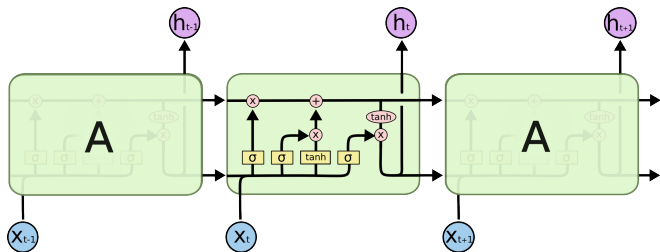# Long Short Term Memory Networks

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, B, A, B, A

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

- prediction output:

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

I don't want you to compute every cell

- Will get very boring quickly
- Goal is to understand big picture, not tiny details

Parameters that take x_t as input

Input Gate

$$W_{ii} = \begin{bmatrix} 30.00 & 0.00 \\ 0.00 & 0.00 \end{bmatrix}$$

$$b_{ii} = \begin{bmatrix} 0.00 \\ 0.00 \end{bmatrix}$$

Memory Cell

$$W_{ic} = \begin{bmatrix} 30.00 & 0.00 \\ 0.00 & 30.00 \end{bmatrix}$$

$$b_{ic} = \begin{bmatrix} 0.00 \\ 0.00 \end{bmatrix}$$

Forget Gate

$$W_{if} = \begin{bmatrix} 60.00 & 0.00 \\ 0.00 & 0.00 \end{bmatrix}$$

$$b_{if} = \begin{bmatrix} 0.00 \\ 0.00 \end{bmatrix}$$

Output Gate

$$W_{io} = \begin{bmatrix} 0.00 & 0.00 \\ 0.00 & 0.00 \end{bmatrix}$$

$$b_{io} = \begin{bmatrix} 30.00 \\ 30.00 \end{bmatrix}$$

Parameters that take h_{t-1} as input

Input Gate

$$W_{hi} = \begin{bmatrix} 0.00 & 0.00 \\ 60.00 & 0.00 \end{bmatrix}$$

$$b_{hi} = \begin{bmatrix} 0.00 \\ -30.00 \end{bmatrix}$$

Memory Cell

$$W_{hc} = \begin{bmatrix} 0.00 & 0.00 \\ 0.00 & 0.00 \end{bmatrix}$$

$$b_{hc} = \begin{bmatrix} 0.00 \\ 0.00 \end{bmatrix}$$

Forget Gate

$$W_{hf} = \begin{bmatrix} 0.00 & 0.00 \\ 0.00 & -30.00 \end{bmatrix}$$

$$b_{hf} = \begin{bmatrix} -30.00 \\ 0.00 \end{bmatrix}$$

Output Gate

$$W_{ho} = \begin{bmatrix} 0.00 & 0.00 \\ 0.00 & 0.00 \end{bmatrix}$$

$$b_{ho} = \begin{bmatrix} 0.00 \\ 0.00 \end{bmatrix}$$

What gate has been disabled?

What gate has been disabled?

- Output gate ignores x_t and h_{t-1} (all zero parameters)
- Only has positive bias term
- Thus will always be identity:

$$h_t = \tanh(c_t) \qquad (1)$$

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: A, A, B, B, A, B, A

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

Step 1 Task

What is c_1?

Input Gate at $t = 1$: i_1

$$W_{ii} = \begin{bmatrix} 30.00 & 0.00 \\ 0.00 & 0.00 \end{bmatrix} \quad b_{ii} = \begin{bmatrix} 0.00 \\ 0.00 \end{bmatrix} \qquad W_{hi} = \begin{bmatrix} 0.00 & 0.00 \\ 60.00 & 0.00 \end{bmatrix} \quad b_{hi} = \begin{bmatrix} 0.00 \\ -30.00 \end{bmatrix}$$

$$x^{(1)} = [1.00, 0.00]^\top \qquad\qquad h^{(0)} = [0.00, 0.00]^\top$$

Input Gate at $t = 1$: i_1

$$W_{ii} = \begin{bmatrix} 30.00 & 0.00 \\ 0.00 & 0.00 \end{bmatrix} \quad b_{ii} = \begin{bmatrix} 0.00 \\ 0.00 \end{bmatrix} \quad W_{hi} = \begin{bmatrix} 0.00 & 0.00 \\ 60.00 & 0.00 \end{bmatrix} \quad b_{hi} = \begin{bmatrix} 0.00 \\ -30.00 \end{bmatrix}$$

$$x^{(1)} = [1.00, 0.00]^\top \qquad\qquad h^{(0)} = [0.00, 0.00]^\top$$

$$i^{(1)} = \sigma(W_{ii}x^{(1)} + b_{ii} + W_{hi}h^{(0)} + b_{hi}) \tag{2}$$
$$= \sigma([30.00, -30.00]^\top) \tag{3}$$
$$= [1.00, 0.00]^\top \tag{4}$$

Forget Gate at $t = 1$: $f^{(1)}$

$$W_{if} = \begin{bmatrix} 60.00 & 0.00 \\ 0.00 & 0.00 \end{bmatrix} \quad b_{if} = \begin{bmatrix} 0.00 \\ 0.00 \end{bmatrix} \quad W_{hf} = \begin{bmatrix} 0.00 & 0.00 \\ 0.00 & -30.00 \end{bmatrix} b_{hf} = \begin{bmatrix} -30.00 \\ 0.00 \end{bmatrix}$$

$$x^{(1)} = [1.00, 0.00]^{\top} \qquad h^{(0)} = [0.00, 0.00]^{\top}$$

Forget Gate at $t = 1$: $f^{(1)}$

$$W_{if} = \begin{bmatrix} 60.00 & 0.00 \\ 0.00 & 0.00 \end{bmatrix} \quad b_{if} = \begin{bmatrix} 0.00 \\ 0.00 \end{bmatrix} \quad W_{hf} = \begin{bmatrix} 0.00 & 0.00 \\ 0.00 & -30.00 \end{bmatrix} b_{hf} = \begin{bmatrix} -30.00 \\ 0.00 \end{bmatrix}$$

$$x^{(1)} = [1.00, 0.00]^\top \qquad\qquad h^{(0)} = [0.00, 0.00]^\top$$

$$f^{(1)} = \sigma\left(W_{if}x^{(1)} + b_{if} + W_{hf}h^{(0)} + b_{hf}\right) \tag{5}$$

$$= \sigma\left([30.00, 0.00]^\top\right) \tag{6}$$

$$= [1.00, 0.50]^\top \tag{7}$$

Output Gate at $t = 1$: $o^{(1)}$

$$W_{io} = \begin{bmatrix} 0.00 & 0.00 \\ 0.00 & 0.00 \end{bmatrix} \quad b_{io} = \begin{bmatrix} 30.00 \\ 30.00 \end{bmatrix} \quad W_{ho} = \begin{bmatrix} 0.00 & 0.00 \\ 0.00 & 0.00 \end{bmatrix} \quad b_{ho} = \begin{bmatrix} 0.00 \\ 0.00 \end{bmatrix}$$

$$x^{(1)} = [1.00, 0.00]^\top \qquad\qquad h^{(0)} = [0.00, 0.00]^\top$$

Output Gate at $t = 1$: $o^{(1)}$

$$W_{io} = \begin{bmatrix} 0.00 & 0.00 \\ 0.00 & 0.00 \end{bmatrix} \quad b_{io} = \begin{bmatrix} 30.00 \\ 30.00 \end{bmatrix} \quad W_{ho} = \begin{bmatrix} 0.00 & 0.00 \\ 0.00 & 0.00 \end{bmatrix} \quad b_{ho} = \begin{bmatrix} 0.00 \\ 0.00 \end{bmatrix}$$

$$x^{(1)} = [1.00, 0.00]^\top \qquad\qquad h^{(0)} = [0.00, 0.00]^\top$$

$$\begin{align}
o^{(1)} &= \sigma(W_{io}x^{(1)} + b_{io} + W_{ho}h^{(0)} + b_{ho}) \tag{8} \\
&= \sigma([30.00, 30.00]^\top) \tag{9} \\
&= [1.00, 1.00]^\top \tag{10}
\end{align}$$

Memory Contribution at $t = 1$: $\tilde{c}^{(1)}$

$$W_{i\tilde{c}} = \begin{bmatrix} 30.00 & 0.00 \\ 0.00 & 30.00 \end{bmatrix} \quad b_{i\tilde{c}} = \begin{bmatrix} 0.00 \\ 0.00 \end{bmatrix} \qquad W_{h\tilde{c}} = \begin{bmatrix} 0.00 & 0.00 \\ 0.00 & 0.00 \end{bmatrix} \quad b_{h\tilde{c}} = \begin{bmatrix} 0.00 \\ 0.00 \end{bmatrix}$$

$$x^{(1)} = [1.00, 0.00]^{\top} \qquad\qquad h^{(0)} = [0.00, 0.00]^{\top}$$

Memory Contribution at $t = 1$: $\tilde{c}^{(1)}$

$$W_{i\tilde{c}} = \begin{bmatrix} 30.00 & 0.00 \\ 0.00 & 30.00 \end{bmatrix} b_{i\tilde{c}} = \begin{bmatrix} 0.00 \\ 0.00 \end{bmatrix} \qquad W_{h\tilde{c}} = \begin{bmatrix} 0.00 & 0.00 \\ 0.00 & 0.00 \end{bmatrix} b_{h\tilde{c}} = \begin{bmatrix} 0.00 \\ 0.00 \end{bmatrix}$$

$$x^{(1)} = [1.00, 0.00]^\top \qquad\qquad h^{(0)} = [0.00, 0.00]^\top$$

$$\tilde{c}^{(1)} = \tanh(W_{i\tilde{c}} x^{(1)} + b_{i\tilde{c}} + W_{h\tilde{c}} h^{(0)} + b_{h\tilde{c}}) \tag{11}$$

$$= \tanh([30.00, 0.00]^\top) \tag{12}$$

$$= [1.00, 0.00]^\top \tag{13}$$

Forward message at time step 1

f_1	c_0	i_1	$\tilde{c}_1$
$[1.00, 0.50]^\top$	$[0.00, 0.00]^\top$	$[1.00, 0.00]^\top$	$[1.00, 0.00]^\top$

- Message forward (c_1)

$$c_1 = f_1 \circ c_0 + i_1 \circ \tilde{c}_1 \tag{14}$$

$$\tag{15}$$

Forward message at time step 1

f_1	c_0	i_1	$\tilde{c}_1$
$[1.00, 0.50]^\top$	$[0.00, 0.00]^\top$	$[1.00, 0.00]^\top$	$[1.00, 0.00]^\top$

- Message forward (c_1)

$$c_1 = f_1 \circ c_0 + i_1 \circ \tilde{c}_1 \tag{14}$$
$$= [1.00, 0.50]^\top \circ [0.00, 0.00]^\top + [1.00, 0.00]^\top \circ [1.00, 0.00]^\top \tag{15}$$
$$\tag{16}$$

Forward message at time step 1

f_1	c_0	i_1	$\tilde{c}_1$
$[1.00, 0.50]^\top$	$[0.00, 0.00]^\top$	$[1.00, 0.00]^\top$	$[1.00, 0.00]^\top$

- Message forward (c_1)

$$c_1 = f_1 \circ c_0 + i_1 \circ \tilde{c}_1 \tag{14}$$

$$= [1.00, 0.50]^\top \circ [0.00, 0.00]^\top + [1.00, 0.00]^\top \circ [1.00, 0.00]^\top \tag{15}$$

$$= [1.00, 0.00]^\top \tag{16}$$

Forward message at time step 1

f_1	c_0	i_1	$\tilde{c}_1$
$[1.00, 0.50]^\top$	$[0.00, 0.00]^\top$	$[1.00, 0.00]^\top$	$[1.00, 0.00]^\top$

- Message forward (c_1)

$$c_1 = [1.00, 0.00]^\top \tag{14}$$

- New hidden (h_1)

$$h_1 \tag{15}$$

Forward message at time step 1

f_1	c_0	i_1	$\tilde{c}_1$
$[1.00, 0.50]^\top$	$[0.00, 0.00]^\top$	$[1.00, 0.00]^\top$	$[1.00, 0.00]^\top$

- Message forward (c_1)

$$c_1 = [1.00, 0.00]^\top \tag{14}$$

- New hidden (h_1)

$$h_1 = o_1 \circ \tanh(c_1) \tag{15}$$

$$\tag{16}$$

Forward message at time step 1

f_1	c_0	i_1	$\tilde{c}_1$
$[1.00, 0.50]^\top$	$[0.00, 0.00]^\top$	$[1.00, 0.00]^\top$	$[1.00, 0.00]^\top$

- Message forward (c_1)

$$c_1 = [1.00, 0.00]^\top \tag{14}$$

- New hidden (h_1)

$$h_1 = o_1 \circ \tanh(c_1) \tag{15}$$
$$= [1.00, 1.00]^\top \circ \tanh([1.00, 0.00]^\top) \tag{16}$$
$$\tag{17}$$

Forward message at time step 1

f_1	c_0	i_1	$\tilde{c}_1$
$[1.00, 0.50]^\top$	$[0.00, 0.00]^\top$	$[1.00, 0.00]^\top$	$[1.00, 0.00]^\top$

- Message forward (c_1)

$$c_1 = [1.00, 0.00]^\top \tag{14}$$

- New hidden (h_1)

$$h_1 = o_1 \circ \tanh(c_1) \tag{15}$$
$$= [1.00, 1.00]^\top \circ \tanh([1.00, 0.00]^\top) \tag{16}$$
$$= [0.76, 0.00]^\top \tag{17}$$

Forward message at time step 1

f_1	c_0	i_1	$\tilde{c}_1$
$[1.00, 0.50]^\top$	$[0.00, 0.00]^\top$	$[1.00, 0.00]^\top$	$[1.00, 0.00]^\top$

- Message forward (c_1)
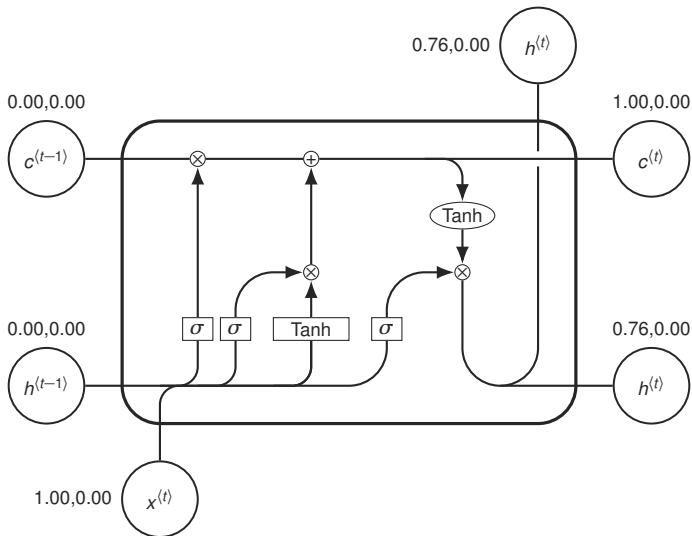
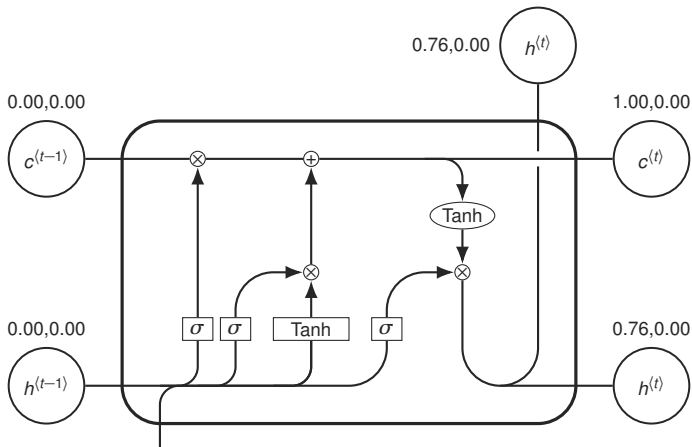$$c_1 = [1.00, 0.00]^\top \qquad (14)$$

- New hidden (h_1)

$$h_1 = [0.76, 0.00]^\top \qquad (15)$$

- Prediction $y_1 = \mathrm{softmax}(h_1) = 0$

Summary at $t = 1$

Summary at $t = 1$

Step 1 Task

What is c_2? This is easier to compute because you figure most of it out already? What will be different from the previous time step?

Step 1 Task

What is c_2? This is easier to compute because you figure most of it out already? What will be different from the previous time step?

- i_2 (because hidden state was different)
- c_2

Input Gate at $t = 2$: i_1

$$W_{ii} = \begin{bmatrix} 30.00 & 0.00 \\ 0.00 & 0.00 \end{bmatrix} \quad b_{ii} = \begin{bmatrix} 0.00 \\ 0.00 \end{bmatrix} \qquad W_{hi} = \begin{bmatrix} 0.00 & 0.00 \\ 60.00 & 0.00 \end{bmatrix} \quad b_{hi} = \begin{bmatrix} 0.00 \\ -30.00 \end{bmatrix}$$

$$x^{(2)} = [1.00, 0.00]^\top \qquad\qquad h^{(1)} = [0.76, 0.00]^\top$$

$$W_{ii} = \begin{bmatrix} 30.00 & 0.00 \\ 0.00 & 0.00 \end{bmatrix} \quad b_{ii} = \begin{bmatrix} 0.00 \\ 0.00 \end{bmatrix} \quad W_{hi} = \begin{bmatrix} 0.00 & 0.00 \\ 60.00 & 0.00 \end{bmatrix} \quad b_{hi} = \begin{bmatrix} 0.00 \\ -30.00 \end{bmatrix}$$

$$x^{(2)} = [1.00, 0.00]^\top \qquad\qquad h^{(1)} = [0.76, 0.00]^\top$$

$$i^{(2)} = \sigma(W_{ii}x^{(2)} + b_{ii} + W_{hi}h^{(1)} + b_{hi}) \tag{16}$$
$$= \sigma([30.00, 15.70]^\top) \tag{17}$$
$$= [1.00, 1.00]^\top \tag{18}$$

Forget Gate at $t = 2$: $f^{(2)}$

$$W_{if} = \begin{bmatrix} 60.00 & 0.00 \\ 0.00 & 0.00 \end{bmatrix} \quad b_{if} = \begin{bmatrix} 0.00 \\ 0.00 \end{bmatrix} \quad W_{hf} = \begin{bmatrix} 0.00 & 0.00 \\ 0.00 & -30.00 \end{bmatrix} \quad b_{hf} = \begin{bmatrix} -30.00 \\ 0.00 \end{bmatrix}$$

$$x^{(2)} = [1.00, 0.00]^\top \qquad\qquad h^{(1)} = [0.76, 0.00]^\top$$

Forget Gate at $t = 2$: $f^{(2)}$

$$W_{if} = \begin{bmatrix} 60.00 & 0.00 \\ 0.00 & 0.00 \end{bmatrix} \quad b_{if} = \begin{bmatrix} 0.00 \\ 0.00 \end{bmatrix} \quad W_{hf} = \begin{bmatrix} 0.00 & 0.00 \\ 0.00 & -30.00 \end{bmatrix} b_{hf} = \begin{bmatrix} -30.00 \\ 0.00 \end{bmatrix}$$

$$x^{(2)} = [1.00, 0.00]^{\top} \qquad h^{(1)} = [0.76, 0.00]^{\top}$$

$$f^{(2)} = \sigma(W_{if}x^{(2)} + b_{if} + W_{hf}h^{(1)} + b_{hf}) \tag{19}$$

$$= \sigma([30.00, 0.00]^{\top}) \tag{20}$$

$$= [1.00, 0.50]^{\top} \tag{21}$$

Output Gate at $t = 2$: $o^{(2)}$

$$W_{io} = \begin{bmatrix} 0.00 & 0.00 \\ 0.00 & 0.00 \end{bmatrix} \quad b_{io} = \begin{bmatrix} 30.00 \\ 30.00 \end{bmatrix} \quad W_{ho} = \begin{bmatrix} 0.00 & 0.00 \\ 0.00 & 0.00 \end{bmatrix} \quad b_{ho} = \begin{bmatrix} 0.00 \\ 0.00 \end{bmatrix}$$

$$x^{(2)} = [1.00, 0.00]^{\top} \qquad\qquad h^{(1)} = [0.76, 0.00]^{\top}$$

Output Gate at $t = 2$: $o^{(2)}$

$$W_{io} = \begin{bmatrix} 0.00 & 0.00 \\ 0.00 & 0.00 \end{bmatrix} \quad b_{io} = \begin{bmatrix} 30.00 \\ 30.00 \end{bmatrix} \quad W_{ho} = \begin{bmatrix} 0.00 & 0.00 \\ 0.00 & 0.00 \end{bmatrix} \quad b_{ho} = \begin{bmatrix} 0.00 \\ 0.00 \end{bmatrix}$$

$$x^{(2)} = [1.00, 0.00]^\top \qquad\qquad h^{(1)} = [0.76, 0.00]^\top$$

$$o^{(2)} = \sigma(W_{io} x^{(2)} + b_{io} + W_{ho} h^{(1)} + b_{ho}) \tag{22}$$

$$= \sigma([30.00, 30.00]^\top) \tag{23}$$

$$= [1.00, 1.00]^\top \tag{24}$$

Memory Contribution at $t = 2$: $\tilde{c}^{(2)}$

$$W_{i\tilde{c}} = \begin{bmatrix} 30.00 & 0.00 \\ 0.00 & 30.00 \end{bmatrix} \quad b_{i\tilde{c}} = \begin{bmatrix} 0.00 \\ 0.00 \end{bmatrix} \qquad W_{h\tilde{c}} = \begin{bmatrix} 0.00 & 0.00 \\ 0.00 & 0.00 \end{bmatrix} \quad b_{h\tilde{c}} = \begin{bmatrix} 0.00 \\ 0.00 \end{bmatrix}$$

$$x^{(2)} = [1.00, 0.00]^\top \qquad\qquad h^{(1)} = [0.76, 0.00]^\top$$

$$W_{i\tilde{c}} = \begin{bmatrix} 30.00 & 0.00 \\ 0.00 & 30.00 \end{bmatrix} \quad b_{i\tilde{c}} = \begin{bmatrix} 0.00 \\ 0.00 \end{bmatrix} \qquad W_{h\tilde{c}} = \begin{bmatrix} 0.00 & 0.00 \\ 0.00 & 0.00 \end{bmatrix} \quad b_{h\tilde{c}} = \begin{bmatrix} 0.00 \\ 0.00 \end{bmatrix}$$

$$x^{(2)} = [1.00, 0.00]^{\top} \qquad\qquad h^{(1)} = [0.76, 0.00]^{\top}$$

$$\tilde{c}^{(2)} = \tanh(W_{i\tilde{c}}x^{(2)} + b_{i\tilde{c}} + W_{h\tilde{c}}h^{(1)} + b_{h\tilde{c}}) \tag{25}$$

$$= \tanh([30.00, 0.00]^{\top}) \tag{26}$$

$$= [1.00, 0.00]^{\top} \tag{27}$$

Forward message at time step 2

f_2	c_1	i_2	$\tilde{c}_2$
$[1.00, 0.50]^\top$	$[1.00, 0.00]^\top$	$[1.00, 1.00]^\top$	$[1.00, 0.00]^\top$

- Message forward (c_2)

$$c_2 = f_2 \circ c_1 + i_2 \circ \tilde{c}_2 \qquad (28)$$

$$\qquad (29)$$

Forward message at time step 2

f_2	c_1	i_2	$\tilde{c}_2$
$[1.00, 0.50]^\top$	$[1.00, 0.00]^\top$	$[1.00, 1.00]^\top$	$[1.00, 0.00]^\top$

- Message forward (c_2)

$$c_2 = f_2 \circ c_1 + i_2 \circ \tilde{c}_2 \tag{28}$$
$$= [1.00, 0.50]^\top \circ [1.00, 0.00]^\top + [1.00, 1.00]^\top \circ [1.00, 0.00]^\top \tag{29}$$
$$\tag{30}$$

Forward message at time step 2

f_2	c_1	i_2	$\tilde{c}_2$
$[1.00, 0.50]^\top$	$[1.00, 0.00]^\top$	$[1.00, 1.00]^\top$	$[1.00, 0.00]^\top$

- Message forward (c_2)

$$c_2 = f_2 \circ c_1 + i_2 \circ \tilde{c}_2 \tag{28}$$
$$= [1.00, 0.50]^\top \circ [1.00, 0.00]^\top + [1.00, 1.00]^\top \circ [1.00, 0.00]^\top \tag{29}$$
$$= [2.00, 0.00]^\top \tag{30}$$

Forward message at time step 2

f_2	c_1	i_2	$\tilde{c}_2$
$[1.00, 0.50]^\top$	$[1.00, 0.00]^\top$	$[1.00, 1.00]^\top$	$[1.00, 0.00]^\top$

- Message forward (c_2)

$$c_2 = [2.00, 0.00]^\top \tag{28}$$

- New hidden (h_2)

$$h_2 \tag{29}$$

f_2	c_1	i_2	$\tilde{c}_2$
$[1.00, 0.50]^\top$	$[1.00, 0.00]^\top$	$[1.00, 1.00]^\top$	$[1.00, 0.00]^\top$

- Message forward (c_2)

$$c_2 = [2.00, 0.00]^\top \tag{28}$$

- New hidden (h_2)

$$h_2 = o_2 \circ \tanh(c_2) \tag{29}$$

$$\tag{30}$$

Forward message at time step 2

f_2	c_1	i_2	$\tilde{c}_2$
$[1.00, 0.50]^\top$	$[1.00, 0.00]^\top$	$[1.00, 1.00]^\top$	$[1.00, 0.00]^\top$

- Message forward (c_2)

$$c_2 = [2.00, 0.00]^\top \tag{28}$$

- New hidden (h_2)

$$h_2 = o_2 \circ \tanh(c_2) \tag{29}$$

$$= [1.00, 1.00]^\top \circ \tanh([2.00, 0.00]^\top) \tag{30}$$

$$\tag{31}$$

Forward message at time step 2

f_2	c_1	i_2	$\tilde{c}_2$
$[1.00, 0.50]^\top$	$[1.00, 0.00]^\top$	$[1.00, 1.00]^\top$	$[1.00, 0.00]^\top$

- Message forward (c_2)

$$c_2 = [2.00, 0.00]^\top \tag{28}$$

- New hidden (h_2)

$$
\begin{align}
h_2 &= o_2 \circ \tanh(c_2) \tag{29} \\
&= [1.00, 1.00]^\top \circ \tanh([2.00, 0.00]^\top) \tag{30} \\
&= [0.96, 0.00]^\top \tag{31}
\end{align}
$$

f_2	c_1	i_2	$\tilde{c}_2$
$[1.00, 0.50]^\top$	$[1.00, 0.00]^\top$	$[1.00, 1.00]^\top$	$[1.00, 0.00]^\top$

- Message forward (c_2)

$$c_2 = [2.00, 0.00]^\top \tag{28}$$
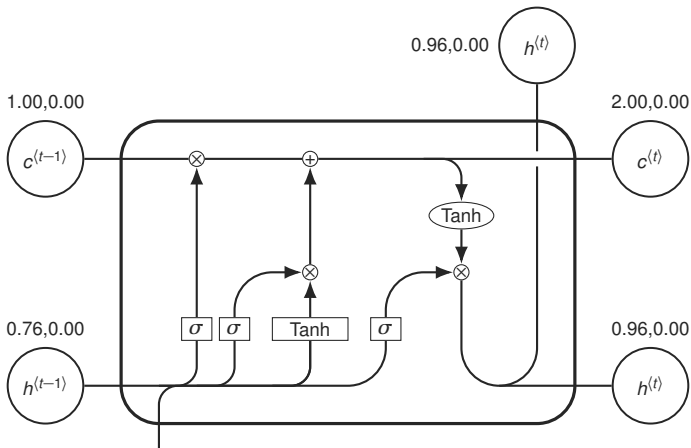
- New hidden (h_2)

$$h_2 = [0.96, 0.00]^\top \tag{29}$$

- Prediction $y_2 = \mathrm{softmax}(h_2) = 0$

Summary at $t = 2$
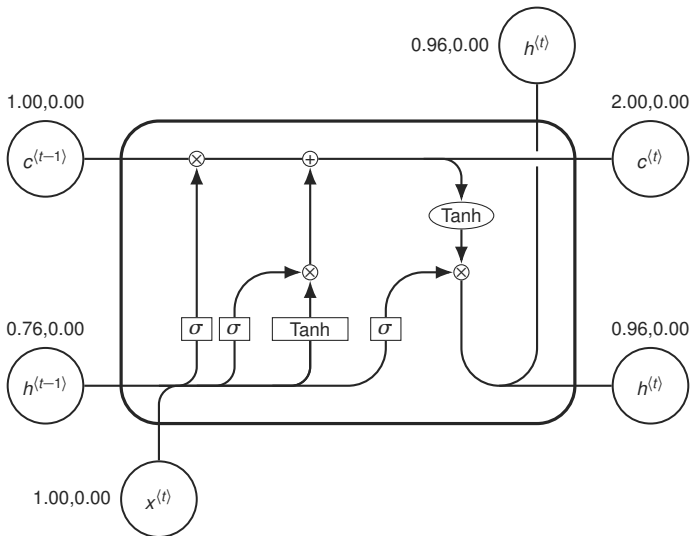
Summary at $t = 2$

After another A as input, message forward becomes (2, 0)

What happens now?

- $\tilde{c}_3$ is very different, as is f_3
- Input looks a little different

Input Gate at $t = 3$: i_1

$$W_{ii} = \begin{bmatrix} 30.00 & 0.00 \\ 0.00 & 0.00 \end{bmatrix} \quad b_{ii} = \begin{bmatrix} 0.00 \\ 0.00 \end{bmatrix} \qquad W_{hi} = \begin{bmatrix} 0.00 & 0.00 \\ 60.00 & 0.00 \end{bmatrix} \quad b_{hi} = \begin{bmatrix} 0.00 \\ -30.00 \end{bmatrix}$$

$$x^{(3)} = [0.00, 1.00]^\top \qquad\qquad h^{(2)} = [0.96, 0.00]^\top$$

Input Gate at $t = 3$: i_1

$$W_{ii} = \begin{bmatrix} 30.00 & 0.00 \\ 0.00 & 0.00 \end{bmatrix} \quad b_{ii} = \begin{bmatrix} 0.00 \\ 0.00 \end{bmatrix} \quad W_{hi} = \begin{bmatrix} 0.00 & 0.00 \\ 60.00 & 0.00 \end{bmatrix} b_{hi} = \begin{bmatrix} 0.00 \\ -30.00 \end{bmatrix}$$

$$x^{(3)} = [0.00, 1.00]^\top \qquad\qquad h^{(2)} = [0.96, 0.00]^\top$$

$$i^{(3)} = \sigma(W_{ii}x^{(3)} + b_{ii} + W_{hi}h^{(2)} + b_{hi}) \tag{30}$$
$$= \sigma([0.00, 27.84]^\top) \tag{31}$$
$$= [0.50, 1.00]^\top \tag{32}$$

Forget Gate at $t = 3$: $f^{(3)}$

$$W_{if} = \begin{bmatrix} 60.00 & 0.00 \\ 0.00 & 0.00 \end{bmatrix} \quad b_{if} = \begin{bmatrix} 0.00 \\ 0.00 \end{bmatrix} \qquad W_{hf} = \begin{bmatrix} 0.00 & 0.00 \\ 0.00 & -30.00 \end{bmatrix} \quad b_{hf} = \begin{bmatrix} -30.00 \\ 0.00 \end{bmatrix}$$

$$x^{(3)} = [0.00, 1.00]^\top \qquad\qquad h^{(2)} = [0.96, 0.00]^\top$$

Forget Gate at $t = 3$: $f^{(3)}$

$$W_{if} = \begin{bmatrix} 60.00 & 0.00 \\ 0.00 & 0.00 \end{bmatrix} \quad b_{if} = \begin{bmatrix} 0.00 \\ 0.00 \end{bmatrix} \qquad W_{hf} = \begin{bmatrix} 0.00 & 0.00 \\ 0.00 & -30.00 \end{bmatrix} b_{hf} = \begin{bmatrix} -30.00 \\ 0.00 \end{bmatrix}$$

$$x^{(3)} = [0.00, 1.00]^\top \qquad\qquad h^{(2)} = [0.96, 0.00]^\top$$

$$f^{(3)} = \sigma\left(W_{if}x^{(3)} + b_{if} + W_{hf}h^{(2)} + b_{hf}\right) \tag{33}$$

$$= \sigma\left([-30.00, 0.00]^\top\right) \tag{34}$$

$$= [0.00, 0.50]^\top \tag{35}$$

Output Gate at $t = 3$: $o^{(3)}$

$$W_{io} = \begin{bmatrix} 0.00 & 0.00 \\ 0.00 & 0.00 \end{bmatrix} \quad b_{io} = \begin{bmatrix} 30.00 \\ 30.00 \end{bmatrix} \quad W_{ho} = \begin{bmatrix} 0.00 & 0.00 \\ 0.00 & 0.00 \end{bmatrix} \quad b_{ho} = \begin{bmatrix} 0.00 \\ 0.00 \end{bmatrix}$$

$$x^{(3)} = [0.00, 1.00]^\top \qquad h^{(2)} = [0.96, 0.00]^\top$$

Output Gate at $t = 3$: $o^{(3)}$

$$W_{io} = \begin{bmatrix} 0.00 & 0.00 \\ 0.00 & 0.00 \end{bmatrix} \quad b_{io} = \begin{bmatrix} 30.00 \\ 30.00 \end{bmatrix} \quad W_{ho} = \begin{bmatrix} 0.00 & 0.00 \\ 0.00 & 0.00 \end{bmatrix} \quad b_{ho} = \begin{bmatrix} 0.00 \\ 0.00 \end{bmatrix}$$

$$x^{(3)} = [0.00, 1.00]^\top \qquad\qquad h^{(2)} = [0.96, 0.00]^\top$$

$$o^{(3)} = \sigma(W_{io}x^{(3)} + b_{io} + W_{ho}h^{(2)} + b_{ho}) \tag{36}$$
$$= \sigma([30.00, 30.00]^\top) \tag{37}$$
$$= [1.00, 1.00]^\top \tag{38}$$

Memory Contribution at $t = 3$: $\tilde{c}^{(3)}$

$$W_{i\tilde{c}} = \begin{bmatrix} 30.00 & 0.00 \\ 0.00 & 30.00 \end{bmatrix} \quad b_{i\tilde{c}} = \begin{bmatrix} 0.00 \\ 0.00 \end{bmatrix}$$

$$x^{(3)} = [0.00, 1.00]^\top$$

$$W_{h\tilde{c}} = \begin{bmatrix} 0.00 & 0.00 \\ 0.00 & 0.00 \end{bmatrix} \quad b_{h\tilde{c}} = \begin{bmatrix} 0.00 \\ 0.00 \end{bmatrix}$$

$$h^{(2)} = [0.96, 0.00]^\top$$

$$W_{i\tilde{c}} = \begin{bmatrix} 30.00 & 0.00 \\ 0.00 & 30.00 \end{bmatrix} \; b_{i\tilde{c}} = \begin{bmatrix} 0.00 \\ 0.00 \end{bmatrix} \qquad W_{h\tilde{c}} = \begin{bmatrix} 0.00 & 0.00 \\ 0.00 & 0.00 \end{bmatrix} \; b_{h\tilde{c}} = \begin{bmatrix} 0.00 \\ 0.00 \end{bmatrix}$$

$$x^{(3)} = [0.00, 1.00]^{\top} \qquad\qquad\qquad h^{(2)} = [0.96, 0.00]^{\top}$$

$$\tilde{c}^{(3)} = \tanh(W_{i\tilde{c}} x^{(3)} + b_{i\tilde{c}} + W_{h\tilde{c}} h^{(2)} + b_{h\tilde{c}}) \tag{39}$$
$$= \tanh([0.00, 30.00]^{\top}) \tag{40}$$
$$= [0.00, 1.00]^{\top} \tag{41}$$

Forward message at time step 3

f_3	c_2	i_3	$\tilde{c}_3$
$[0.00, 0.50]^\top$	$[2.00, 0.00]^\top$	$[0.50, 1.00]^\top$	$[0.00, 1.00]^\top$

- Message forward (c_3)

$$c_3 = f_3 \circ c_2 + i_3 \circ \tilde{c}_3 \tag{42}$$

$$\tag{43}$$

Forward message at time step 3

f_3	c_2	i_3	$\tilde{c}_3$
$[0.00, 0.50]^\top$	$[2.00, 0.00]^\top$	$[0.50, 1.00]^\top$	$[0.00, 1.00]^\top$

- Message forward (c_3)

$$c_3 = f_3 \circ c_2 + i_3 \circ \tilde{c}_3 \tag{42}$$
$$= [0.00, 0.50]^\top \circ [2.00, 0.00]^\top + [0.50, 1.00]^\top \circ [0.00, 1.00]^\top \tag{43}$$
$$\tag{44}$$

28

f_3	c_2	i_3	$\tilde{c}_3$
$[0.00, 0.50]^\top$	$[2.00, 0.00]^\top$	$[0.50, 1.00]^\top$	$[0.00, 1.00]^\top$

- Message forward (c_3)

$$c_3 = f_3 \circ c_2 + i_3 \circ \tilde{c}_3 \tag{42}$$
$$= [0.00, 0.50]^\top \circ [2.00, 0.00]^\top + [0.50, 1.00]^\top \circ [0.00, 1.00]^\top \tag{43}$$
$$= [0.00, 1.00]^\top \tag{44}$$

Forward message at time step 3

f_3	c_2	i_3	$\tilde{c}_3$
$[0.00, 0.50]^\top$	$[2.00, 0.00]^\top$	$[0.50, 1.00]^\top$	$[0.00, 1.00]^\top$

- Message forward (c_3)

$$c_3 = [0.00, 1.00]^\top \qquad (42)$$

- New hidden (h_3)

$$h_3 \qquad (43)$$

Forward message at time step 3

f_3	c_2	i_3	$\tilde{c}_3$
$[0.00, 0.50]^\top$	$[2.00, 0.00]^\top$	$[0.50, 1.00]^\top$	$[0.00, 1.00]^\top$

- Message forward (c_3)

$$c_3 = [0.00, 1.00]^\top \tag{42}$$

- New hidden (h_3)

$$h_3 = o_3 \circ \tanh(c_3) \tag{43}$$

$$\tag{44}$$

Forward message at time step 3

f_3	c_2	i_3	$\tilde{c}_3$
$[0.00, 0.50]^\top$	$[2.00, 0.00]^\top$	$[0.50, 1.00]^\top$	$[0.00, 1.00]^\top$

- Message forward (c_3)

$$c_3 = [0.00, 1.00]^\top \tag{42}$$

- New hidden (h_3)

$$h_3 = o_3 \circ \tanh(c_3) \tag{43}$$
$$= [1.00, 1.00]^\top \circ \tanh([0.00, 1.00]^\top) \tag{44}$$
$$\tag{45}$$

Forward message at time step 3

f_3	c_2	i_3	$\tilde{c}_3$
$[0.00, 0.50]^\top$	$[2.00, 0.00]^\top$	$[0.50, 1.00]^\top$	$[0.00, 1.00]^\top$

- Message forward (c_3)

$$c_3 = [0.00, 1.00]^\top \tag{42}$$

- New hidden (h_3)

$$h_3 = o_3 \circ \tanh(c_3) \tag{43}$$
$$= [1.00, 1.00]^\top \circ \tanh([0.00, 1.00]^\top) \tag{44}$$
$$= [0.00, 0.76]^\top \tag{45}$$

Forward message at time step 3

f_3	c_2	i_3	$\tilde{c}_3$
$[0.00, 0.50]^\top$	$[2.00, 0.00]^\top$	$[0.50, 1.00]^\top$	$[0.00, 1.00]^\top$

- Message forward (c_3)
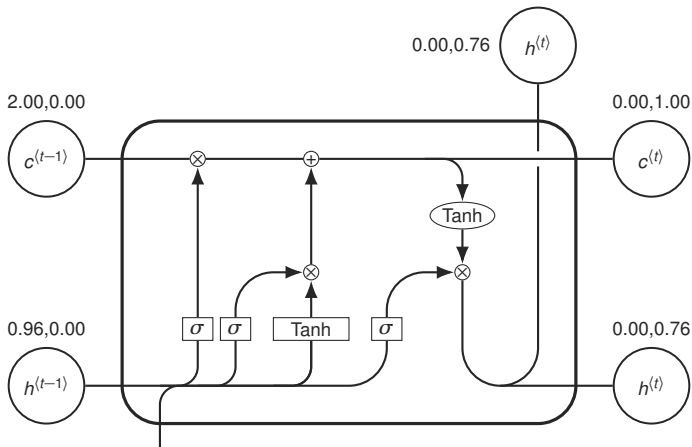
$$c_3 = [0.00, 1.00]^\top \tag{42}$$

- New hidden (h_3)

$$h_3 = [0.00, 0.76]^\top \tag{43}$$

- Prediction $y_3 = \mathrm{softmax}(h_3) = 1$
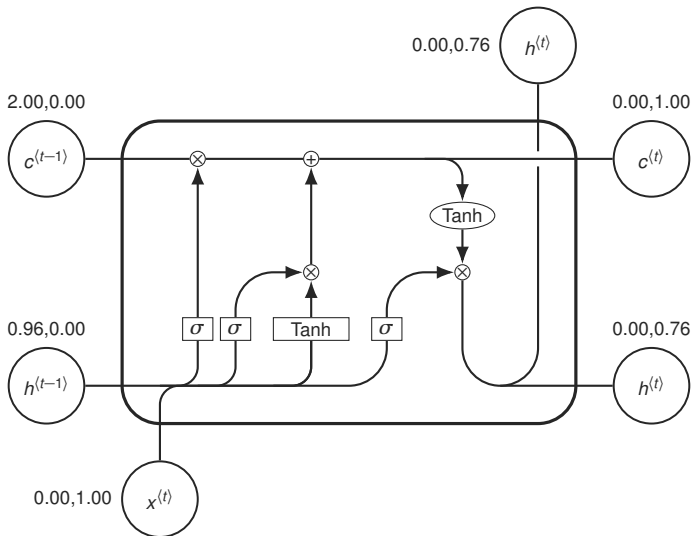
Summary at $t = 3$

But after we get a B, the message forward flips to (0, 1)

$$W_{ii} = \begin{bmatrix} 30.00 & 0.00 \\ 0.00 & 0.00 \end{bmatrix} \quad b_{ii} = \begin{bmatrix} 0.00 \\ 0.00 \end{bmatrix} \qquad W_{hi} = \begin{bmatrix} 0.00 & 0.00 \\ 60.00 & 0.00 \end{bmatrix} \quad b_{hi} = \begin{bmatrix} 0.00 \\ -30.00 \end{bmatrix}$$

$$x^{(4)} = [0.00, 1.00]^{\top} \qquad\qquad h^{(3)} = [0.00, 0.76]^{\top}$$

Input Gate at $t = 4$: i_1

$$W_{ii} = \begin{bmatrix} 30.00 & 0.00 \\ 0.00 & 0.00 \end{bmatrix} \quad b_{ii} = \begin{bmatrix} 0.00 \\ 0.00 \end{bmatrix} \quad W_{hi} = \begin{bmatrix} 0.00 & 0.00 \\ 60.00 & 0.00 \end{bmatrix} b_{hi} = \begin{bmatrix} 0.00 \\ -30.00 \end{bmatrix}$$

$$x^{(4)} = [0.00, 1.00]^\top \qquad\qquad h^{(3)} = [0.00, 0.76]^\top$$

$$i^{(4)} = \sigma(W_{ii}x^{(4)} + b_{ii} + W_{hi}h^{(3)} + b_{hi}) \tag{44}$$
$$= \sigma([0.00, -30.00]^\top) \tag{45}$$
$$= [0.50, 0.00]^\top \tag{46}$$

Forget Gate at $t = 4$: $f^{(4)}$

$$W_{if} = \begin{bmatrix} 60.00 & 0.00 \\ 0.00 & 0.00 \end{bmatrix} \quad b_{if} = \begin{bmatrix} 0.00 \\ 0.00 \end{bmatrix} \qquad W_{hf} = \begin{bmatrix} 0.00 & 0.00 \\ 0.00 & -30.00 \end{bmatrix} b_{hf} = \begin{bmatrix} -30.00 \\ 0.00 \end{bmatrix}$$

$$x^{(4)} = [0.00, 1.00]^{\top} \qquad\qquad h^{(3)} = [0.00, 0.76]^{\top}$$

Forget Gate at $t = 4$: $f^{(4)}$

$$W_{if} = \begin{bmatrix} 60.00 & 0.00 \\ 0.00 & 0.00 \end{bmatrix} \quad b_{if} = \begin{bmatrix} 0.00 \\ 0.00 \end{bmatrix} \quad W_{hf} = \begin{bmatrix} 0.00 & 0.00 \\ 0.00 & -30.00 \end{bmatrix} b_{hf} = \begin{bmatrix} -30.00 \\ 0.00 \end{bmatrix}$$

$$x^{(4)} = [0.00, 1.00]^\top \qquad\qquad h^{(3)} = [0.00, 0.76]^\top$$

$$f^{(4)} = \sigma\left(W_{if}x^{(4)} + b_{if} + W_{hf}h^{(3)} + b_{hf}\right) \qquad (47)$$

$$= \sigma\left([-30.00, -22.85]^\top\right) \qquad (48)$$

$$= [0.00, 0.00]^\top \qquad (49)$$

$$W_{io} = \begin{bmatrix} 0.00 & 0.00 \\ 0.00 & 0.00 \end{bmatrix} \quad b_{io} = \begin{bmatrix} 30.00 \\ 30.00 \end{bmatrix} \qquad W_{ho} = \begin{bmatrix} 0.00 & 0.00 \\ 0.00 & 0.00 \end{bmatrix} \quad b_{ho} = \begin{bmatrix} 0.00 \\ 0.00 \end{bmatrix}$$

$$x^{(4)} = [0.00, 1.00]^\top \qquad\qquad h^{(3)} = [0.00, 0.76]^\top$$

Output Gate at $t = 4$: $o^{(4)}$

$$W_{io} = \begin{bmatrix} 0.00 & 0.00 \\ 0.00 & 0.00 \end{bmatrix} \quad b_{io} = \begin{bmatrix} 30.00 \\ 30.00 \end{bmatrix} \quad W_{ho} = \begin{bmatrix} 0.00 & 0.00 \\ 0.00 & 0.00 \end{bmatrix} \quad b_{ho} = \begin{bmatrix} 0.00 \\ 0.00 \end{bmatrix}$$

$$x^{(4)} = [0.00, 1.00]^\top \qquad\qquad h^{(3)} = [0.00, 0.76]^\top$$

$$o^{(4)} = \sigma(W_{io}x^{(4)} + b_{io} + W_{ho}h^{(3)} + b_{ho}) \tag{50}$$

$$= \sigma([30.00, 30.00]^\top) \tag{51}$$

$$= [1.00, 1.00]^\top \tag{52}$$

Memory Contribution at $t = 4$: $\tilde{c}^{(4)}$

$$W_{i\tilde{c}} = \begin{bmatrix} 30.00 & 0.00 \\ 0.00 & 30.00 \end{bmatrix} \quad b_{i\tilde{c}} = \begin{bmatrix} 0.00 \\ 0.00 \end{bmatrix} \qquad W_{h\tilde{c}} = \begin{bmatrix} 0.00 & 0.00 \\ 0.00 & 0.00 \end{bmatrix} \quad b_{h\tilde{c}} = \begin{bmatrix} 0.00 \\ 0.00 \end{bmatrix}$$

$$x^{(4)} = [0.00, 1.00]^{\top} \qquad\qquad h^{(3)} = [0.00, 0.76]^{\top}$$

Memory Contribution at $t = 4$: $\tilde{c}^{(4)}$

$$W_{i\tilde{c}} = \begin{bmatrix} 30.00 & 0.00 \\ 0.00 & 30.00 \end{bmatrix} \quad b_{i\tilde{c}} = \begin{bmatrix} 0.00 \\ 0.00 \end{bmatrix} \qquad W_{h\tilde{c}} = \begin{bmatrix} 0.00 & 0.00 \\ 0.00 & 0.00 \end{bmatrix} \quad b_{h\tilde{c}} = \begin{bmatrix} 0.00 \\ 0.00 \end{bmatrix}$$

$$x^{(4)} = [0.00, 1.00]^\top \qquad\qquad h^{(3)} = [0.00, 0.76]^\top$$

$$
\begin{aligned}
\tilde{c}^{(4)} &= \tanh(W_{i\tilde{c}} x^{(4)} + b_{i\tilde{c}} + W_{h\tilde{c}} h^{(3)} + b_{h\tilde{c}}) && (53) \\
&= \tanh([0.00, 30.00]^\top) && (54) \\
&= [0.00, 1.00]^\top && (55)
\end{aligned}
$$

Forward message at time step 4

f_4	c_3	i_4	$\tilde{c}_4$
$[0.00, 0.00]^\top$	$[0.00, 1.00]^\top$	$[0.50, 0.00]^\top$	$[0.00, 1.00]^\top$

- Message forward (c_4)

$$c_4 = f_4 \circ c_3 + i_4 \circ \tilde{c}_4 \qquad (56)$$

$$(57)$$

Forward message at time step 4

f_4	c_3	i_4	$\tilde{c}_4$
$[0.00, 0.00]^\top$	$[0.00, 1.00]^\top$	$[0.50, 0.00]^\top$	$[0.00, 1.00]^\top$

- Message forward (c_4)

$$c_4 = f_4 \circ c_3 + i_4 \circ \tilde{c}_4 \tag{56}$$
$$= [0.00, 0.00]^\top \circ [0.00, 1.00]^\top + [0.50, 0.00]^\top \circ [0.00, 1.00]^\top \tag{57}$$
$$\tag{58}$$

Forward message at time step 4

f_4	c_3	i_4	$\tilde{c}_4$
$[0.00, 0.00]^\top$	$[0.00, 1.00]^\top$	$[0.50, 0.00]^\top$	$[0.00, 1.00]^\top$

- Message forward (c_4)

$$c_4 = f_4 \circ c_3 + i_4 \circ \tilde{c}_4 \tag{56}$$

$$= [0.00, 0.00]^\top \circ [0.00, 1.00]^\top + [0.50, 0.00]^\top \circ [0.00, 1.00]^\top \tag{57}$$

$$= [0.00, 0.00]^\top \tag{58}$$

f_4	c_3	i_4	$\tilde{c}_4$
$[0.00, 0.00]^\top$	$[0.00, 1.00]^\top$	$[0.50, 0.00]^\top$	$[0.00, 1.00]^\top$

- Message forward (c_4)

$$c_4 = [0.00, 0.00]^\top \qquad (56)$$

- New hidden (h_4)

$$h_4 \qquad (57)$$

f_4	c_3	i_4	$\tilde{c}_4$
$[0.00, 0.00]^\top$	$[0.00, 1.00]^\top$	$[0.50, 0.00]^\top$	$[0.00, 1.00]^\top$

- Message forward (c_4)

$$c_4 = [0.00, 0.00]^\top \qquad (56)$$

- New hidden (h_4)

$$h_4 = o_4 \circ \tanh(c_4) \qquad (57)$$

$$(58)$$

Forward message at time step 4

f_4	c_3	i_4	$\tilde{c}_4$
$[0.00, 0.00]^\top$	$[0.00, 1.00]^\top$	$[0.50, 0.00]^\top$	$[0.00, 1.00]^\top$

- Message forward (c_4)

$$c_4 = [0.00, 0.00]^\top \tag{56}$$

- New hidden (h_4)

$$h_4 = o_4 \circ \tanh(c_4) \tag{57}$$
$$= [1.00, 1.00]^\top \circ \tanh([0.00, 0.00]^\top) \tag{58}$$
$$\tag{59}$$

Forward message at time step 4

f_4	c_3	i_4	$\tilde{c}_4$
$[0.00, 0.00]^\top$	$[0.00, 1.00]^\top$	$[0.50, 0.00]^\top$	$[0.00, 1.00]^\top$

- Message forward (c_4)

$$c_4 = [0.00, 0.00]^\top \tag{56}$$

- New hidden (h_4)

$$h_4 = o_4 \circ \tanh(c_4) \tag{57}$$
$$= [1.00, 1.00]^\top \circ \tanh([0.00, 0.00]^\top) \tag{58}$$
$$= [0.00, 0.00]^\top \tag{59}$$

f_4	c_3	i_4	$\tilde{c}_4$
$[0.00, 0.00]^\top$	$[0.00, 1.00]^\top$	$[0.50, 0.00]^\top$	$[0.00, 1.00]^\top$

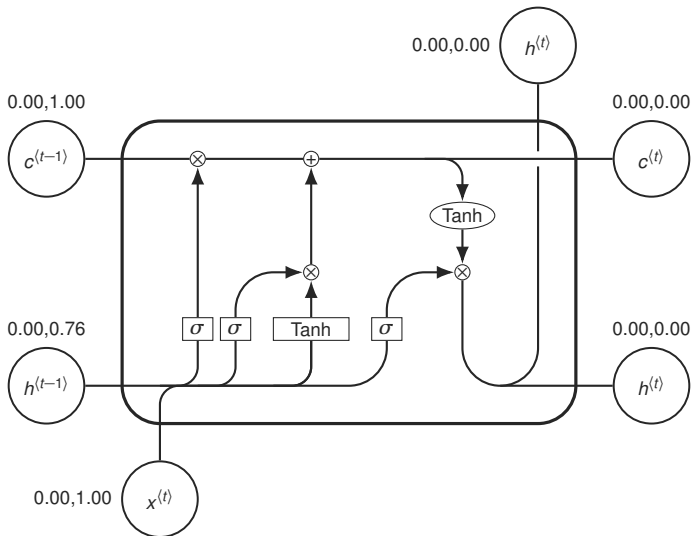- Message forward (c_4)

$$c_4 = [0.00, 0.00]^\top \tag{56}$$

- New hidden (h_4)

$$h_4 = [0.00, 0.00]^\top \tag{57}$$

- Prediction $y_4 = \text{softmax}(h_4) = 1$

Summary at $t = 4$

0.00,0.00 $h^{\langle t \rangle}$

0.00,1.00 $c^{\langle t-1 \rangle}$

0.00,0.00 $c^{\langle t \rangle}$

Tanh

0.00,0.76 $h^{\langle t-1 \rangle}$

σ σ Tanh σ

Tanh
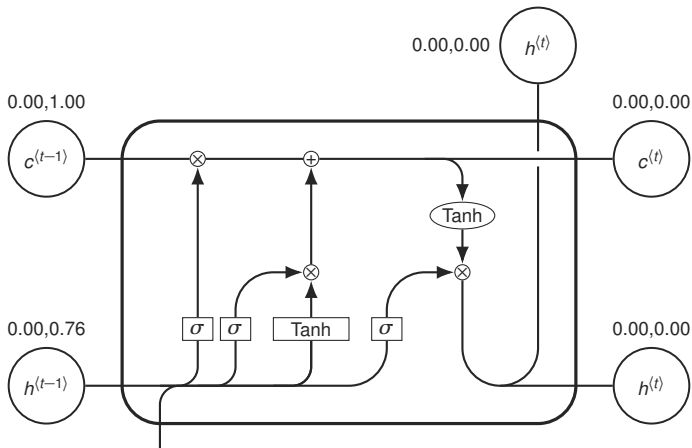
0.00,0.00 $h^{\langle t \rangle}$

Another B turns the message forward to (0, 0), same as start

$$W_{ii} = \begin{bmatrix} 30.00 & 0.00 \\ 0.00 & 0.00 \end{bmatrix} \quad b_{ii} = \begin{bmatrix} 0.00 \\ 0.00 \end{bmatrix} \qquad W_{hi} = \begin{bmatrix} 0.00 & 0.00 \\ 60.00 & 0.00 \end{bmatrix} \quad b_{hi} = \begin{bmatrix} 0.00 \\ -30.00 \end{bmatrix}$$

$$x^{(5)} = [1.00, 0.00]^\top \qquad h^{(4)} = [0.00, 0.00]^\top$$

Input Gate at $t = 5$: i_1

$$W_{ii} = \begin{bmatrix} 30.00 & 0.00 \\ 0.00 & 0.00 \end{bmatrix} \quad b_{ii} = \begin{bmatrix} 0.00 \\ 0.00 \end{bmatrix} \quad W_{hi} = \begin{bmatrix} 0.00 & 0.00 \\ 60.00 & 0.00 \end{bmatrix} b_{hi} = \begin{bmatrix} 0.00 \\ -30.00 \end{bmatrix}$$

$$x^{(5)} = [1.00, 0.00]^\top \qquad\qquad h^{(4)} = [0.00, 0.00]^\top$$

$$
\begin{align}
i^{(5)} &= \sigma(W_{ii}x^{(5)} + b_{ii} + W_{hi}h^{(4)} + b_{hi}) \tag{58} \\
&= \sigma([30.00, -30.00]^\top) \tag{59} \\
&= [1.00, 0.00]^\top \tag{60}
\end{align}
$$

Forget Gate at $t = 5$: $f^{(5)}$

$$W_{if} = \begin{bmatrix} 60.00 & 0.00 \\ 0.00 & 0.00 \end{bmatrix} \quad b_{if} = \begin{bmatrix} 0.00 \\ 0.00 \end{bmatrix} \qquad W_{hf} = \begin{bmatrix} 0.00 & 0.00 \\ 0.00 & -30.00 \end{bmatrix} b_{hf} = \begin{bmatrix} -30.00 \\ 0.00 \end{bmatrix}$$

$$x^{(5)} = [1.00, 0.00]^\top \qquad\qquad h^{(4)} = [0.00, 0.00]^\top$$

Forget Gate at $t = 5$: $f^{(5)}$

$$W_{if} = \begin{bmatrix} 60.00 & 0.00 \\ 0.00 & 0.00 \end{bmatrix} \quad b_{if} = \begin{bmatrix} 0.00 \\ 0.00 \end{bmatrix} \qquad W_{hf} = \begin{bmatrix} 0.00 & 0.00 \\ 0.00 & -30.00 \end{bmatrix} b_{hf} = \begin{bmatrix} -30.00 \\ 0.00 \end{bmatrix}$$

$$x^{(5)} = [1.00, 0.00]^\top \qquad\qquad h^{(4)} = [0.00, 0.00]^\top$$

$$f^{(5)} = \sigma\left(W_{if}x^{(5)} + b_{if} + W_{hf}h^{(4)} + b_{hf}\right) \tag{61}$$

$$= \sigma\left([30.00, -0.00]^\top\right) \tag{62}$$

$$= [1.00, 0.50]^\top \tag{63}$$

Output Gate at $t = 5$: $o^{(5)}$

$$W_{io} = \begin{bmatrix} 0.00 & 0.00 \\ 0.00 & 0.00 \end{bmatrix} \quad b_{io} = \begin{bmatrix} 30.00 \\ 30.00 \end{bmatrix} \qquad W_{ho} = \begin{bmatrix} 0.00 & 0.00 \\ 0.00 & 0.00 \end{bmatrix} \quad b_{ho} = \begin{bmatrix} 0.00 \\ 0.00 \end{bmatrix}$$

$$x^{(5)} = [1.00, 0.00]^{\top} \qquad\qquad h^{(4)} = [0.00, 0.00]^{\top}$$

Output Gate at $t = 5$: $o^{(5)}$

$$W_{io} = \begin{bmatrix} 0.00 & 0.00 \\ 0.00 & 0.00 \end{bmatrix} \quad b_{io} = \begin{bmatrix} 30.00 \\ 30.00 \end{bmatrix} \quad W_{ho} = \begin{bmatrix} 0.00 & 0.00 \\ 0.00 & 0.00 \end{bmatrix} \quad b_{ho} = \begin{bmatrix} 0.00 \\ 0.00 \end{bmatrix}$$

$$x^{(5)} = [1.00, 0.00]^\top \qquad\qquad h^{(4)} = [0.00, 0.00]^\top$$

$$o^{(5)} = \sigma(W_{io}x^{(5)} + b_{io} + W_{ho}h^{(4)} + b_{ho}) \qquad (64)$$

$$= \sigma([30.00, 30.00]^\top) \qquad (65)$$

$$= [1.00, 1.00]^\top \qquad (66)$$

Memory Contribution at $t = 5$: $\tilde{c}^{(5)}$

$$W_{i\tilde{c}} = \begin{bmatrix} 30.00 & 0.00 \\ 0.00 & 30.00 \end{bmatrix} \quad b_{i\tilde{c}} = \begin{bmatrix} 0.00 \\ 0.00 \end{bmatrix} \qquad W_{h\tilde{c}} = \begin{bmatrix} 0.00 & 0.00 \\ 0.00 & 0.00 \end{bmatrix} \quad b_{h\tilde{c}} = \begin{bmatrix} 0.00 \\ 0.00 \end{bmatrix}$$

$$x^{(5)} = [1.00, 0.00]^{\top} \qquad\qquad h^{(4)} = [0.00, 0.00]^{\top}$$

$$W_{i\tilde{c}} = \begin{bmatrix} 30.00 & 0.00 \\ 0.00 & 30.00 \end{bmatrix} \quad b_{i\tilde{c}} = \begin{bmatrix} 0.00 \\ 0.00 \end{bmatrix} \qquad W_{h\tilde{c}} = \begin{bmatrix} 0.00 & 0.00 \\ 0.00 & 0.00 \end{bmatrix} \quad b_{h\tilde{c}} = \begin{bmatrix} 0.00 \\ 0.00 \end{bmatrix}$$

$$x^{(5)} = [1.00, 0.00]^{\top} \qquad\qquad\qquad h^{(4)} = [0.00, 0.00]^{\top}$$

$$\tilde{c}^{(5)} = \tanh(W_{i\tilde{c}}x^{(5)} + b_{i\tilde{c}} + W_{h\tilde{c}}h^{(4)} + b_{h\tilde{c}}) \tag{67}$$

$$= \tanh([30.00, 0.00]^{\top}) \tag{68}$$

$$= [1.00, 0.00]^{\top} \tag{69}$$

Forward message at time step 5

f_5	c_4	i_5	$\tilde{c}_5$
$[1.00, 0.50]^\top$	$[0.00, 0.00]^\top$	$[1.00, 0.00]^\top$	$[1.00, 0.00]^\top$

- Message forward (c_5)

$$c_5 = f_5 \circ c_4 + i_5 \circ \tilde{c}_5 \tag{70}$$

$$\tag{71}$$

Forward message at time step 5

f_5	c_4	i_5	$\tilde{c}_5$
$[1.00, 0.50]^\top$	$[0.00, 0.00]^\top$	$[1.00, 0.00]^\top$	$[1.00, 0.00]^\top$

- Message forward (c_5)

$$c_5 = f_5 \circ c_4 + i_5 \circ \tilde{c}_5 \tag{70}$$
$$= [1.00, 0.50]^\top \circ [0.00, 0.00]^\top + [1.00, 0.00]^\top \circ [1.00, 0.00]^\top \tag{71}$$
$$\tag{72}$$

f_5	c_4	i_5	$\tilde{c}_5$
$[1.00, 0.50]^\top$	$[0.00, 0.00]^\top$	$[1.00, 0.00]^\top$	$[1.00, 0.00]^\top$

- Message forward (c_5)

$$c_5 = f_5 \circ c_4 + i_5 \circ \tilde{c}_5 \tag{70}$$
$$= [1.00, 0.50]^\top \circ [0.00, 0.00]^\top + [1.00, 0.00]^\top \circ [1.00, 0.00]^\top \tag{71}$$
$$= [1.00, 0.00]^\top \tag{72}$$

Forward message at time step 5

f_5	c_4	i_5	$\tilde{c}_5$
$[1.00, 0.50]^\top$	$[0.00, 0.00]^\top$	$[1.00, 0.00]^\top$	$[1.00, 0.00]^\top$

- Message forward (c_5)

$$c_5 = [1.00, 0.00]^\top \tag{70}$$

- New hidden (h_5)

$$h_5 \tag{71}$$

f_5	c_4	i_5	$\tilde{c}_5$
$[1.00, 0.50]^\top$	$[0.00, 0.00]^\top$	$[1.00, 0.00]^\top$	$[1.00, 0.00]^\top$

- Message forward (c_5)

$$c_5 = [1.00, 0.00]^\top \tag{70}$$

- New hidden (h_5)

$$h_5 = o_5 \circ \tanh(c_5) \tag{71}$$

$$\tag{72}$$

Forward message at time step 5

f_5	c_4	i_5	$\tilde{c}_5$
$[1.00, 0.50]^\top$	$[0.00, 0.00]^\top$	$[1.00, 0.00]^\top$	$[1.00, 0.00]^\top$

- Message forward (c_5)

$$c_5 = [1.00, 0.00]^\top \tag{70}$$

- New hidden (h_5)

$$h_5 = o_5 \circ \tanh(c_5) \tag{71}$$
$$= [1.00, 1.00]^\top \circ \tanh([1.00, 0.00]^\top) \tag{72}$$
$$\tag{73}$$

Forward message at time step 5

f_5	c_4	i_5	$\tilde{c}_5$
$[1.00, 0.50]^\top$	$[0.00, 0.00]^\top$	$[1.00, 0.00]^\top$	$[1.00, 0.00]^\top$

- Message forward (c_5)

$$c_5 = [1.00, 0.00]^\top \tag{70}$$

- New hidden (h_5)

$$h_5 = o_5 \circ \tanh(c_5) \tag{71}$$
$$= [1.00, 1.00]^\top \circ \tanh([1.00, 0.00]^\top) \tag{72}$$
$$= [0.76, 0.00]^\top \tag{73}$$

f_5	c_4	i_5	$\tilde{c}_5$
$[1.00, 0.50]^\top$	$[0.00, 0.00]^\top$	$[1.00, 0.00]^\top$	$[1.00, 0.00]^\top$

- Message forward (c_5)

$$c_5 = [1.00, 0.00]^\top \tag{70}$$

- New hidden (h_5)

$$h_5 = [0.76, 0.00]^\top \tag{71}$$

- Prediction $y_5 = \mathrm{softmax}(h_5) = 0$

0.76,0.00 $h^{\langle t \rangle}$

0.00,0.00 $c^{\langle t-1 \rangle}$

1.00,0.00 $c^{\langle t \rangle}$

Tanh

σ σ Tanh σ

0.00,0.00 $h^{\langle t-1 \rangle}$
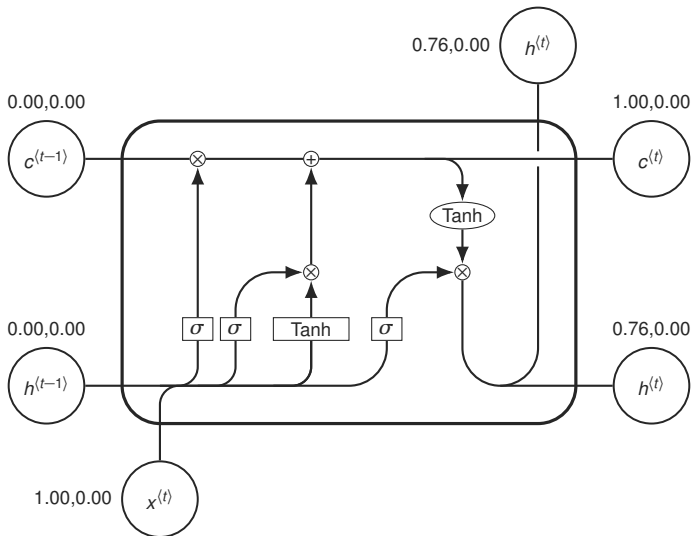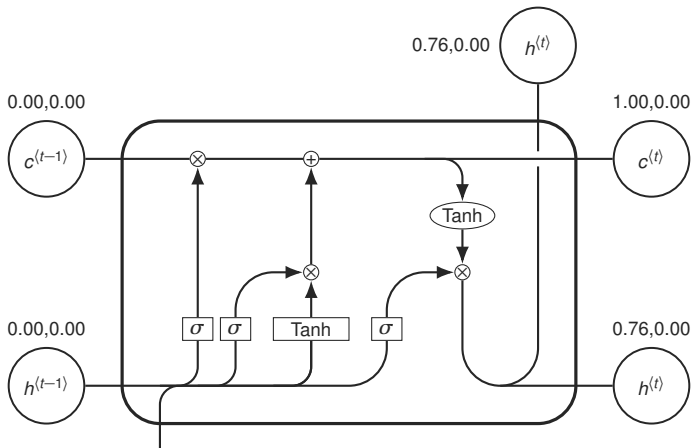
0.76,0.00 $h^{\langle t \rangle}$

Another A as input turns the message to (1, 0)

Input Gate at $t = 6$: i_1

$$W_{ii} = \begin{bmatrix} 30.00 & 0.00 \\ 0.00 & 0.00 \end{bmatrix} \quad b_{ii} = \begin{bmatrix} 0.00 \\ 0.00 \end{bmatrix} \qquad W_{hi} = \begin{bmatrix} 0.00 & 0.00 \\ 60.00 & 0.00 \end{bmatrix} \quad b_{hi} = \begin{bmatrix} 0.00 \\ -30.00 \end{bmatrix}$$

$$x^{(6)} = [0.00, 1.00]^\top \qquad\qquad h^{(5)} = [0.76, 0.00]^\top$$

Input Gate at $t = 6$: i_1

$$W_{ii} = \begin{bmatrix} 30.00 & 0.00 \\ 0.00 & 0.00 \end{bmatrix} \quad b_{ii} = \begin{bmatrix} 0.00 \\ 0.00 \end{bmatrix} \quad W_{hi} = \begin{bmatrix} 0.00 & 0.00 \\ 60.00 & 0.00 \end{bmatrix} b_{hi} = \begin{bmatrix} 0.00 \\ -30.00 \end{bmatrix}$$

$$x^{(6)} = [0.00, 1.00]^\top \qquad\qquad h^{(5)} = [0.76, 0.00]^\top$$

$$i^{(6)} = \sigma(W_{ii}x^{(6)} + b_{ii} + W_{hi}h^{(5)} + b_{hi}) \tag{72}$$
$$= \sigma([0.00, 15.70]^\top) \tag{73}$$
$$= [0.50, 1.00]^\top \tag{74}$$

Forget Gate at $t = 6$: $f^{(6)}$

$$W_{if} = \begin{bmatrix} 60.00 & 0.00 \\ 0.00 & 0.00 \end{bmatrix} \quad b_{if} = \begin{bmatrix} 0.00 \\ 0.00 \end{bmatrix} \quad W_{hf} = \begin{bmatrix} 0.00 & 0.00 \\ 0.00 & -30.00 \end{bmatrix} b_{hf} = \begin{bmatrix} -30.00 \\ 0.00 \end{bmatrix}$$

$$x^{(6)} = [0.00, 1.00]^\top \qquad\qquad h^{(5)} = [0.76, 0.00]^\top$$

Forget Gate at $t = 6$: $f^{(6)}$

$$W_{if} = \begin{bmatrix} 60.00 & 0.00 \\ 0.00 & 0.00 \end{bmatrix} \quad b_{if} = \begin{bmatrix} 0.00 \\ 0.00 \end{bmatrix} \quad W_{hf} = \begin{bmatrix} 0.00 & 0.00 \\ 0.00 & -30.00 \end{bmatrix} \quad b_{hf} = \begin{bmatrix} -30.00 \\ 0.00 \end{bmatrix}$$

$$x^{(6)} = [0.00, 1.00]^\top \qquad\qquad h^{(5)} = [0.76, 0.00]^\top$$

$$
\begin{align}
f^{(6)} &= \sigma \left(W_{if} x^{(6)} + b_{if} + W_{hf} h^{(5)} + b_{hf} \right) \tag{75} \\
&= \sigma \left([-30.00, -0.00]^\top \right) \tag{76} \\
&= [0.00, 0.50]^\top \tag{77}
\end{align}
$$

$$W_{io} = \begin{bmatrix} 0.00 & 0.00 \\ 0.00 & 0.00 \end{bmatrix} \quad b_{io} = \begin{bmatrix} 30.00 \\ 30.00 \end{bmatrix} \quad W_{ho} = \begin{bmatrix} 0.00 & 0.00 \\ 0.00 & 0.00 \end{bmatrix} \quad b_{ho} = \begin{bmatrix} 0.00 \\ 0.00 \end{bmatrix}$$

$$x^{(6)} = [0.00, 1.00]^{\top} \qquad\qquad h^{(5)} = [0.76, 0.00]^{\top}$$

$$W_{io} = \begin{bmatrix} 0.00 & 0.00 \\ 0.00 & 0.00 \end{bmatrix} \quad b_{io} = \begin{bmatrix} 30.00 \\ 30.00 \end{bmatrix} \quad W_{ho} = \begin{bmatrix} 0.00 & 0.00 \\ 0.00 & 0.00 \end{bmatrix} \quad b_{ho} = \begin{bmatrix} 0.00 \\ 0.00 \end{bmatrix}$$

$$x^{(6)} = [0.00, 1.00]^{\top} \qquad\qquad h^{(5)} = [0.76, 0.00]^{\top}$$

$$o^{(6)} = \sigma(W_{io}x^{(6)} + b_{io} + W_{ho}h^{(5)} + b_{ho}) \tag{78}$$

$$= \sigma([30.00, 30.00]^{\top}) \tag{79}$$

$$= [1.00, 1.00]^{\top} \tag{80}$$

$$W_{i\tilde{c}} = \begin{bmatrix} 30.00 & 0.00 \\ 0.00 & 30.00 \end{bmatrix} \quad b_{i\tilde{c}} = \begin{bmatrix} 0.00 \\ 0.00 \end{bmatrix}$$

$$x^{(6)} = [0.00, 1.00]^\top$$

$$W_{h\tilde{c}} = \begin{bmatrix} 0.00 & 0.00 \\ 0.00 & 0.00 \end{bmatrix} \quad b_{h\tilde{c}} = \begin{bmatrix} 0.00 \\ 0.00 \end{bmatrix}$$

$$h^{(5)} = [0.76, 0.00]^\top$$

$$W_{i\tilde{c}} = \begin{bmatrix} 30.00 & 0.00 \\ 0.00 & 30.00 \end{bmatrix} \quad b_{i\tilde{c}} = \begin{bmatrix} 0.00 \\ 0.00 \end{bmatrix} \qquad W_{h\tilde{c}} = \begin{bmatrix} 0.00 & 0.00 \\ 0.00 & 0.00 \end{bmatrix} \quad b_{h\tilde{c}} = \begin{bmatrix} 0.00 \\ 0.00 \end{bmatrix}$$

$$x^{(6)} = [0.00, 1.00]^{\top} \qquad\qquad h^{(5)} = [0.76, 0.00]^{\top}$$

$$\tilde{c}^{(6)} = \tanh(W_{i\tilde{c}} x^{(6)} + b_{i\tilde{c}} + W_{h\tilde{c}} h^{(5)} + b_{h\tilde{c}}) \qquad (81)$$

$$= \tanh([0.00, 30.00]^{\top}) \qquad (82)$$

$$= [0.00, 1.00]^{\top} \qquad (83)$$

Forward message at time step 6

f_6	c_5	i_6	$\tilde{c}_6$
$[0.00, 0.50]^\top$	$[1.00, 0.00]^\top$	$[0.50, 1.00]^\top$	$[0.00, 1.00]^\top$

- Message forward (c_6)

$$c_6 = f_6 \circ c_5 + i_6 \circ \tilde{c}_6 \tag{84}$$

$$\tag{85}$$

f_6	c_5	i_6	$\tilde{c}_6$
$[0.00, 0.50]^\top$	$[1.00, 0.00]^\top$	$[0.50, 1.00]^\top$	$[0.00, 1.00]^\top$

- Message forward (c_6)

$$c_6 = f_6 \circ c_5 + i_6 \circ \tilde{c}_6 \tag{84}$$
$$= [0.00, 0.50]^\top \circ [1.00, 0.00]^\top + [0.50, 1.00]^\top \circ [0.00, 1.00]^\top \tag{85}$$
$$\tag{86}$$

Forward message at time step 6

f_6	c_5	i_6	$\tilde{c}_6$
$[0.00, 0.50]^\top$	$[1.00, 0.00]^\top$	$[0.50, 1.00]^\top$	$[0.00, 1.00]^\top$

- Message forward (c_6)

$$c_6 = f_6 \circ c_5 + i_6 \circ \tilde{c}_6 \tag{84}$$
$$= [0.00, 0.50]^\top \circ [1.00, 0.00]^\top + [0.50, 1.00]^\top \circ [0.00, 1.00]^\top \tag{85}$$
$$= [0.00, 1.00]^\top \tag{86}$$

Forward message at time step 6

f_6	c_5	i_6	$\tilde{c}_6$
$[0.00, 0.50]^\top$	$[1.00, 0.00]^\top$	$[0.50, 1.00]^\top$	$[0.00, 1.00]^\top$

- Message forward (c_6)

$$c_6 = [0.00, 1.00]^\top \qquad (84)$$

- New hidden (h_6)

$$h_6 \qquad (85)$$

Forward message at time step 6

f_6	c_5	i_6	$\tilde{c}_6$
$[0.00, 0.50]^\top$	$[1.00, 0.00]^\top$	$[0.50, 1.00]^\top$	$[0.00, 1.00]^\top$

- Message forward (c_6)

$$c_6 = [0.00, 1.00]^\top \tag{84}$$

- New hidden (h_6)

$$h_6 = o_6 \circ \tanh(c_6) \tag{85}$$

$$\tag{86}$$

Forward message at time step 6

f_6	c_5	i_6	$\tilde{c}_6$
$[0.00, 0.50]^\top$	$[1.00, 0.00]^\top$	$[0.50, 1.00]^\top$	$[0.00, 1.00]^\top$

- Message forward (c_6)

$$c_6 = [0.00, 1.00]^\top \tag{84}$$

- New hidden (h_6)

$$h_6 = o_6 \circ \tanh(c_6) \tag{85}$$
$$= [1.00, 1.00]^\top \circ \tanh([0.00, 1.00]^\top) \tag{86}$$
$$\tag{87}$$

f_6	c_5	i_6	$\tilde{c}_6$
$[0.00, 0.50]^\top$	$[1.00, 0.00]^\top$	$[0.50, 1.00]^\top$	$[0.00, 1.00]^\top$

- Message forward (c_6)

$$c_6 = [0.00, 1.00]^\top \tag{84}$$

- New hidden (h_6)

$$\begin{align}
h_6 &= o_6 \circ \tanh(c_6) \tag{85}\\
&= [1.00, 1.00]^\top \circ \tanh([0.00, 1.00]^\top) \tag{86}\\
&= [0.00, 0.76]^\top \tag{87}
\end{align}$$

f_6	c_5	i_6	$\tilde{c}_6$
$[0.00, 0.50]^\top$	$[1.00, 0.00]^\top$	$[0.50, 1.00]^\top$	$[0.00, 1.00]^\top$

- Message forward (c_6)

$$c_6 = [0.00, 1.00]^\top \tag{84}$$

- New hidden (h_6)

$$h_6 = [0.00, 0.76]^\top \tag{85}$$

- Prediction $y_6 = \mathrm{softmax}(h_6) = 1$

Summary at $t = 6$

47

Input Gate at $t = 7$: i_1

$$W_{ii} = \begin{bmatrix} 30.00 & 0.00 \\ 0.00 & 0.00 \end{bmatrix} \quad b_{ii} = \begin{bmatrix} 0.00 \\ 0.00 \end{bmatrix} \qquad W_{hi} = \begin{bmatrix} 0.00 & 0.00 \\ 60.00 & 0.00 \end{bmatrix} \quad b_{hi} = \begin{bmatrix} 0.00 \\ -30.00 \end{bmatrix}$$

$$x^{(7)} = [1.00, 0.00]^\top \qquad\qquad h^{(6)} = [0.00, 0.76]^\top$$

Input Gate at $t = 7$: i_1

$$W_{ii} = \begin{bmatrix} 30.00 & 0.00 \\ 0.00 & 0.00 \end{bmatrix} \quad b_{ii} = \begin{bmatrix} 0.00 \\ 0.00 \end{bmatrix} \quad W_{hi} = \begin{bmatrix} 0.00 & 0.00 \\ 60.00 & 0.00 \end{bmatrix} \quad b_{hi} = \begin{bmatrix} 0.00 \\ -30.00 \end{bmatrix}$$

$$x^{(7)} = [1.00, 0.00]^\top \qquad\qquad h^{(6)} = [0.00, 0.76]^\top$$

$$
\begin{align}
i^{(7)} &= \sigma(W_{ii}x^{(7)} + b_{ii} + W_{hi}h^{(6)} + b_{hi}) \tag{86} \\
&= \sigma([30.00, -30.00]^\top) \tag{87} \\
&= [1.00, 0.00]^\top \tag{88}
\end{align}
$$

Forget Gate at $t = 7$: $f^{(7)}$

$$W_{if} = \begin{bmatrix} 60.00 & 0.00 \\ 0.00 & 0.00 \end{bmatrix} \quad b_{if} = \begin{bmatrix} 0.00 \\ 0.00 \end{bmatrix} \quad W_{hf} = \begin{bmatrix} 0.00 & 0.00 \\ 0.00 & -30.00 \end{bmatrix} b_{hf} = \begin{bmatrix} -30.00 \\ 0.00 \end{bmatrix}$$

$$x^{(7)} = [1.00, 0.00]^{\top} \qquad\qquad h^{(6)} = [0.00, 0.76]^{\top}$$

Forget Gate at $t = 7$: $f^{(7)}$

$$W_{if} = \begin{bmatrix} 60.00 & 0.00 \\ 0.00 & 0.00 \end{bmatrix} \quad b_{if} = \begin{bmatrix} 0.00 \\ 0.00 \end{bmatrix} \qquad W_{hf} = \begin{bmatrix} 0.00 & 0.00 \\ 0.00 & -30.00 \end{bmatrix} b_{hf} = \begin{bmatrix} -30.00 \\ 0.00 \end{bmatrix}$$

$$x^{(7)} = [1.00, 0.00]^\top \qquad\qquad h^{(6)} = [0.00, 0.76]^\top$$

$$f^{(7)} = \sigma(W_{if}x^{(7)} + b_{if} + W_{hf}h^{(6)} + b_{hf}) \tag{89}$$

$$= \sigma([30.00, -22.85]^\top) \tag{90}$$

$$= [1.00, 0.00]^\top \tag{91}$$

$$W_{io} = \begin{bmatrix} 0.00 & 0.00 \\ 0.00 & 0.00 \end{bmatrix} \quad b_{io} = \begin{bmatrix} 30.00 \\ 30.00 \end{bmatrix} \quad W_{ho} = \begin{bmatrix} 0.00 & 0.00 \\ 0.00 & 0.00 \end{bmatrix} \quad b_{ho} = \begin{bmatrix} 0.00 \\ 0.00 \end{bmatrix}$$

$$x^{(7)} = [1.00, 0.00]^\top \qquad h^{(6)} = [0.00, 0.76]^\top$$

Output Gate at $t = 7$: $o^{(7)}$

$$W_{io} = \begin{bmatrix} 0.00 & 0.00 \\ 0.00 & 0.00 \end{bmatrix} \quad b_{io} = \begin{bmatrix} 30.00 \\ 30.00 \end{bmatrix} \quad W_{ho} = \begin{bmatrix} 0.00 & 0.00 \\ 0.00 & 0.00 \end{bmatrix} \quad b_{ho} = \begin{bmatrix} 0.00 \\ 0.00 \end{bmatrix}$$

$$x^{(7)} = [1.00, 0.00]^\top \qquad\qquad h^{(6)} = [0.00, 0.76]^\top$$

$$o^{(7)} = \sigma(W_{io}x^{(7)} + b_{io} + W_{ho}h^{(6)} + b_{ho}) \tag{92}$$

$$= \sigma([30.00, 30.00]^\top) \tag{93}$$

$$= [1.00, 1.00]^\top \tag{94}$$

$$W_{i\tilde{c}} = \begin{bmatrix} 30.00 & 0.00 \\ 0.00 & 30.00 \end{bmatrix} \quad b_{i\tilde{c}} = \begin{bmatrix} 0.00 \\ 0.00 \end{bmatrix}$$

$$x^{(7)} = [1.00, 0.00]^{\top}$$

$$W_{h\tilde{c}} = \begin{bmatrix} 0.00 & 0.00 \\ 0.00 & 0.00 \end{bmatrix} \quad b_{h\tilde{c}} = \begin{bmatrix} 0.00 \\ 0.00 \end{bmatrix}$$

$$h^{(6)} = [0.00, 0.76]^{\top}$$

Memory Contribution at $t = 7$: $\tilde{c}^{(7)}$

$$W_{i\tilde{c}} = \begin{bmatrix} 30.00 & 0.00 \\ 0.00 & 30.00 \end{bmatrix} b_{i\tilde{c}} = \begin{bmatrix} 0.00 \\ 0.00 \end{bmatrix} \qquad W_{h\tilde{c}} = \begin{bmatrix} 0.00 & 0.00 \\ 0.00 & 0.00 \end{bmatrix} \quad b_{h\tilde{c}} = \begin{bmatrix} 0.00 \\ 0.00 \end{bmatrix}$$

$$x^{(7)} = [1.00, 0.00]^{\top} \qquad\qquad h^{(6)} = [0.00, 0.76]^{\top}$$

$$
\begin{align}
\tilde{c}^{(7)} &= \tanh(W_{i\tilde{c}} x^{(7)} + b_{i\tilde{c}} + W_{h\tilde{c}} h^{(6)} + b_{h\tilde{c}}) \tag{95}\\
&= \tanh([30.00, 0.00]^{\top}) \tag{96}\\
&= [1.00, 0.00]^{\top} \tag{97}
\end{align}
$$

f_7	c_6	i_7	$\tilde{c}_7$
$[1.00, 0.00]^\top$	$[0.00, 1.00]^\top$	$[1.00, 0.00]^\top$	$[1.00, 0.00]^\top$

- Message forward (c_7)

$$c_7 = f_7 \circ c_6 + i_7 \circ \tilde{c}_7 \tag{98}$$

$$\tag{99}$$

Forward message at time step 7

f_7	c_6	i_7	$\tilde{c}_7$
$[1.00, 0.00]^\top$	$[0.00, 1.00]^\top$	$[1.00, 0.00]^\top$	$[1.00, 0.00]^\top$

- Message forward (c_7)

$$c_7 = f_7 \circ c_6 + i_7 \circ \tilde{c}_7 \tag{98}$$
$$= [1.00, 0.00]^\top \circ [0.00, 1.00]^\top + [1.00, 0.00]^\top \circ [1.00, 0.00]^\top \tag{99}$$
$$\tag{100}$$

Forward message at time step 7

f_7	c_6	i_7	$\tilde{c}_7$
$[1.00, 0.00]^\top$	$[0.00, 1.00]^\top$	$[1.00, 0.00]^\top$	$[1.00, 0.00]^\top$

- Message forward (c_7)

$$c_7 = f_7 \circ c_6 + i_7 \circ \tilde{c}_7 \tag{98}$$
$$= [1.00, 0.00]^\top \circ [0.00, 1.00]^\top + [1.00, 0.00]^\top \circ [1.00, 0.00]^\top \tag{99}$$
$$= [1.00, 0.00]^\top \tag{100}$$

f_7	c_6	i_7	$\tilde{c}_7$
$[1.00, 0.00]^\top$	$[0.00, 1.00]^\top$	$[1.00, 0.00]^\top$	$[1.00, 0.00]^\top$

- Message forward (c_7)

$$c_7 = [1.00, 0.00]^\top \tag{98}$$

- New hidden (h_7)

$$h_7 \tag{99}$$

Forward message at time step 7

f_7	c_6	i_7	$\tilde{c}_7$
$[1.00, 0.00]^\top$	$[0.00, 1.00]^\top$	$[1.00, 0.00]^\top$	$[1.00, 0.00]^\top$

- Message forward (c_7)

$$c_7 = [1.00, 0.00]^\top \tag{98}$$

- New hidden (h_7)

$$h_7 = o_7 \circ \tanh(c_7) \tag{99}$$

$$\tag{100}$$

f_7	c_6	i_7	$\tilde{c}_7$
$[1.00, 0.00]^\top$	$[0.00, 1.00]^\top$	$[1.00, 0.00]^\top$	$[1.00, 0.00]^\top$

- Message forward (c_7)

$$c_7 = [1.00, 0.00]^\top \tag{98}$$

- New hidden (h_7)

$$h_7 = o_7 \circ \tanh(c_7) \tag{99}$$
$$= [1.00, 1.00]^\top \circ \tanh([1.00, 0.00]^\top) \tag{100}$$
$$\tag{101}$$

f_7	c_6	i_7	$\tilde{c}_7$
$[1.00, 0.00]^\top$	$[0.00, 1.00]^\top$	$[1.00, 0.00]^\top$	$[1.00, 0.00]^\top$

- Message forward (c_7)

$$c_7 = [1.00, 0.00]^\top \tag{98}$$

- New hidden (h_7)

$$
\begin{aligned}
h_7 &= o_7 \circ \tanh(c_7) \tag{99} \\
&= [1.00, 1.00]^\top \circ \tanh([1.00, 0.00]^\top) \tag{100} \\
&= [0.76, 0.00]^\top \tag{101}
\end{aligned}
$$

Forward message at time step 7

f_7	c_6	i_7	$\tilde{c}_7$
$[1.00, 0.00]^\top$	$[0.00, 1.00]^\top$	$[1.00, 0.00]^\top$	$[1.00, 0.00]^\top$

- Message forward (c_7)
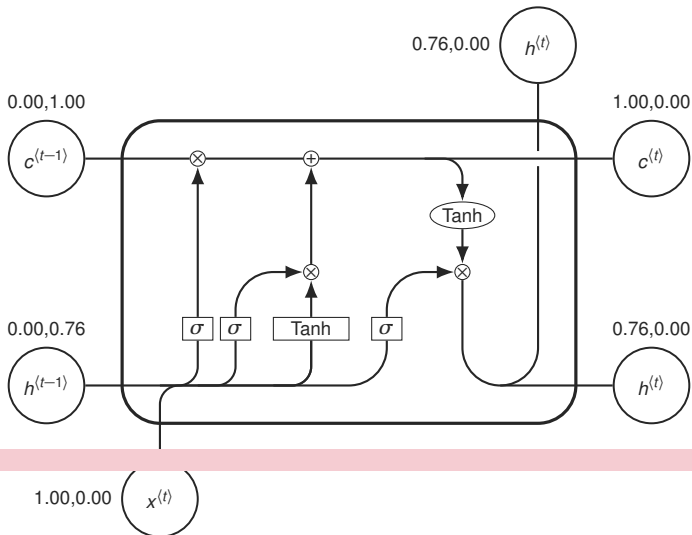
$$c_7 = [1.00, 0.00]^\top \tag{98}$$

- New hidden (h_7)

$$h_7 = [0.76, 0.00]^\top \tag{99}$$

- Prediction $y_7 = \mathrm{softmax}(h_7) = 0$
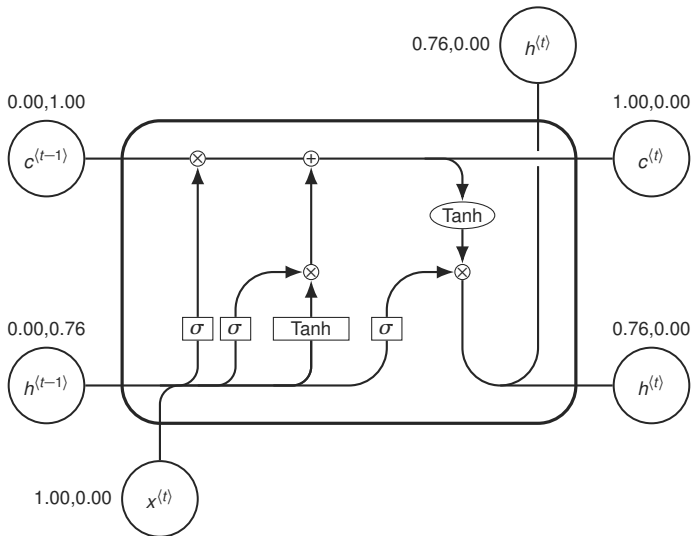
Summary at $t = 7$

What's going on?

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