$

$$= \sigma\left(\begin{bmatrix} 5 & 5 \\ 3 & 5 \end{bmatrix} \times \begin{bmatrix} 1.0 \\ 0.0 \end{bmatrix}\right)$$

Forwards at time step 1: o_1

output gate

$$W_{io} = \begin{bmatrix} 5 & 5 \\ 3 & 5 \end{bmatrix} \tag{21}$$

output gate

$$W_{ho} = \begin{bmatrix} 1 & 0 \\ 2 & 1 \end{bmatrix} \tag{22}$$

- $o_1 = \sigma(W_{io}x_1 + W_{ho}h_0)$

$$= \sigma\left(\begin{bmatrix} 5 & 5 \\ 3 & 5 \end{bmatrix} \times \begin{bmatrix} 1.0 \\ 0.0 \end{bmatrix}\right) \quad = \sigma([5.0, 3.0]^\top)$$

Forwards at time step 1: o_1

output gate

$$W_{io} = \begin{bmatrix} 5 & 5 \\ 3 & 5 \end{bmatrix} \tag{21}$$

output gate

$$W_{ho} = \begin{bmatrix} 1 & 0 \\ 2 & 1 \end{bmatrix} \tag{22}$$

- $o_1 = \sigma(W_{io}x_1 + W_{ho}h_0)$

$$= \sigma\left(\begin{bmatrix} 5 & 5 \\ 3 & 5 \end{bmatrix} \times \begin{bmatrix} 1.0 \\ 0.0 \end{bmatrix}\right) \quad = \sigma([5.0, 3.0]^\top)$$

$$= [1.0, 1.0]^\top$$

Forwards at time step 1: $\tilde{c}_1$

cell params

$$W_{ic} = \begin{bmatrix} 1 & 3 \\ 0 & -3 \end{bmatrix} \tag{23}$$

cell params

$$W_{hc} = \begin{bmatrix} -4 & -8 \\ 4 & 3 \end{bmatrix} \tag{24}$$

- $\tilde{c}_1 = \tanh(W_{ic}x_1 + W_{hc}h_0)$

Forwards at time step 1: $\tilde{c}_1$

cell params

$$W_{ic} = \begin{bmatrix} 1 & 3 \\ 0 & -3 \end{bmatrix} \tag{23}$$

cell params

$$W_{hc} = \begin{bmatrix} -4 & -8 \\ 4 & 3 \end{bmatrix} \tag{24}$$

- $\tilde{c}_1 = \tanh(W_{ic}x_1 + W_{hc}h_0)$

$$= \tanh\left(\begin{bmatrix} 1 & 3 \\ 0 & -3 \end{bmatrix} \times \begin{bmatrix} 1.0 \\ 0.0 \end{bmatrix}\right)$$

Forwards at time step 1: $\tilde{c}_1$

cell params

$$W_{ic} = \begin{bmatrix} 1 & 3 \\ 0 & -3 \end{bmatrix} \tag{23}$$

cell params

$$W_{hc} = \begin{bmatrix} -4 & -8 \\ 4 & 3 \end{bmatrix} \tag{24}$$

- $\tilde{c}_1 = \tanh(W_{ic}x_1 + W_{hc}h_0)$

$$= \tanh\left(\begin{bmatrix} 1 & 3 \\ 0 & -3 \end{bmatrix} \times \begin{bmatrix} 1.0 \\ 0.0 \end{bmatrix}\right) \quad = \tanh([1.0, 0.0]^\top)$$

Forwards at time step 1: $\tilde{c}_1$

cell params

$$W_{ic} = \begin{bmatrix} 1 & 3 \\ 0 & -3 \end{bmatrix} \tag{23}$$

cell params

$$W_{hc} = \begin{bmatrix} -4 & -8 \\ 4 & 3 \end{bmatrix} \tag{24}$$

- $\tilde{c}_1 = \tanh(W_{ic}x_1 + W_{hc}h_0)$

$$= \tanh\left(\begin{bmatrix} 1 & 3 \\ 0 & -3 \end{bmatrix} \times \begin{bmatrix} 1.0 \\ 0.0 \end{bmatrix}\right) \quad = \tanh([1.0, 0.0]^\top) \quad = [0.8, 0.0]^\top$$

Forwards at time step 1

f_1	c_0	i_1	$\tilde{c}_1$
$[0.1, 0.9]^\top$	$[0.0, 0.0]^\top$	$[1.0, 0.9]^\top$	$[0.8, 0.0]^\top$

- Message forward (c_1)

$$c_1 = f_1 \circ c_0 + i_1 \circ \tilde{c}_1 \tag{25}$$

(26)

Forwards at time step 1

f_1	c_0	i_1	$\tilde{c}_1$
$[0.1, 0.9]^\top$	$[0.0, 0.0]^\top$	$[1.0, 0.9]^\top$	$[0.8, 0.0]^\top$

- Message forward (c_1)

$$c_1 = f_1 \circ c_0 + i_1 \circ \tilde{c}_1 \tag{25}$$

$$= [1.0, 0.9]^\top \circ [0.8, 0.0]^\top \tag{26}$$

$$\tag{27}$$

Forwards at time step 1

f_1	c_0	i_1	$\tilde{c}_1$
$[0.1, 0.9]^\top$	$[0.0, 0.0]^\top$	$[1.0, 0.9]^\top$	$[0.8, 0.0]^\top$

- Message forward (c_1)

$$c_1 = f_1 \circ c_0 + i_1 \circ \tilde{c}_1 \tag{25}$$

$$= [1.0, 0.9]^\top \circ [0.8, 0.0]^\top \tag{26}$$

$$\tag{27}$$

Forwards at time step 1

f_1	c_0	i_1	$\tilde{c}_1$
$[0.1, 0.9]^\top$	$[0.0, 0.0]^\top$	$[1.0, 0.9]^\top$	$[0.8, 0.0]^\top$

- Message forward (c_1)

$$c_1 = [0.8, 0.0]^\top \tag{25}$$

- New hidden (h_1)

$$h_1 \tag{26}$$

Forwards at time step 1

f_1	c_0	i_1	$\tilde{c}_1$
$[0.1, 0.9]^\top$	$[0.0, 0.0]^\top$	$[1.0, 0.9]^\top$	$[0.8, 0.0]^\top$

- Message forward (c_1)

$$c_1 = [0.8, 0.0]^\top \tag{25}$$

- New hidden (h_1)

$$h_1 = o_1 \circ \tanh(c_1) \tag{26}$$

$$\tag{27}$$

Forwards at time step 1

f_1	c_0	i_1	$\tilde{c}_1$
$[0.1, 0.9]^\top$	$[0.0, 0.0]^\top$	$[1.0, 0.9]^\top$	$[0.8, 0.0]^\top$

- Message forward (c_1)

$$c_1 = [0.8, 0.0]^\top \tag{25}$$

- New hidden (h_1)

$$h_1 = o_1 \circ \tanh(c_1) \tag{26}$$
$$= [1.0, 1.0]^\top \circ \tanh([0.8, 0.0]^\top) \tag{27}$$
$$\tag{28}$$

Forwards at time step 1

f_1	c_0	i_1	$\tilde{c}_1$
$[0.1, 0.9]^\top$	$[0.0, 0.0]^\top$	$[1.0, 0.9]^\top$	$[0.8, 0.0]^\top$

- Message forward (c_1)

$$c_1 = [0.8, 0.0]^\top \tag{25}$$

- New hidden (h_1)

$$\begin{aligned} h_1 &= o_1 \circ \tanh(c_1) && (26)\\ &= [1.0, 1.0]^\top \circ \tanh([0.8, 0.0]^\top) && (27)\\ &= [0.7, 0.0]^\top && (28) \end{aligned}$$

Forwards at time step 1

f_1	c_0	i_1	$\tilde{c}_1$
$[0.1, 0.9]^\top$	$[0.0, 0.0]^\top$	$[1.0, 0.9]^\top$	$[0.8, 0.0]^\top$

- Message forward (c_1)

$$c_1 = [0.8, 0.0]^\top \tag{25}$$

- New hidden (h_1)

$$h_1 = [0.7, 0.0]^\top \tag{26}$$

- Prediction $y_1 = \mathrm{softmax}(h_1)$

Forwards at time step 2

$t = 2$ State

$x_2 = [1.0, 0.0]^\top;\ c_1 = [0.8, 0.0]^\top;\ h_1 = [0.7, 0.0]^\top$

Input's input gate

$$W_{ii} = \begin{bmatrix} 4 & 4 \\ 2 & 2 \end{bmatrix} \tag{27}$$

input gate

$$W_{hi} = \begin{bmatrix} 1 & 0 \\ 4 & -2 \end{bmatrix} \tag{28}$$

$$i_2 = \sigma\left(W_{ii} x_2 + W_{hi} h_1\right) \tag{29}$$

(30)

Forwards at time step 2

$t = 2$ State

$x_2 = [1.0, 0.0]^\top;\ c_1 = [0.8, 0.0]^\top;\ h_1 = [0.7, 0.0]^\top$

Input's input gate

$$W_{ii} = \begin{bmatrix} 4 & 4 \\ 2 & 2 \end{bmatrix} \tag{27}$$

input gate

$$W_{hi} = \begin{bmatrix} 1 & 0 \\ 4 & -2 \end{bmatrix} \tag{28}$$

$$i_2 = \sigma\left(W_{ii} x_2 + W_{hi} h_1\right) \tag{29}$$

$$= \sigma\left(\begin{bmatrix} 4 & 4 \\ 2 & 2 \end{bmatrix} \times \begin{bmatrix} 1.0 \\ 0.0 \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ 4 & -2 \end{bmatrix} \times \begin{bmatrix} 0.7 \\ 0.0 \end{bmatrix}\right) \tag{30}$$

$$\tag{31}$$

Forwards at time step 2

$t = 2$ State

$x_2 = [1.0, 0.0]^\top;\ c_1 = [0.8, 0.0]^\top;\ h_1 = [0.7, 0.0]^\top$

Input's input gate

$$W_{ii} = \begin{bmatrix} 4 & 4 \\ 2 & 2 \end{bmatrix} \tag{27}$$

input gate

$$W_{hi} = \begin{bmatrix} 1 & 0 \\ 4 & -2 \end{bmatrix} \tag{28}$$

$$i_2 = \sigma(W_{ii} x_2 + W_{hi} h_1) \tag{29}$$

$$= \sigma\left(\begin{bmatrix} 4 & 4 \\ 2 & 2 \end{bmatrix} \times \begin{bmatrix} 1.0 \\ 0.0 \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ 4 & -2 \end{bmatrix} \times \begin{bmatrix} 0.7 \\ 0.0 \end{bmatrix}\right) \tag{30}$$

$$= \sigma([4.0, 2.0]^\top + [0.7, 2.8]^\top) \tag{31}$$

$$\tag{32}$$

Forwards at time step 2

$t = 2$ State

$x_2 = [1.0, 0.0]^\top;\ c_1 = [0.8, 0.0]^\top;\ h_1 = [0.7, 0.0]^\top$

Input's input gate

$$W_{ii} = \begin{bmatrix} 4 & 4 \\ 2 & 2 \end{bmatrix} \tag{27}$$

input gate

$$W_{hi} = \begin{bmatrix} 1 & 0 \\ 4 & -2 \end{bmatrix} \tag{28}$$

$$i_2 = \sigma(W_{ii} x_2 + W_{hi} h_1) \tag{29}$$

$$= \sigma\left(\begin{bmatrix} 4 & 4 \\ 2 & 2 \end{bmatrix} \times \begin{bmatrix} 1.0 \\ 0.0 \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ 4 & -2 \end{bmatrix} \times \begin{bmatrix} 0.7 \\ 0.0 \end{bmatrix}\right) \tag{30}$$

$$= \sigma([4.0, 2.0]^\top + [0.7, 2.8]^\top) = \sigma([4.7, 4.8]^\top) \tag{31}$$

$$= [1.0, 1.0]^\top \tag{32}$$

Forwards at time step 2

$t = 2$ State

$x_2 = [1.0, 0.0]^\top;\ c_1 = [0.8, 0.0]^\top;\ h_1 = [0.7, 0.0]^\top$

Input's input gate

$$W_{ii} = \begin{bmatrix} 4 & 4 \\ 2 & 2 \end{bmatrix} \tag{33}$$

input gate

$$W_{hi} = \begin{bmatrix} 1 & 0 \\ 4 & -2 \end{bmatrix} \tag{34}$$

$$f_2 = \sigma\left(W_{if} x_2 + W_{hf} h_1\right) \tag{35}$$

(36)

Forwards at time step 2

$t = 2$ State

$x_2 = [1.0, 0.0]^\top;\ c_1 = [0.8, 0.0]^\top;\ h_1 = [0.7, 0.0]^\top$

Input's input gate

$$W_{ii} = \begin{bmatrix} 4 & 4 \\ 2 & 2 \end{bmatrix} \tag{33}$$

input gate

$$W_{hi} = \begin{bmatrix} 1 & 0 \\ 4 & -2 \end{bmatrix} \tag{34}$$

$$f_2 = \sigma\left(W_{if} x_2 + W_{hf} h_1\right) \tag{35}$$

$$= \sigma\left(\begin{bmatrix} -2 & 3 \\ 2 & 3 \end{bmatrix} \times \begin{bmatrix} 1.0 \\ 0.0 \end{bmatrix} + \begin{bmatrix} -1 & -2 \\ 0 & 0 \end{bmatrix} \times \begin{bmatrix} 0.7 \\ 0.0 \end{bmatrix}\right) \tag{36}$$

(37)

Forwards at time step 2

$t = 2$ State

$x_2 = [1.0, 0.0]^\top;\ c_1 = [0.8, 0.0]^\top;\ h_1 = [0.7, 0.0]^\top$

Input's input gate

$$W_{ii} = \begin{bmatrix} 4 & 4 \\ 2 & 2 \end{bmatrix} \tag{33}$$

input gate

$$W_{hi} = \begin{bmatrix} 1 & 0 \\ 4 & -2 \end{bmatrix} \tag{34}$$

$$f_2 = \sigma(W_{if} x_2 + W_{hf} h_1) \tag{35}$$

$$= \sigma\left(\begin{bmatrix} -2 & 3 \\ 2 & 3 \end{bmatrix} \times \begin{bmatrix} 1.0 \\ 0.0 \end{bmatrix} + \begin{bmatrix} -1 & -2 \\ 0 & 0 \end{bmatrix} \times \begin{bmatrix} 0.7 \\ 0.0 \end{bmatrix}\right) \tag{36}$$

$$= \sigma([-2.0, 2.0]^\top + [-0.7, 0.0]^\top) \tag{37}$$

$$\tag{38}$$

Forwards at time step 2

$t = 2$ State

$x_2 = [1.0, 0.0]^\top;\ c_1 = [0.8, 0.0]^\top;\ h_1 = [0.7, 0.0]^\top$

Input's input gate

$$W_{ii} = \begin{bmatrix} 4 & 4 \\ 2 & 2 \end{bmatrix} \tag{33}$$

input gate

$$W_{hi} = \begin{bmatrix} 1 & 0 \\ 4 & -2 \end{bmatrix} \tag{34}$$

$$f_2 = \sigma(W_{if} x_2 + W_{hf} h_1) \tag{35}$$

$$= \sigma\left(\begin{bmatrix} -2 & 3 \\ 2 & 3 \end{bmatrix} \times \begin{bmatrix} 1.0 \\ 0.0 \end{bmatrix} + \begin{bmatrix} -1 & -2 \\ 0 & 0 \end{bmatrix} \times \begin{bmatrix} 0.7 \\ 0.0 \end{bmatrix}\right) \tag{36}$$

$$= \sigma([-2.0, 2.0]^\top + [-0.7, 0.0]^\top) \tag{37}$$

$$= \sigma([-2.7, 2.0]^\top) = [0.1, 0.9]^\top \tag{38}$$

Forwards at time step 2

$t = 2$ State

$x_2 = [1.0, 0.0]^\top;\ c_1 = [0.8, 0.0]^\top;\ h_1 = [0.7, 0.0]^\top$

output gate

$$W_{io} = \begin{bmatrix} 5 & 5 \\ 3 & 5 \end{bmatrix} \tag{39}$$

output gate

$$W_{ho} = \begin{bmatrix} 1 & 0 \\ 2 & 1 \end{bmatrix} \tag{40}$$

$$o_2 = \sigma\left(W_{io} x_2 + W_{ho} h_1\right) \tag{41}$$

(42)

Forwards at time step 2

$t = 2$ State

$x_2 = [1.0, 0.0]^\top;\ c_1 = [0.8, 0.0]^\top;\ h_1 = [0.7, 0.0]^\top$

output gate

$$W_{io} = \begin{bmatrix} 5 & 5 \\ 3 & 5 \end{bmatrix} \tag{39}$$

output gate

$$W_{ho} = \begin{bmatrix} 1 & 0 \\ 2 & 1 \end{bmatrix} \tag{40}$$

$$o_2 = \sigma\left(W_{io} x_2 + W_{ho} h_1\right) \tag{41}$$

$$= \sigma\left(\begin{bmatrix} 5 & 5 \\ 3 & 5 \end{bmatrix} \times \begin{bmatrix} 1.0 \\ 0.0 \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ 2 & 1 \end{bmatrix} \times \begin{bmatrix} 0.7 \\ 0.0 \end{bmatrix}\right) \tag{42}$$

$$\tag{43}$$

Forwards at time step 2

$t = 2$ State

$x_2 = [1.0, 0.0]^\top;\ c_1 = [0.8, 0.0]^\top;\ h_1 = [0.7, 0.0]^\top$

output gate

$$W_{io} = \begin{bmatrix} 5 & 5 \\ 3 & 5 \end{bmatrix} \tag{39}$$

output gate

$$W_{ho} = \begin{bmatrix} 1 & 0 \\ 2 & 1 \end{bmatrix} \tag{40}$$

$$o_2 = \sigma(W_{io}x_2 + W_{ho}h_1) \tag{41}$$

$$= \sigma\left(\begin{bmatrix} 5 & 5 \\ 3 & 5 \end{bmatrix} \times \begin{bmatrix} 1.0 \\ 0.0 \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ 2 & 1 \end{bmatrix} \times \begin{bmatrix} 0.7 \\ 0.0 \end{bmatrix}\right) \tag{42}$$

$$= \sigma([5.0, 3.0]^\top + [0.7, 1.4]^\top) \tag{43}$$

$$\tag{44}$$

Forwards at time step 2

$t = 2$ State

$x_2 = [1.0, 0.0]^\top;\ c_1 = [0.8, 0.0]^\top;\ h_1 = [0.7, 0.0]^\top$

output gate

$$W_{io} = \begin{bmatrix} 5 & 5 \\ 3 & 5 \end{bmatrix} \tag{39}$$

output gate

$$W_{ho} = \begin{bmatrix} 1 & 0 \\ 2 & 1 \end{bmatrix} \tag{40}$$

$$o_2 = \sigma(W_{io}x_2 + W_{ho}h_1) \tag{41}$$

$$= \sigma\left(\begin{bmatrix} 5 & 5 \\ 3 & 5 \end{bmatrix} \times \begin{bmatrix} 1.0 \\ 0.0 \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ 2 & 1 \end{bmatrix} \times \begin{bmatrix} 0.7 \\ 0.0 \end{bmatrix}\right) \tag{42}$$

$$= \sigma([5.0, 3.0]^\top + [0.7, 1.4]^\top) \tag{43}$$

$$= \sigma([5.7, 4.4]^\top) = [1.0, 1.0]^\top \tag{44}$$

Forwards at time step 2

$t = 2$ State

$x_2 = [1.0, 0.0]^\top;\ c_1 = [0.8, 0.0]^\top;\ h_1 = [0.7, 0.0]^\top$

cell params

$$W_{ic} = \begin{bmatrix} 1 & 3 \\ 0 & -3 \end{bmatrix} \tag{45}$$

cell params

$$W_{hc} = \begin{bmatrix} -4 & -8 \\ 4 & 3 \end{bmatrix} \tag{46}$$

$$\tilde{c_2} = \tanh\left(W_{ic} x_2 + W_{hc} h_1 \right) \tag{47}$$

(48)

Forwards at time step 2

$t = 2$ State

$x_2 = [1.0, 0.0]^\top;\ c_1 = [0.8, 0.0]^\top;\ h_1 = [0.7, 0.0]^\top$

cell params

$$W_{ic} = \begin{bmatrix} 1 & 3 \\ 0 & -3 \end{bmatrix} \tag{45}$$

cell params

$$W_{hc} = \begin{bmatrix} -4 & -8 \\ 4 & 3 \end{bmatrix} \tag{46}$$

$$\tilde{c_2} = \tanh\left(W_{ic}x_2 + W_{hc}h_1\right) \tag{47}$$

$$= \tanh\left(\begin{bmatrix} 1 & 3 \\ 0 & -3 \end{bmatrix} \times \begin{bmatrix} 1.0 \\ 0.0 \end{bmatrix} + \begin{bmatrix} -4 & -8 \\ 4 & 3 \end{bmatrix} \times \begin{bmatrix} 0.7 \\ 0.0 \end{bmatrix}\right) \tag{48}$$

$$\tag{49}$$

Forwards at time step 2

$t = 2$ State

$x_2 = [1.0, 0.0]^\top;\ c_1 = [0.8, 0.0]^\top;\ h_1 = [0.7, 0.0]^\top$

cell params

$$W_{ic} = \begin{bmatrix} 1 & 3 \\ 0 & -3 \end{bmatrix} \tag{45}$$

cell params

$$W_{hc} = \begin{bmatrix} -4 & -8 \\ 4 & 3 \end{bmatrix} \tag{46}$$

$$\tilde{c_2} = \tanh(W_{ic}x_2 + W_{hc}h_1) \tag{47}$$

$$= \tanh\left(\begin{bmatrix} 1 & 3 \\ 0 & -3 \end{bmatrix} \times \begin{bmatrix} 1.0 \\ 0.0 \end{bmatrix} + \begin{bmatrix} -4 & -8 \\ 4 & 3 \end{bmatrix} \times \begin{bmatrix} 0.7 \\ 0.0 \end{bmatrix}\right) \tag{48}$$

$$= \tanh([1.0, 0.0]^\top + [-2.8, 2.8]^\top) \tag{49}$$

$$\tag{50}$$

Forwards at time step 2

$t = 2$ State

$x_2 = [1.0, 0.0]^\top;\ c_1 = [0.8, 0.0]^\top;\ h_1 = [0.7, 0.0]^\top$

cell params

$$W_{ic} = \begin{bmatrix} 1 & 3 \\ 0 & -3 \end{bmatrix} \tag{45}$$

cell params

$$W_{hc} = \begin{bmatrix} -4 & -8 \\ 4 & 3 \end{bmatrix} \tag{46}$$

$$\tilde{c_2} = \tanh(W_{ic}x_2 + W_{hc}h_1) \tag{47}$$

$$= \tanh\left(\begin{bmatrix} 1 & 3 \\ 0 & -3 \end{bmatrix} \times \begin{bmatrix} 1.0 \\ 0.0 \end{bmatrix} + \begin{bmatrix} -4 & -8 \\ 4 & 3 \end{bmatrix} \times \begin{bmatrix} 0.7 \\ 0.0 \end{bmatrix}\right) \tag{48}$$

$$= \tanh([1.0, 0.0]^\top + [-2.8, 2.8]^\top) \tag{49}$$

$$= \tanh([-1.8, 2.8]^\top) = [-0.9, 1.0]^\top \tag{50}$$

Forwards at time step 2

$\tilde{c}_2$	i_2	f_2	c_1
$[-0.9, 1.0]^\top$	$[1.0, 1.0]^\top$	$[0.1, 0.9]^\top$	$[0.8, 0.0]^\top$

- Message

$$c_2 = f_2 \circ c_1 + i_2 \circ \tilde{c}_2 \tag{51}$$

(52)

- Hidden

Forwards at time step 2

$\tilde{c}_2$	i_2	f_2	c_1
$[-0.9, 1.0]^\top$	$[1.0, 1.0]^\top$	$[0.1, 0.9]^\top$	$[0.8, 0.0]^\top$

- Message

$$c_2 = f_2 \circ c_1 + i_2 \circ \tilde{c}_2 \tag{51}$$
$$= [0.1, 0.9]^\top \circ [0.8, 0.0]^\top + [1.0, 1.0]^\top \circ [-0.9, 1.0]^\top \tag{52}$$
$$\tag{53}$$

- Hidden

Forwards at time step 2

$\tilde{c}_2$	i_2	f_2	c_1
$[-0.9, 1.0]^\top$	$[1.0, 1.0]^\top$	$[0.1, 0.9]^\top$	$[0.8, 0.0]^\top$

- Message

$$
\begin{aligned}
c_2 &= f_2 \circ c_1 + i_2 \circ \tilde{c}_2 && (51)\\
&= [0.1, 0.9]^\top \circ [0.8, 0.0]^\top + [1.0, 1.0]^\top \circ [-0.9, 1.0]^\top && (52)\\
&= [-0.8, 1.0]^\top && (53)\\
& && (54)
\end{aligned}
$$

- Hidden

Forwards at time step 2

$\tilde{c}_2$	i_2	f_2	c_1
$[-0.9, 1.0]^\top$	$[1.0, 1.0]^\top$	$[0.1, 0.9]^\top$	$[0.8, 0.0]^\top$

- Message

$$c_2 = f_2 \circ c_1 + i_2 \circ \tilde{c}_2 \tag{51}$$
$$= [0.1, 0.9]^\top \circ [0.8, 0.0]^\top + [1.0, 1.0]^\top \circ [-0.9, 1.0]^\top \tag{52}$$
$$= [-0.8, 1.0]^\top \tag{53}$$
$$\tag{54}$$

- Hidden

$$h_2 = o_2 \circ \tanh(c_2) \tag{55}$$
$$\tag{56}$$

Forwards at time step 2

$\tilde{c}_2$	i_2	f_2	c_1
$[-0.9, 1.0]^\top$	$[1.0, 1.0]^\top$	$[0.1, 0.9]^\top$	$[0.8, 0.0]^\top$

- Message

$$c_2 = f_2 \circ c_1 + i_2 \circ \tilde{c}_2 \tag{51}$$
$$= [-0.8, 1.0]^\top \tag{52}$$
$$\tag{53}$$

- Hidden

$$h_2 = o_2 \circ \tanh(c_2) \tag{54}$$
$$= [1.0, 1.0]^\top \circ \tanh([-0.8, 1.0]^\top) \tag{55}$$
$$\tag{56}$$

Forwards at time step 2

$\tilde{c}_2$	i_2	f_2	c_1
$[-0.9, 1.0]^\top$	$[1.0, 1.0]^\top$	$[0.1, 0.9]^\top$	$[0.8, 0.0]^\top$

- Message

$$c_2 = f_2 \circ c_1 + i_2 \circ \tilde{c}_2 \tag{51}$$
$$= [-0.8, 1.0]^\top \tag{52}$$
$$\tag{53}$$

- Hidden

$$h_2 = o_2 \circ \tanh(c_2) \tag{54}$$
$$= [1.0, 1.0]^\top \circ \tanh([-0.8, 1.0]^\top) \tag{55}$$
$$= [-0.7, 0.8]^\top \tag{56}$$

Forwards at time step 2

$\tilde{c}_2$	i_2	f_2	c_1
$[-0.9, 1.0]^\top$	$[1.0, 1.0]^\top$	$[0.1, 0.9]^\top$	$[0.8, 0.0]^\top$

- Message

$$c_2 = f_2 \circ c_1 + i_2 \circ \tilde{c}_2 \tag{51}$$
$$= [-0.8, 1.0]^\top \tag{52}$$
$$\tag{53}$$

- Hidden

$$h_2 = o_2 \circ \tanh(c_2) \tag{54}$$
$$= [-0.7, 0.8]^\top \tag{55}$$

- Output target$_2 = [0.0, 1.0]^\top$

Next time step . . .

- $i_3 = [0.4, 0.0]^\top$
- $f_3 = [0.4, 0.6]^\top$
- $o_3 = [0.5, 0.5]^\top$
- $\tilde{c}_3 = [-1.0, -0.6]^\top$
- $c_3 = [-0.7, 0.6]^\top$
- $h_3 = [-0.3, 0.3]^\top$

- Classify target$_3 = [0.0, 1.0]^\top$

What's going on?

- What's the classification?
- What inputs are important?
- When can things be forgotten?
- How would other sequences be classified?

Training

- The parameters of LSTM showed in this example are obtained by training with cross-entropy loss function: (T=3)

$$\sum_{i=1}^{N}\sum_{t=1}^{T} H(y_{it},\ \text{target}_{it})$$

  - 0: accumulated number of A at time t is no larger than 1
  - 1: accumulated number of A at time t is larger than 1
  - Converted to binary classification problem:

$$\text{target}_1 = [1.0, 0.0] \quad \text{target}_2 = [0.0, 1.0] \quad \text{target}_3 = [0.0, 1.0]$$