

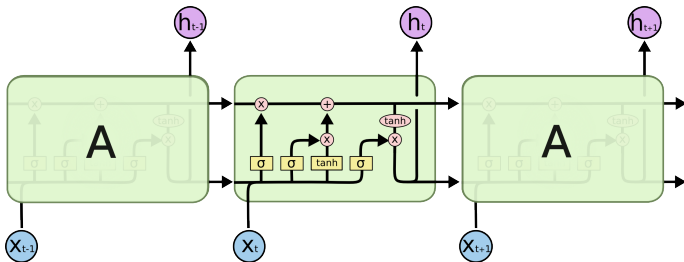


Long Short Term Memory Networks

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)

$$i_t = \sigma(W_{ij}x_t + b_{ij} + W_{hi}h_{t-1} + b_{hi})$$

$$f_t = \sigma(W_{if}x_t + b_{if} + W_{hf}h_{t-1} + b_{hf})$$

$$o_t = \sigma(W_{io}x_t + b_{io} + W_{ho}h_{t-1} + b_{ho})$$

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	
1.0	0.0

B	
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:

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

Parameters that take x_t as input

Input Gate

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

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

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

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

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

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

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

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

Inputs

- Initial hidden states:

$$h_0 = [0.0, 0.0]^T$$

- Initial memory input:

$$c_0 = [0.0, 0.0]^T$$

- Input sequences in time: A, A, B, B, A, B, A

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

Input Gate at $t = 1$: i_1

$$W_{ii} = \begin{bmatrix} 30.00 & 0.00 \\ 0.00 & 0.00 \end{bmatrix} \quad b_{ii} = \begin{bmatrix} 0.00 \\ 0.00 \end{bmatrix}$$
$$x^{(1)} = [1.00, 0.00]^\top$$

$$W_{hi} = \begin{bmatrix} 0.00 & 0.00 \\ 60.00 & 0.00 \end{bmatrix} \quad b_{hi} = \begin{bmatrix} 0.00 \\ -30.00 \end{bmatrix}$$
$$h^{(0)} = [0.00, 0.00]^\top$$

Input Gate at $t = 1$: i_1

$$W_{ii} = \begin{bmatrix} 30.00 & 0.00 \\ 0.00 & 0.00 \end{bmatrix} \quad b_{ii} = \begin{bmatrix} 0.00 \\ 0.00 \end{bmatrix} \quad W_{hi} = \begin{bmatrix} 0.00 & 0.00 \\ 60.00 & 0.00 \end{bmatrix} \quad b_{hi} = \begin{bmatrix} 0.00 \\ -30.00 \end{bmatrix}$$
$$x^{(1)} = [1.00, 0.00]^\top \quad h^{(0)} = [0.00, 0.00]^\top$$

$$i^{(1)} = \sigma(W_{ii}x^{(1)} + b_{ii} + W_{hi}h^{(0)} + b_{hi}) \quad (1)$$

$$= \sigma([30.00, -30.00]^\top) \quad (2)$$

$$= [1.00, 0.00]^\top \quad (3)$$

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

$$W_{if} = \begin{bmatrix} 0.00 & 0.00 \\ 0.00 & 0.00 \end{bmatrix} \quad b_{if} = \begin{bmatrix} 0.00 \\ 0.00 \end{bmatrix}$$
$$x^{(1)} = [1.00, 0.00]^\top$$

$$W_{hf} = \begin{bmatrix} 0.00 & 0.00 \\ 0.00 & -30.00 \end{bmatrix} \quad b_{hf} = \begin{bmatrix} -30.00 \\ 0.00 \end{bmatrix}$$
$$h^{(0)} = [0.00, 0.00]^\top$$

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

$$W_{if} = \begin{bmatrix} 0.00 & 0.00 \\ 0.00 & 0.00 \end{bmatrix} \quad b_{if} = \begin{bmatrix} 0.00 \\ 0.00 \end{bmatrix} \quad W_{hf} = \begin{bmatrix} 0.00 & 0.00 \\ 0.00 & -30.00 \end{bmatrix} \quad b_{hf} = \begin{bmatrix} -30.00 \\ 0.00 \end{bmatrix}$$
$$x^{(1)} = [1.00, 0.00]^\top \quad h^{(0)} = [0.00, 0.00]^\top$$

$$f^{(1)} = \sigma(W_{if}x^{(1)} + b_{if} + W_{hf}h^{(0)} + b_{hf}) \quad (4)$$

$$= \sigma([-30.00, 0.00]^\top) \quad (5)$$

$$= [0.00, 0.50]^\top \quad (6)$$

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

$$W_{io} = \begin{bmatrix} 0.00 & 0.00 \\ 0.00 & 0.00 \end{bmatrix} \quad b_{io} = \begin{bmatrix} 30.00 \\ 30.00 \end{bmatrix}$$
$$x^{(1)} = [1.00, 0.00]^\top$$

$$W_{ho} = \begin{bmatrix} 0.00 & 0.00 \\ 0.00 & 0.00 \end{bmatrix} \quad b_{ho} = \begin{bmatrix} 0.00 \\ 0.00 \end{bmatrix}$$
$$h^{(0)} = [0.00, 0.00]^\top$$

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

$$W_{io} = \begin{bmatrix} 0.00 & 0.00 \\ 0.00 & 0.00 \end{bmatrix} \quad b_{io} = \begin{bmatrix} 30.00 \\ 30.00 \end{bmatrix} \quad W_{ho} = \begin{bmatrix} 0.00 & 0.00 \\ 0.00 & 0.00 \end{bmatrix} \quad b_{ho} = \begin{bmatrix} 0.00 \\ 0.00 \end{bmatrix}$$
$$x^{(1)} = [1.00, 0.00]^\top \quad h^{(0)} = [0.00, 0.00]^\top$$

$$o^{(1)} = \sigma(W_{io}x^{(1)} + b_{io} + W_{ho}h^{(0)} + b_{ho}) \quad (7)$$

$$= \sigma([30.00, 30.00]^\top) \quad (8)$$

$$= [1.00, 1.00]^\top \quad (9)$$

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

$$W_{i\tilde{c}} = \begin{bmatrix} 30.00 & 0.00 \\ 0.00 & 30.00 \end{bmatrix} \quad b_{i\tilde{c}} = \begin{bmatrix} 0.00 \\ 0.00 \end{bmatrix}$$
$$x^{(1)} = [1.00, 0.00]^\top$$

$$W_{h\tilde{c}} = \begin{bmatrix} 0.00 & 0.00 \\ 0.00 & 0.00 \end{bmatrix} \quad b_{h\tilde{c}} = \begin{bmatrix} 0.00 \\ 0.00 \end{bmatrix}$$
$$h^{(0)} = [0.00, 0.00]^\top$$

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

$$W_{i\tilde{c}} = \begin{bmatrix} 30.00 & 0.00 \\ 0.00 & 30.00 \end{bmatrix} \quad b_{i\tilde{c}} = \begin{bmatrix} 0.00 \\ 0.00 \end{bmatrix} \quad W_{h\tilde{c}} = \begin{bmatrix} 0.00 & 0.00 \\ 0.00 & 0.00 \end{bmatrix} \quad b_{h\tilde{c}} = \begin{bmatrix} 0.00 \\ 0.00 \end{bmatrix}$$
$$x^{(1)} = [1.00, 0.00]^\top \quad h^{(0)} = [0.00, 0.00]^\top$$

$$\tilde{c}^{(1)} = \tanh(W_{i\tilde{c}}x^{(1)} + b_{i\tilde{c}} + W_{h\tilde{c}}h^{(0)} + b_{h\tilde{c}}) \quad (10)$$

$$= \tanh([30.00, 0.00]^\top) \quad (11)$$

$$= [1.00, 0.00]^\top \quad (12)$$

Forward message at time step 1

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

- Message forward (c_1)

$$c_1 = f_1 \circ c_0 + i_1 \circ \tilde{c}_1 \quad (13)$$

$$(14)$$

Forward message at time step 1

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

■ Message forward (c_1)

$$c_1 = f_1 \circ c_0 + i_1 \circ \tilde{c}_1 \quad (13)$$

$$= [0.00, 0.50]^\top \circ [0.00, 0.00]^\top + [1.00, 0.00]^\top \circ [1.00, 0.00]^\top \quad (14)$$

$$(15)$$

Forward message at time step 1

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

■ Message forward (c_1)

$$c_1 = f_1 \circ c_0 + i_1 \circ \tilde{c}_1 \quad (13)$$

$$= [0.00, 0.50]^\top \circ [0.00, 0.00]^\top + [1.00, 0.00]^\top \circ [1.00, 0.00]^\top \quad (14)$$

$$= [1.00, 0.00]^\top \quad (15)$$

Forward message at time step 1

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

- Message forward (c_1)

$$c_1 = [1.00, 0.00]^\top \quad (13)$$

- New hidden (h_1)

$$h_1 \quad (14)$$

Forward message at time step 1

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

- Message forward (c_1)

$$c_1 = [1.00, 0.00]^\top \quad (13)$$

- New hidden (h_1)

$$h_1 = o_1 \circ \tanh(c_1) \quad (14)$$

$$(15)$$

Forward message at time step 1

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

- Message forward (c_1)

$$c_1 = [1.00, 0.00]^\top \quad (13)$$

- New hidden (h_1)

$$h_1 = o_1 \circ \tanh(c_1) \quad (14)$$

$$= [1.00, 1.00]^\top \circ \tanh([1.00, 0.00]^\top) \quad (15)$$

$$(16)$$

Forward message at time step 1

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

- Message forward (c_1)

$$c_1 = [1.00, 0.00]^\top \quad (13)$$

- New hidden (h_1)

$$h_1 = o_1 \circ \tanh(c_1) \quad (14)$$

$$= [1.00, 1.00]^\top \circ \tanh([1.00, 0.00]^\top) \quad (15)$$

$$= [0.76, 0.00]^\top \quad (16)$$

Forward message at time step 1

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

- Message forward (c_1)

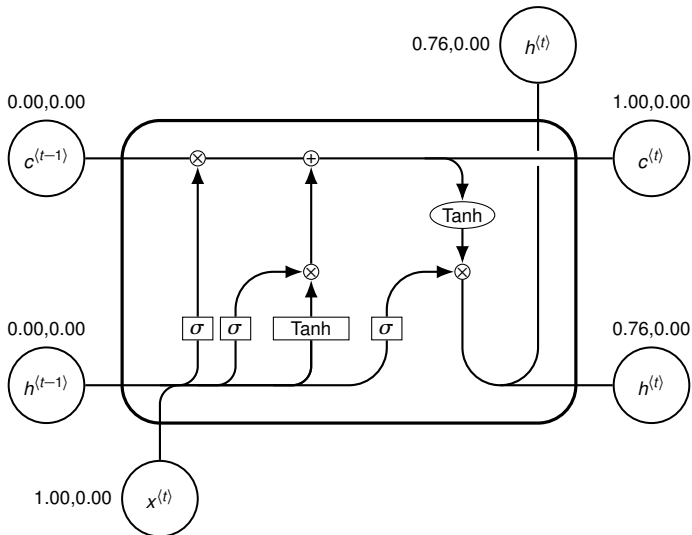
$$c_1 = [1.00, 0.00]^\top \quad (13)$$

- New hidden (h_1)

$$h_1 = [0.76, 0.00]^\top \quad (14)$$

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

Summary at $t = 1$



Input Gate at $t = 2$: i_t

$$W_{ii} = \begin{bmatrix} 30.00 & 0.00 \\ 0.00 & 0.00 \end{bmatrix} \quad b_{ii} = \begin{bmatrix} 0.00 \\ 0.00 \end{bmatrix}$$
$$x^{(2)} = [1.00, 0.00]^\top$$

$$W_{hi} = \begin{bmatrix} 0.00 & 0.00 \\ 60.00 & 0.00 \end{bmatrix} \quad b_{hi} = \begin{bmatrix} 0.00 \\ -30.00 \end{bmatrix}$$
$$h^{(1)} = [0.76, 0.00]^\top$$

Input Gate at $t = 2$: i_t

$$W_{ii} = \begin{bmatrix} 30.00 & 0.00 \\ 0.00 & 0.00 \end{bmatrix} \quad b_{ii} = \begin{bmatrix} 0.00 \\ 0.00 \end{bmatrix} \quad W_{hi} = \begin{bmatrix} 0.00 & 0.00 \\ 60.00 & 0.00 \end{bmatrix} \quad b_{hi} = \begin{bmatrix} 0.00 \\ -30.00 \end{bmatrix}$$
$$x^{(2)} = [1.00, 0.00]^\top \quad h^{(1)} = [0.76, 0.00]^\top$$

$$i^{(2)} = \sigma(W_{ii}x^{(2)} + b_{ii} + W_{hi}h^{(1)} + b_{hi}) \quad (15)$$

$$= \sigma([30.00, 15.70]^\top) \quad (16)$$

$$= [1.00, 1.00]^\top \quad (17)$$

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

$$W_{if} = \begin{bmatrix} 0.00 & 0.00 \\ 0.00 & 0.00 \end{bmatrix} \quad b_{if} = \begin{bmatrix} 0.00 \\ 0.00 \end{bmatrix}$$
$$x^{(2)} = [1.00, 0.00]^\top$$

$$W_{hf} = \begin{bmatrix} 0.00 & 0.00 \\ 0.00 & -30.00 \end{bmatrix} \quad b_{hf} = \begin{bmatrix} -30.00 \\ 0.00 \end{bmatrix}$$
$$h^{(1)} = [0.76, 0.00]^\top$$

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

$$W_{if} = \begin{bmatrix} 0.00 & 0.00 \\ 0.00 & 0.00 \end{bmatrix} \quad b_{if} = \begin{bmatrix} 0.00 \\ 0.00 \end{bmatrix} \quad W_{hf} = \begin{bmatrix} 0.00 & 0.00 \\ 0.00 & -30.00 \end{bmatrix} \quad b_{hf} = \begin{bmatrix} -30.00 \\ 0.00 \end{bmatrix}$$
$$x^{(2)} = [1.00, 0.00]^\top \quad h^{(1)} = [0.76, 0.00]^\top$$

$$f^{(2)} = \sigma(W_{if}x^{(2)} + b_{if} + W_{hf}h^{(1)} + b_{hf}) \quad (18)$$

$$= \sigma([-30.00, 0.00]^\top) \quad (19)$$

$$= [0.00, 0.50]^\top \quad (20)$$

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

$$W_{io} = \begin{bmatrix} 0.00 & 0.00 \\ 0.00 & 0.00 \end{bmatrix} \quad b_{io} = \begin{bmatrix} 30.00 \\ 30.00 \end{bmatrix}$$
$$x^{(2)} = [1.00, 0.00]^\top$$

$$W_{ho} = \begin{bmatrix} 0.00 & 0.00 \\ 0.00 & 0.00 \end{bmatrix} \quad b_{ho} = \begin{bmatrix} 0.00 \\ 0.00 \end{bmatrix}$$
$$h^{(1)} = [0.76, 0.00]^\top$$

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

$$W_{io} = \begin{bmatrix} 0.00 & 0.00 \\ 0.00 & 0.00 \end{bmatrix} \quad b_{io} = \begin{bmatrix} 30.00 \\ 30.00 \end{bmatrix} \quad W_{ho} = \begin{bmatrix} 0.00 & 0.00 \\ 0.00 & 0.00 \end{bmatrix} \quad b_{ho} = \begin{bmatrix} 0.00 \\ 0.00 \end{bmatrix}$$
$$x^{(2)} = [1.00, 0.00]^\top \quad h^{(1)} = [0.76, 0.00]^\top$$

$$o^{(2)} = \sigma(W_{io}x^{(2)} + b_{io} + W_{ho}h^{(1)} + b_{ho}) \quad (21)$$

$$= \sigma([30.00, 30.00]^\top) \quad (22)$$

$$= [1.00, 1.00]^\top \quad (23)$$

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

$$W_{i\tilde{c}} = \begin{bmatrix} 30.00 & 0.00 \\ 0.00 & 30.00 \end{bmatrix} \quad b_{i\tilde{c}} = \begin{bmatrix} 0.00 \\ 0.00 \end{bmatrix}$$
$$x^{(2)} = [1.00, 0.00]^\top$$

$$W_{h\tilde{c}} = \begin{bmatrix} 0.00 & 0.00 \\ 0.00 & 0.00 \end{bmatrix} \quad b_{h\tilde{c}} = \begin{bmatrix} 0.00 \\ 0.00 \end{bmatrix}$$
$$h^{(1)} = [0.76, 0.00]^\top$$

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

$$W_{i\tilde{c}} = \begin{bmatrix} 30.00 & 0.00 \\ 0.00 & 30.00 \end{bmatrix} \quad b_{i\tilde{c}} = \begin{bmatrix} 0.00 \\ 0.00 \end{bmatrix} \quad W_{h\tilde{c}} = \begin{bmatrix} 0.00 & 0.00 \\ 0.00 & 0.00 \end{bmatrix} \quad b_{h\tilde{c}} = \begin{bmatrix} 0.00 \\ 0.00 \end{bmatrix}$$
$$x^{(2)} = [1.00, 0.00]^\top \quad h^{(1)} = [0.76, 0.00]^\top$$

$$\tilde{c}^{(2)} = \tanh(W_{i\tilde{c}}x^{(2)} + b_{i\tilde{c}} + W_{h\tilde{c}}h^{(1)} + b_{h\tilde{c}}) \quad (24)$$

$$= \tanh([30.00, 0.00]^\top) \quad (25)$$

$$= [1.00, 0.00]^\top \quad (26)$$

Forward message at time step 2

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

- Message forward (c_2)

$$c_2 = f_2 \circ c_1 + i_2 \circ \tilde{c}_2 \quad (27)$$

$$(28)$$

Forward message at time step 2

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

- Message forward (c_2)

$$c_2 = f_2 \circ c_1 + i_2 \circ \tilde{c}_2 \quad (27)$$

$$= [0.00, 0.50]^\top \circ [1.00, 0.00]^\top + [1.00, 1.00]^\top \circ [1.00, 0.00]^\top \quad (28)$$

$$(29)$$

Forward message at time step 2

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

■ Message forward (c_2)

$$c_2 = f_2 \circ c_1 + i_2 \circ \tilde{c}_2 \quad (27)$$

$$= [0.00, 0.50]^\top \circ [1.00, 0.00]^\top + [1.00, 1.00]^\top \circ [1.00, 0.00]^\top \quad (28)$$

$$= [1.00, 0.00]^\top \quad (29)$$

Forward message at time step 2

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

- Message forward (c_2)

$$c_2 = [1.00, 0.00]^\top \quad (27)$$

- New hidden (h_2)

$$h_2 \quad (28)$$

Forward message at time step 2

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

- Message forward (c_2)

$$c_2 = [1.00, 0.00]^\top \quad (27)$$

- New hidden (h_2)

$$h_2 = o_2 \circ \tanh(c_2) \quad (28)$$

$$(29)$$

Forward message at time step 2

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

- Message forward (c_2)

$$c_2 = [1.00, 0.00]^\top \quad (27)$$

- New hidden (h_2)

$$h_2 = o_2 \circ \tanh(c_2) \quad (28)$$

$$= [1.00, 1.00]^\top \circ \tanh([1.00, 0.00]^\top) \quad (29)$$

$$(30)$$

Forward message at time step 2

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

- Message forward (c_2)

$$c_2 = [1.00, 0.00]^\top \quad (27)$$

- New hidden (h_2)

$$h_2 = o_2 \circ \tanh(c_2) \quad (28)$$

$$= [1.00, 1.00]^\top \circ \tanh([1.00, 0.00]^\top) \quad (29)$$

$$= [0.76, 0.00]^\top \quad (30)$$

Forward message at time step 2

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

- Message forward (c_2)

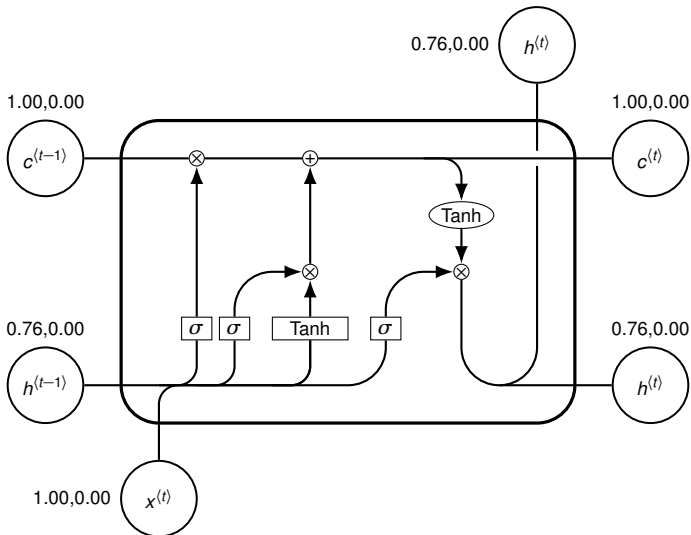
$$c_2 = [1.00, 0.00]^\top \quad (27)$$

- New hidden (h_2)

$$h_2 = [0.76, 0.00]^\top \quad (28)$$

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

Summary at $t = 2$



Input Gate at $t = 3$: i_t

$$W_{ii} = \begin{bmatrix} 30.00 & 0.00 \\ 0.00 & 0.00 \end{bmatrix} \quad b_{ii} = \begin{bmatrix} 0.00 \\ 0.00 \end{bmatrix}$$
$$x^{(3)} = [0.00, 1.00]^\top$$

$$W_{hi} = \begin{bmatrix} 0.00 & 0.00 \\ 60.00 & 0.00 \end{bmatrix} \quad b_{hi} = \begin{bmatrix} 0.00 \\ -30.00 \end{bmatrix}$$
$$h^{(2)} = [0.76, 0.00]^\top$$

Input Gate at $t = 3$: i_t

$$W_{ii} = \begin{bmatrix} 30.00 & 0.00 \\ 0.00 & 0.00 \end{bmatrix} \quad b_{ii} = \begin{bmatrix} 0.00 \\ 0.00 \end{bmatrix} \quad W_{hi} = \begin{bmatrix} 0.00 & 0.00 \\ 60.00 & 0.00 \end{bmatrix} \quad b_{hi} = \begin{bmatrix} 0.00 \\ -30.00 \end{bmatrix}$$
$$x^{(3)} = [0.00, 1.00]^\top \quad h^{(2)} = [0.76, 0.00]^\top$$

$$i^{(3)} = \sigma(W_{ii}x^{(3)} + b_{ii} + W_{hi}h^{(2)} + b_{hi}) \quad (29)$$

$$= \sigma([0.00, 15.70]^\top) \quad (30)$$

$$= [0.50, 1.00]^\top \quad (31)$$

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

$$W_{if} = \begin{bmatrix} 0.00 & 0.00 \\ 0.00 & 0.00 \end{bmatrix} \quad b_{if} = \begin{bmatrix} 0.00 \\ 0.00 \end{bmatrix}$$
$$x^{(3)} = [0.00, 1.00]^\top$$

$$W_{hf} = \begin{bmatrix} 0.00 & 0.00 \\ 0.00 & -30.00 \end{bmatrix} \quad b_{hf} = \begin{bmatrix} -30.00 \\ 0.00 \end{bmatrix}$$
$$h^{(2)} = [0.76, 0.00]^\top$$

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

$$W_{if} = \begin{bmatrix} 0.00 & 0.00 \\ 0.00 & 0.00 \end{bmatrix} \quad b_{if} = \begin{bmatrix} 0.00 \\ 0.00 \end{bmatrix} \quad W_{hf} = \begin{bmatrix} 0.00 & 0.00 \\ 0.00 & -30.00 \end{bmatrix} \quad b_{hf} = \begin{bmatrix} -30.00 \\ 0.00 \end{bmatrix}$$
$$x^{(3)} = [0.00, 1.00]^\top \quad h^{(2)} = [0.76, 0.00]^\top$$

$$f^{(3)} = \sigma(W_{if}x^{(3)} + b_{if} + W_{hf}h^{(2)} + b_{hf}) \quad (32)$$

$$= \sigma([-30.00, 0.00]^\top) \quad (33)$$

$$= [0.00, 0.50]^\top \quad (34)$$

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

$$W_{io} = \begin{bmatrix} 0.00 & 0.00 \\ 0.00 & 0.00 \end{bmatrix} \quad b_{io} = \begin{bmatrix} 30.00 \\ 30.00 \end{bmatrix}$$
$$x^{(3)} = [0.00, 1.00]^\top$$

$$W_{ho} = \begin{bmatrix} 0.00 & 0.00 \\ 0.00 & 0.00 \end{bmatrix} \quad b_{ho} = \begin{bmatrix} 0.00 \\ 0.00 \end{bmatrix}$$
$$h^{(2)} = [0.76, 0.00]^\top$$

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

$$W_{io} = \begin{bmatrix} 0.00 & 0.00 \\ 0.00 & 0.00 \end{bmatrix} \quad b_{io} = \begin{bmatrix} 30.00 \\ 30.00 \end{bmatrix} \quad W_{ho} = \begin{bmatrix} 0.00 & 0.00 \\ 0.00 & 0.00 \end{bmatrix} \quad b_{ho} = \begin{bmatrix} 0.00 \\ 0.00 \end{bmatrix}$$
$$x^{(3)} = [0.00, 1.00]^\top \quad h^{(2)} = [0.76, 0.00]^\top$$

$$o^{(3)} = \sigma(W_{io}x^{(3)} + b_{io} + W_{ho}h^{(2)} + b_{ho}) \quad (35)$$

$$= \sigma([30.00, 30.00]^\top) \quad (36)$$

$$= [1.00, 1.00]^\top \quad (37)$$

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

$$W_{i\tilde{c}} = \begin{bmatrix} 30.00 & 0.00 \\ 0.00 & 30.00 \end{bmatrix} \quad b_{i\tilde{c}} = \begin{bmatrix} 0.00 \\ 0.00 \end{bmatrix}$$
$$x^{(3)} = [0.00, 1.00]^\top$$

$$W_{h\tilde{c}} = \begin{bmatrix} 0.00 & 0.00 \\ 0.00 & 0.00 \end{bmatrix} \quad b_{h\tilde{c}} = \begin{bmatrix} 0.00 \\ 0.00 \end{bmatrix}$$
$$h^{(2)} = [0.76, 0.00]^\top$$

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

$$W_{i\tilde{c}} = \begin{bmatrix} 30.00 & 0.00 \\ 0.00 & 30.00 \end{bmatrix} \quad b_{i\tilde{c}} = \begin{bmatrix} 0.00 \\ 0.00 \end{bmatrix} \quad W_{h\tilde{c}} = \begin{bmatrix} 0.00 & 0.00 \\ 0.00 & 0.00 \end{bmatrix} \quad b_{h\tilde{c}} = \begin{bmatrix} 0.00 \\ 0.00 \end{bmatrix}$$
$$x^{(3)} = [0.00, 1.00]^\top \quad h^{(2)} = [0.76, 0.00]^\top$$

$$\tilde{c}^{(3)} = \tanh(W_{i\tilde{c}}x^{(3)} + b_{i\tilde{c}} + W_{h\tilde{c}}h^{(2)} + b_{h\tilde{c}}) \quad (38)$$

$$= \tanh([0.00, 30.00]^\top) \quad (39)$$

$$= [0.00, 1.00]^\top \quad (40)$$

Forward message at time step 3

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

- Message forward (c_3)

$$c_3 = f_3 \circ c_2 + i_3 \circ \tilde{c}_3 \quad (41)$$

$$(42)$$

Forward message at time step 3

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

■ Message forward (c_3)

$$c_3 = f_3 \circ c_2 + i_3 \circ \tilde{c}_3 \quad (41)$$

$$= [0.00, 0.50]^\top \circ [1.00, 0.00]^\top + [0.50, 1.00]^\top \circ [0.00, 1.00]^\top \quad (42)$$

$$(43)$$

Forward message at time step 3

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

■ Message forward (c_3)

$$c_3 = f_3 \circ c_2 + i_3 \circ \tilde{c}_3 \quad (41)$$

$$= [0.00, 0.50]^\top \circ [1.00, 0.00]^\top + [0.50, 1.00]^\top \circ [0.00, 1.00]^\top \quad (42)$$

$$= [0.00, 1.00]^\top \quad (43)$$

Forward message at time step 3

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

- Message forward (c_3)

$$c_3 = [0.00, 1.00]^\top \quad (41)$$

- New hidden (h_3)

$$h_3 \quad (42)$$

Forward message at time step 3

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

- Message forward (c_3)

$$c_3 = [0.00, 1.00]^\top \quad (41)$$

- New hidden (h_3)

$$h_3 = o_3 \circ \tanh(c_3) \quad (42)$$

$$(43)$$

Forward message at time step 3

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

- Message forward (c_3)

$$c_3 = [0.00, 1.00]^\top \quad (41)$$

- New hidden (h_3)

$$h_3 = o_3 \circ \tanh(c_3) \quad (42)$$

$$= [1.00, 1.00]^\top \circ \tanh([0.00, 1.00]^\top) \quad (43)$$

$$(44)$$

Forward message at time step 3

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

- Message forward (c_3)

$$c_3 = [0.00, 1.00]^\top \quad (41)$$

- New hidden (h_3)

$$h_3 = o_3 \circ \tanh(c_3) \quad (42)$$

$$= [1.00, 1.00]^\top \circ \tanh([0.00, 1.00]^\top) \quad (43)$$

$$= [0.00, 0.76]^\top \quad (44)$$

Forward message at time step 3

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

- Message forward (c_3)

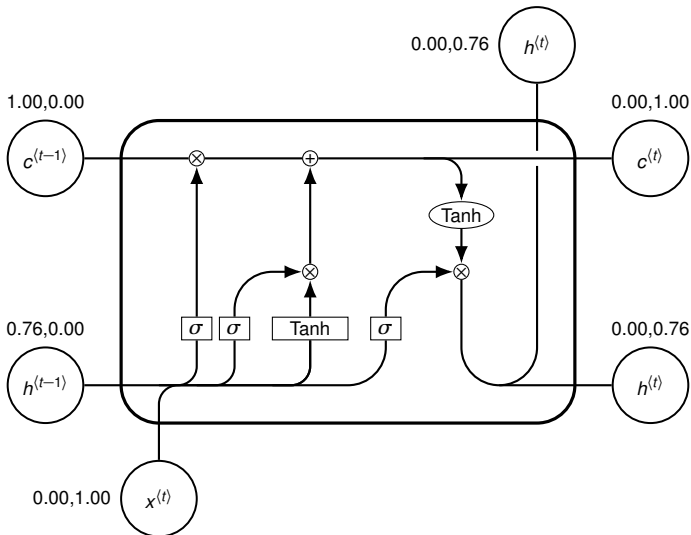
$$c_3 = [0.00, 1.00]^\top \quad (41)$$

- New hidden (h_3)

$$h_3 = [0.00, 0.76]^\top \quad (42)$$

- Prediction $y_3 = \text{softmax}(h_3) = 1$

Summary at $t = 3$



Input Gate at $t = 4$: i_t

$$W_{ii} = \begin{bmatrix} 30.00 & 0.00 \\ 0.00 & 0.00 \end{bmatrix} \quad b_{ii} = \begin{bmatrix} 0.00 \\ 0.00 \end{bmatrix}$$
$$x^{(4)} = [0.00, 1.00]^\top$$

$$W_{hi} = \begin{bmatrix} 0.00 & 0.00 \\ 60.00 & 0.00 \end{bmatrix} \quad b_{hi} = \begin{bmatrix} 0.00 \\ -30.00 \end{bmatrix}$$
$$h^{(3)} = [0.00, 0.76]^\top$$

Input Gate at $t = 4$: i_t

$$W_{ii} = \begin{bmatrix} 30.00 & 0.00 \\ 0.00 & 0.00 \end{bmatrix} \quad b_{ii} = \begin{bmatrix} 0.00 \\ 0.00 \end{bmatrix} \quad W_{hi} = \begin{bmatrix} 0.00 & 0.00 \\ 60.00 & 0.00 \end{bmatrix} \quad b_{hi} = \begin{bmatrix} 0.00 \\ -30.00 \end{bmatrix}$$
$$x^{(4)} = [0.00, 1.00]^\top \quad h^{(3)} = [0.00, 0.76]^\top$$

$$i^{(4)} = \sigma(W_{ii}x^{(4)} + b_{ii} + W_{hi}h^{(3)} + b_{hi}) \quad (43)$$

$$= \sigma([0.00, -30.00]^\top) \quad (44)$$

$$= [0.50, 0.00]^\top \quad (45)$$

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

$$W_{if} = \begin{bmatrix} 0.00 & 0.00 \\ 0.00 & 0.00 \end{bmatrix} \quad b_{if} = \begin{bmatrix} 0.00 \\ 0.00 \end{bmatrix}$$
$$x^{(4)} = [0.00, 1.00]^\top$$

$$W_{hf} = \begin{bmatrix} 0.00 & 0.00 \\ 0.00 & -30.00 \end{bmatrix} \quad b_{hf} = \begin{bmatrix} -30.00 \\ 0.00 \end{bmatrix}$$
$$h^{(3)} = [0.00, 0.76]^\top$$

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

$$W_{if} = \begin{bmatrix} 0.00 & 0.00 \\ 0.00 & 0.00 \end{bmatrix} \quad b_{if} = \begin{bmatrix} 0.00 \\ 0.00 \end{bmatrix} \quad W_{hf} = \begin{bmatrix} 0.00 & 0.00 \\ 0.00 & -30.00 \end{bmatrix} \quad b_{hf} = \begin{bmatrix} -30.00 \\ 0.00 \end{bmatrix}$$
$$x^{(4)} = [0.00, 1.00]^\top \quad h^{(3)} = [0.00, 0.76]^\top$$

$$f^{(4)} = \sigma(W_{if}x^{(4)} + b_{if} + W_{hf}h^{(3)} + b_{hf}) \quad (46)$$

$$= \sigma([-30.00, -22.85]^\top) \quad (47)$$

$$= [0.00, 0.00]^\top \quad (48)$$

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

$$W_{io} = \begin{bmatrix} 0.00 & 0.00 \\ 0.00 & 0.00 \end{bmatrix} \quad b_{io} = \begin{bmatrix} 30.00 \\ 30.00 \end{bmatrix}$$
$$x^{(4)} = [0.00, 1.00]^\top$$

$$W_{ho} = \begin{bmatrix} 0.00 & 0.00 \\ 0.00 & 0.00 \end{bmatrix} \quad b_{ho} = \begin{bmatrix} 0.00 \\ 0.00 \end{bmatrix}$$
$$h^{(3)} = [0.00, 0.76]^\top$$

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

$$W_{io} = \begin{bmatrix} 0.00 & 0.00 \\ 0.00 & 0.00 \end{bmatrix} \quad b_{io} = \begin{bmatrix} 30.00 \\ 30.00 \end{bmatrix} \quad W_{ho} = \begin{bmatrix} 0.00 & 0.00 \\ 0.00 & 0.00 \end{bmatrix} \quad b_{ho} = \begin{bmatrix} 0.00 \\ 0.00 \end{bmatrix}$$
$$x^{(4)} = [0.00, 1.00]^\top \quad h^{(3)} = [0.00, 0.76]^\top$$

$$o^{(4)} = \sigma(W_{io}x^{(4)} + b_{io} + W_{ho}h^{(3)} + b_{ho}) \quad (49)$$

$$= \sigma([30.00, 30.00]^\top) \quad (50)$$

$$= [1.00, 1.00]^\top \quad (51)$$

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

$$W_{i\tilde{c}} = \begin{bmatrix} 30.00 & 0.00 \\ 0.00 & 30.00 \end{bmatrix} \quad b_{i\tilde{c}} = \begin{bmatrix} 0.00 \\ 0.00 \end{bmatrix}$$
$$x^{(4)} = [0.00, 1.00]^\top$$

$$W_{h\tilde{c}} = \begin{bmatrix} 0.00 & 0.00 \\ 0.00 & 0.00 \end{bmatrix} \quad b_{h\tilde{c}} = \begin{bmatrix} 0.00 \\ 0.00 \end{bmatrix}$$
$$h^{(3)} = [0.00, 0.76]^\top$$

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

$$W_{i\tilde{c}} = \begin{bmatrix} 30.00 & 0.00 \\ 0.00 & 30.00 \end{bmatrix} \quad b_{i\tilde{c}} = \begin{bmatrix} 0.00 \\ 0.00 \end{bmatrix} \quad W_{h\tilde{c}} = \begin{bmatrix} 0.00 & 0.00 \\ 0.00 & 0.00 \end{bmatrix} \quad b_{h\tilde{c}} = \begin{bmatrix} 0.00 \\ 0.00 \end{bmatrix}$$
$$x^{(4)} = [0.00, 1.00]^\top \quad h^{(3)} = [0.00, 0.76]^\top$$

$$\tilde{c}^{(4)} = \tanh(W_{i\tilde{c}}x^{(4)} + b_{i\tilde{c}} + W_{h\tilde{c}}h^{(3)} + b_{h\tilde{c}}) \quad (52)$$

$$= \tanh([0.00, 30.00]^\top) \quad (53)$$

$$= [0.00, 1.00]^\top \quad (54)$$

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)

$$c_4 = f_4 \circ c_3 + i_4 \circ \tilde{c}_4 \quad (55)$$

$$(56)$$

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)

$$c_4 = f_4 \circ c_3 + i_4 \circ \tilde{c}_4 \quad (55)$$

$$= [0.00, 0.00]^\top \circ [0.00, 1.00]^\top + [0.50, 0.00]^\top \circ [0.00, 1.00]^\top \quad (56)$$

$$(57)$$

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)

$$c_4 = f_4 \circ c_3 + i_4 \circ \tilde{c}_4 \quad (55)$$

$$= [0.00, 0.00]^\top \circ [0.00, 1.00]^\top + [0.50, 0.00]^\top \circ [0.00, 1.00]^\top \quad (56)$$

$$= [0.00, 0.00]^\top \quad (57)$$

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)

$$c_4 = [0.00, 0.00]^\top \quad (55)$$

- New hidden (h_4)

$$h_4 \quad (56)$$

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)

$$c_4 = [0.00, 0.00]^\top \quad (55)$$

- New hidden (h_4)

$$h_4 = o_4 \circ \tanh(c_4) \quad (56)$$

$$(57)$$

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)

$$c_4 = [0.00, 0.00]^\top \quad (55)$$

- New hidden (h_4)

$$h_4 = o_4 \circ \tanh(c_4) \quad (56)$$

$$= [1.00, 1.00]^\top \circ \tanh([0.00, 0.00]^\top) \quad (57)$$

$$(58)$$

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)

$$c_4 = [0.00, 0.00]^\top \quad (55)$$

- New hidden (h_4)

$$h_4 = o_4 \circ \tanh(c_4) \quad (56)$$

$$= [1.00, 1.00]^\top \circ \tanh([0.00, 0.00]^\top) \quad (57)$$

$$= [0.00, 0.00]^\top \quad (58)$$

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)

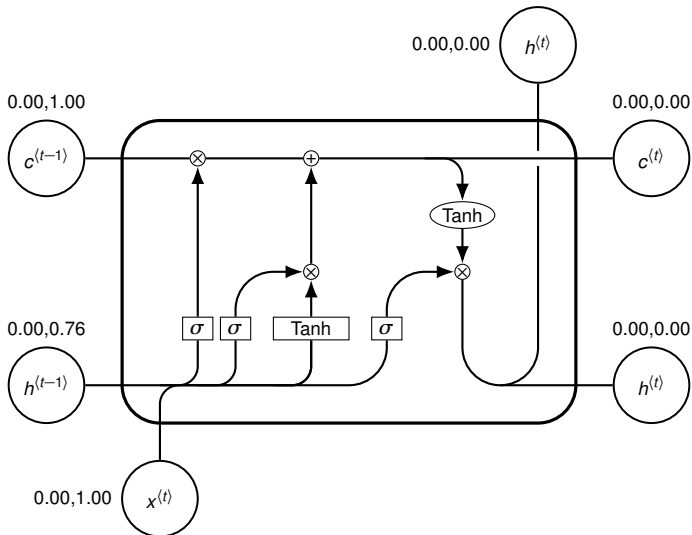
$$c_4 = [0.00, 0.00]^\top \quad (55)$$

- New hidden (h_4)

$$h_4 = [0.00, 0.00]^\top \quad (56)$$

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

Summary at $t = 4$



Input Gate at $t = 5$: i_t

$$W_{ii} = \begin{bmatrix} 30.00 & 0.00 \\ 0.00 & 0.00 \end{bmatrix} \quad b_{ii} = \begin{bmatrix} 0.00 \\ 0.00 \end{bmatrix}$$
$$x^{(5)} = [1.00, 0.00]^\top$$

$$W_{hi} = \begin{bmatrix} 0.00 & 0.00 \\ 60.00 & 0.00 \end{bmatrix} \quad b_{hi} = \begin{bmatrix} 0.00 \\ -30.00 \end{bmatrix}$$
$$h^{(4)} = [0.00, 0.00]^\top$$

Input Gate at $t = 5$: i_t

$$W_{ii} = \begin{bmatrix} 30.00 & 0.00 \\ 0.00 & 0.00 \end{bmatrix} \quad b_{ii} = \begin{bmatrix} 0.00 \\ 0.00 \end{bmatrix} \quad W_{hi} = \begin{bmatrix} 0.00 & 0.00 \\ 60.00 & 0.00 \end{bmatrix} \quad b_{hi} = \begin{bmatrix} 0.00 \\ -30.00 \end{bmatrix}$$
$$x^{(5)} = [1.00, 0.00]^\top \quad h^{(4)} = [0.00, 0.00]^\top$$

$$i^{(5)} = \sigma(W_{ii}x^{(5)} + b_{ii} + W_{hi}h^{(4)} + b_{hi}) \quad (57)$$

$$= \sigma([30.00, -30.00]^\top) \quad (58)$$

$$= [1.00, 0.00]^\top \quad (59)$$

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

$$W_{if} = \begin{bmatrix} 0.00 & 0.00 \\ 0.00 & 0.00 \end{bmatrix} \quad b_{if} = \begin{bmatrix} 0.00 \\ 0.00 \end{bmatrix}$$
$$x^{(5)} = [1.00, 0.00]^\top$$

$$W_{hf} = \begin{bmatrix} 0.00 & 0.00 \\ 0.00 & -30.00 \end{bmatrix} \quad b_{hf} = \begin{bmatrix} -30.00 \\ 0.00 \end{bmatrix}$$
$$h^{(4)} = [0.00, 0.00]^\top$$

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

$$W_{if} = \begin{bmatrix} 0.00 & 0.00 \\ 0.00 & 0.00 \end{bmatrix} \quad b_{if} = \begin{bmatrix} 0.00 \\ 0.00 \end{bmatrix} \quad W_{hf} = \begin{bmatrix} 0.00 & 0.00 \\ 0.00 & -30.00 \end{bmatrix} \quad b_{hf} = \begin{bmatrix} -30.00 \\ 0.00 \end{bmatrix}$$
$$x^{(5)} = [1.00, 0.00]^\top \quad h^{(4)} = [0.00, 0.00]^\top$$

$$f^{(5)} = \sigma(W_{if}x^{(5)} + b_{if} + W_{hf}h^{(4)} + b_{hf}) \quad (60)$$

$$= \sigma([-30.00, -0.00]^\top) \quad (61)$$

$$= [0.00, 0.50]^\top \quad (62)$$

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

$$W_{io} = \begin{bmatrix} 0.00 & 0.00 \\ 0.00 & 0.00 \end{bmatrix} \quad b_{io} = \begin{bmatrix} 30.00 \\ 30.00 \end{bmatrix}$$
$$x^{(5)} = [1.00, 0.00]^\top$$

$$W_{ho} = \begin{bmatrix} 0.00 & 0.00 \\ 0.00 & 0.00 \end{bmatrix} \quad b_{ho} = \begin{bmatrix} 0.00 \\ 0.00 \end{bmatrix}$$
$$h^{(4)} = [0.00, 0.00]^\top$$

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

$$W_{io} = \begin{bmatrix} 0.00 & 0.00 \\ 0.00 & 0.00 \end{bmatrix} \quad b_{io} = \begin{bmatrix} 30.00 \\ 30.00 \end{bmatrix} \quad W_{ho} = \begin{bmatrix} 0.00 & 0.00 \\ 0.00 & 0.00 \end{bmatrix} \quad b_{ho} = \begin{bmatrix} 0.00 \\ 0.00 \end{bmatrix}$$
$$x^{(5)} = [1.00, 0.00]^\top \quad h^{(4)} = [0.00, 0.00]^\top$$

$$o^{(5)} = \sigma(W_{io}x^{(5)} + b_{io} + W_{ho}h^{(4)} + b_{ho}) \quad (63)$$

$$= \sigma([30.00, 30.00]^\top) \quad (64)$$

$$= [1.00, 1.00]^\top \quad (65)$$

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

$$W_{i\tilde{c}} = \begin{bmatrix} 30.00 & 0.00 \\ 0.00 & 30.00 \end{bmatrix} \quad b_{i\tilde{c}} = \begin{bmatrix} 0.00 \\ 0.00 \end{bmatrix}$$
$$x^{(5)} = [1.00, 0.00]^\top$$

$$W_{h\tilde{c}} = \begin{bmatrix} 0.00 & 0.00 \\ 0.00 & 0.00 \end{bmatrix} \quad b_{h\tilde{c}} = \begin{bmatrix} 0.00 \\ 0.00 \end{bmatrix}$$
$$h^{(4)} = [0.00, 0.00]^\top$$

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

$$W_{i\tilde{c}} = \begin{bmatrix} 30.00 & 0.00 \\ 0.00 & 30.00 \end{bmatrix} \quad b_{i\tilde{c}} = \begin{bmatrix} 0.00 \\ 0.00 \end{bmatrix} \quad W_{h\tilde{c}} = \begin{bmatrix} 0.00 & 0.00 \\ 0.00 & 0.00 \end{bmatrix} \quad b_{h\tilde{c}} = \begin{bmatrix} 0.00 \\ 0.00 \end{bmatrix}$$
$$x^{(5)} = [1.00, 0.00]^\top \quad h^{(4)} = [0.00, 0.00]^\top$$

$$\tilde{c}^{(5)} = \tanh(W_{i\tilde{c}}x^{(5)} + b_{i\tilde{c}} + W_{h\tilde{c}}h^{(4)} + b_{h\tilde{c}}) \quad (66)$$

$$= \tanh([30.00, 0.00]^\top) \quad (67)$$

$$= [1.00, 0.00]^\top \quad (68)$$

Forward message at time step 5

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

- Message forward (c_5)

$$c_5 = f_5 \circ c_4 + i_5 \circ \tilde{c}_5 \quad (69)$$

$$(70)$$

Forward message at time step 5

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

■ Message forward (c_5)

$$c_5 = f_5 \circ c_4 + i_5 \circ \tilde{c}_5 \quad (69)$$

$$= [0.00, 0.50]^\top \circ [0.00, 0.00]^\top + [1.00, 0.00]^\top \circ [1.00, 0.00]^\top \quad (70)$$

$$(71)$$

Forward message at time step 5

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

■ Message forward (c_5)

$$c_5 = f_5 \circ c_4 + i_5 \circ \tilde{c}_5 \quad (69)$$

$$= [0.00, 0.50]^\top \circ [0.00, 0.00]^\top + [1.00, 0.00]^\top \circ [1.00, 0.00]^\top \quad (70)$$

$$= [1.00, 0.00]^\top \quad (71)$$

Forward message at time step 5

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

- Message forward (c_5)

$$c_5 = [1.00, 0.00]^\top \quad (69)$$

- New hidden (h_5)

$$h_5 \quad (70)$$

Forward message at time step 5

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

- Message forward (c_5)

$$c_5 = [1.00, 0.00]^\top \quad (69)$$

- New hidden (h_5)

$$h_5 = o_5 \circ \tanh(c_5) \quad (70)$$

$$(71)$$

Forward message at time step 5

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

- Message forward (c_5)

$$c_5 = [1.00, 0.00]^\top \quad (69)$$

- New hidden (h_5)

$$h_5 = o_5 \circ \tanh(c_5) \quad (70)$$

$$= [1.00, 1.00]^\top \circ \tanh([1.00, 0.00]^\top) \quad (71)$$

$$(72)$$

Forward message at time step 5

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

- Message forward (c_5)

$$c_5 = [1.00, 0.00]^\top \quad (69)$$

- New hidden (h_5)

$$h_5 = o_5 \circ \tanh(c_5) \quad (70)$$

$$= [1.00, 1.00]^\top \circ \tanh([1.00, 0.00]^\top) \quad (71)$$

$$= [0.76, 0.00]^\top \quad (72)$$

Forward message at time step 5

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

- Message forward (c_5)

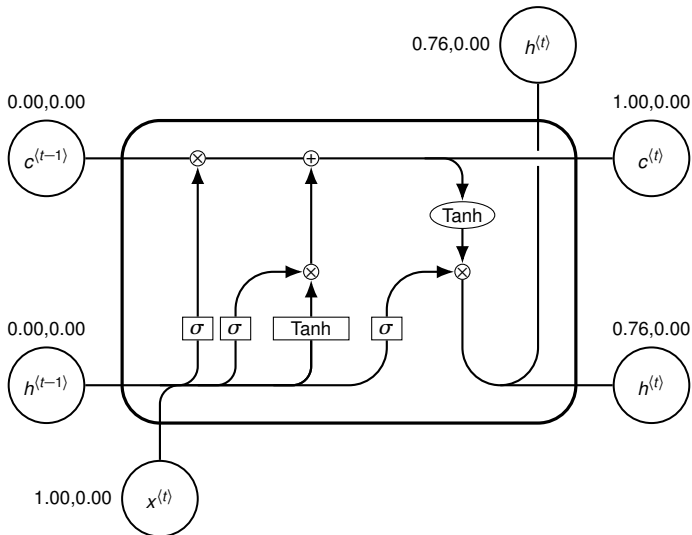
$$c_5 = [1.00, 0.00]^\top \quad (69)$$

- New hidden (h_5)

$$h_5 = [0.76, 0.00]^\top \quad (70)$$

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

Summary at $t = 5$



Input Gate at $t = 6$: i_t

$$W_{ii} = \begin{bmatrix} 30.00 & 0.00 \\ 0.00 & 0.00 \end{bmatrix} \quad b_{ii} = \begin{bmatrix} 0.00 \\ 0.00 \end{bmatrix}$$
$$x^{(6)} = [0.00, 1.00]^\top$$

$$W_{hi} = \begin{bmatrix} 0.00 & 0.00 \\ 60.00 & 0.00 \end{bmatrix} \quad b_{hi} = \begin{bmatrix} 0.00 \\ -30.00 \end{bmatrix}$$
$$h^{(5)} = [0.76, 0.00]^\top$$

Input Gate at $t = 6$: i_t

$$W_{ii} = \begin{bmatrix} 30.00 & 0.00 \\ 0.00 & 0.00 \end{bmatrix} \quad b_{ii} = \begin{bmatrix} 0.00 \\ 0.00 \end{bmatrix} \quad W_{hi} = \begin{bmatrix} 0.00 & 0.00 \\ 60.00 & 0.00 \end{bmatrix} \quad b_{hi} = \begin{bmatrix} 0.00 \\ -30.00 \end{bmatrix}$$
$$x^{(6)} = [0.00, 1.00]^\top \quad h^{(5)} = [0.76, 0.00]^\top$$

$$i^{(6)} = \sigma(W_{ii}x^{(6)} + b_{ii} + W_{hi}h^{(5)} + b_{hi}) \quad (71)$$

$$= \sigma([0.00, 15.70]^\top) \quad (72)$$

$$= [0.50, 1.00]^\top \quad (73)$$

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

$$W_{if} = \begin{bmatrix} 0.00 & 0.00 \\ 0.00 & 0.00 \end{bmatrix} \quad b_{if} = \begin{bmatrix} 0.00 \\ 0.00 \end{bmatrix}$$
$$x^{(6)} = [0.00, 1.00]^\top$$

$$W_{hf} = \begin{bmatrix} 0.00 & 0.00 \\ 0.00 & -30.00 \end{bmatrix} \quad b_{hf} = \begin{bmatrix} -30.00 \\ 0.00 \end{bmatrix}$$
$$h^{(5)} = [0.76, 0.00]^\top$$

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

$$W_{if} = \begin{bmatrix} 0.00 & 0.00 \\ 0.00 & 0.00 \end{bmatrix} \quad b_{if} = \begin{bmatrix} 0.00 \\ 0.00 \end{bmatrix} \quad W_{hf} = \begin{bmatrix} 0.00 & 0.00 \\ 0.00 & -30.00 \end{bmatrix} \quad b_{hf} = \begin{bmatrix} -30.00 \\ 0.00 \end{bmatrix}$$
$$x^{(6)} = [0.00, 1.00]^\top \quad h^{(5)} = [0.76, 0.00]^\top$$

$$f^{(6)} = \sigma(W_{if}x^{(6)} + b_{if} + W_{hf}h^{(5)} + b_{hf}) \quad (74)$$

$$= \sigma([-30.00, -0.00]^\top) \quad (75)$$

$$= [0.00, 0.50]^\top \quad (76)$$

Output Gate at $t = 6$: $o^{(6)}$

$$W_{io} = \begin{bmatrix} 0.00 & 0.00 \\ 0.00 & 0.00 \end{bmatrix} \quad b_{io} = \begin{bmatrix} 30.00 \\ 30.00 \end{bmatrix}$$
$$x^{(6)} = [0.00, 1.00]^\top$$

$$W_{ho} = \begin{bmatrix} 0.00 & 0.00 \\ 0.00 & 0.00 \end{bmatrix} \quad b_{ho} = \begin{bmatrix} 0.00 \\ 0.00 \end{bmatrix}$$
$$h^{(5)} = [0.76, 0.00]^\top$$

Output Gate at $t = 6$: $o^{(6)}$

$$W_{io} = \begin{bmatrix} 0.00 & 0.00 \\ 0.00 & 0.00 \end{bmatrix} \quad b_{io} = \begin{bmatrix} 30.00 \\ 30.00 \end{bmatrix} \quad W_{ho} = \begin{bmatrix} 0.00 & 0.00 \\ 0.00 & 0.00 \end{bmatrix} \quad b_{ho} = \begin{bmatrix} 0.00 \\ 0.00 \end{bmatrix}$$
$$x^{(6)} = [0.00, 1.00]^\top \quad h^{(5)} = [0.76, 0.00]^\top$$

$$o^{(6)} = \sigma(W_{io}x^{(6)} + b_{io} + W_{ho}h^{(5)} + b_{ho}) \quad (77)$$

$$= \sigma([30.00, 30.00]^\top) \quad (78)$$

$$= [1.00, 1.00]^\top \quad (79)$$

Memory Contribution at $t = 6$: $\tilde{c}^{(6)}$

$$W_{i\tilde{c}} = \begin{bmatrix} 30.00 & 0.00 \\ 0.00 & 30.00 \end{bmatrix} \quad b_{i\tilde{c}} = \begin{bmatrix} 0.00 \\ 0.00 \end{bmatrix}$$
$$x^{(6)} = [0.00, 1.00]^\top$$

$$W_{h\tilde{c}} = \begin{bmatrix} 0.00 & 0.00 \\ 0.00 & 0.00 \end{bmatrix} \quad b_{h\tilde{c}} = \begin{bmatrix} 0.00 \\ 0.00 \end{bmatrix}$$
$$h^{(5)} = [0.76, 0.00]^\top$$

Memory Contribution at $t = 6$: $\tilde{c}^{(6)}$

$$W_{i\tilde{c}} = \begin{bmatrix} 30.00 & 0.00 \\ 0.00 & 30.00 \end{bmatrix} \quad b_{i\tilde{c}} = \begin{bmatrix} 0.00 \\ 0.00 \end{bmatrix} \quad W_{h\tilde{c}} = \begin{bmatrix} 0.00 & 0.00 \\ 0.00 & 0.00 \end{bmatrix} \quad b_{h\tilde{c}} = \begin{bmatrix} 0.00 \\ 0.00 \end{bmatrix}$$
$$x^{(6)} = [0.00, 1.00]^\top \quad h^{(5)} = [0.76, 0.00]^\top$$

$$\tilde{c}^{(6)} = \tanh(W_{i\tilde{c}}x^{(6)} + b_{i\tilde{c}} + W_{h\tilde{c}}h^{(5)} + b_{h\tilde{c}}) \quad (80)$$

$$= \tanh([0.00, 30.00]^\top) \quad (81)$$

$$= [0.00, 1.00]^\top \quad (82)$$

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)

$$c_6 = f_6 \circ c_5 + i_6 \circ \tilde{c}_6 \quad (83)$$

$$(84)$$

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)

$$c_6 = f_6 \circ c_5 + i_6 \circ \tilde{c}_6 \quad (83)$$

$$= [0.00, 0.50]^\top \circ [1.00, 0.00]^\top + [0.50, 1.00]^\top \circ [0.00, 1.00]^\top \quad (84)$$

$$(85)$$

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)

$$c_6 = f_6 \circ c_5 + i_6 \circ \tilde{c}_6 \quad (83)$$

$$= [0.00, 0.50]^\top \circ [1.00, 0.00]^\top + [0.50, 1.00]^\top \circ [0.00, 1.00]^\top \quad (84)$$

$$= [0.00, 1.00]^\top \quad (85)$$

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)

$$c_6 = [0.00, 1.00]^\top \quad (83)$$

- New hidden (h_6)

$$h_6 \quad (84)$$

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)

$$c_6 = [0.00, 1.00]^\top \quad (83)$$

- New hidden (h_6)

$$h_6 = o_6 \circ \tanh(c_6) \quad (84)$$

$$(85)$$

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)

$$c_6 = [0.00, 1.00]^\top \quad (83)$$

- New hidden (h_6)

$$h_6 = o_6 \circ \tanh(c_6) \quad (84)$$

$$= [1.00, 1.00]^\top \circ \tanh([0.00, 1.00]^\top) \quad (85)$$

$$(86)$$

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)

$$c_6 = [0.00, 1.00]^\top \quad (83)$$

- New hidden (h_6)

$$h_6 = o_6 \circ \tanh(c_6) \quad (84)$$

$$= [1.00, 1.00]^\top \circ \tanh([0.00, 1.00]^\top) \quad (85)$$

$$= [0.00, 0.76]^\top \quad (86)$$

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)

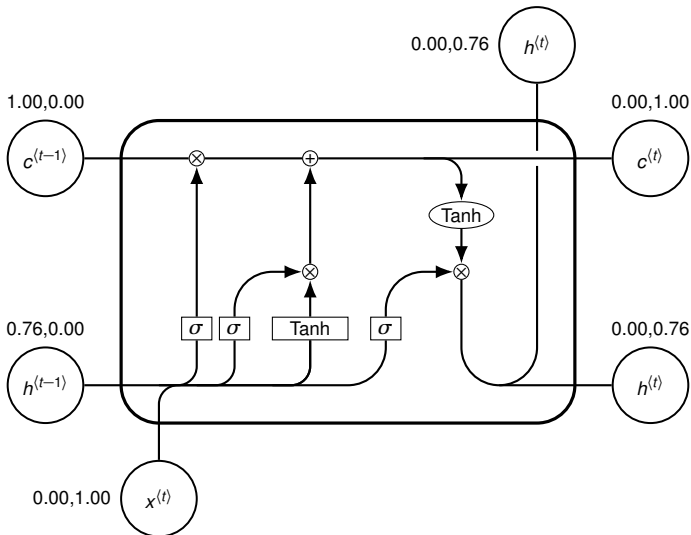
$$c_6 = [0.00, 1.00]^\top \quad (83)$$

- New hidden (h_6)

$$h_6 = [0.00, 0.76]^\top \quad (84)$$

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

Summary at $t = 6$



Input Gate at $t = 7$: i_t

$$W_{ii} = \begin{bmatrix} 30.00 & 0.00 \\ 0.00 & 0.00 \end{bmatrix} \quad b_{ii} = \begin{bmatrix} 0.00 \\ 0.00 \end{bmatrix}$$
$$x^{(7)} = [1.00, 0.00]^\top$$

$$W_{hi} = \begin{bmatrix} 0.00 & 0.00 \\ 60.00 & 0.00 \end{bmatrix} \quad b_{hi} = \begin{bmatrix} 0.00 \\ -30.00 \end{bmatrix}$$
$$h^{(6)} = [0.00, 0.76]^\top$$

Input Gate at $t = 7$: i_t

$$W_{ii} = \begin{bmatrix} 30.00 & 0.00 \\ 0.00 & 0.00 \end{bmatrix} \quad b_{ii} = \begin{bmatrix} 0.00 \\ 0.00 \end{bmatrix} \quad W_{hi} = \begin{bmatrix} 0.00 & 0.00 \\ 60.00 & 0.00 \end{bmatrix} \quad b_{hi} = \begin{bmatrix} 0.00 \\ -30.00 \end{bmatrix}$$
$$x^{(7)} = [1.00, 0.00]^\top \quad h^{(6)} = [0.00, 0.76]^\top$$

$$i^{(7)} = \sigma(W_{ii}x^{(7)} + b_{ii} + W_{hi}h^{(6)} + b_{hi}) \quad (85)$$

$$= \sigma([30.00, -30.00]^\top) \quad (86)$$

$$= [1.00, 0.00]^\top \quad (87)$$

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

$$W_{if} = \begin{bmatrix} 0.00 & 0.00 \\ 0.00 & 0.00 \end{bmatrix} \quad b_{if} = \begin{bmatrix} 0.00 \\ 0.00 \end{bmatrix}$$
$$x^{(7)} = [1.00, 0.00]^\top$$

$$W_{hf} = \begin{bmatrix} 0.00 & 0.00 \\ 0.00 & -30.00 \end{bmatrix} \quad b_{hf} = \begin{bmatrix} -30.00 \\ 0.00 \end{bmatrix}$$
$$h^{(6)} = [0.00, 0.76]^\top$$

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

$$W_{if} = \begin{bmatrix} 0.00 & 0.00 \\ 0.00 & 0.00 \end{bmatrix} \quad b_{if} = \begin{bmatrix} 0.00 \\ 0.00 \end{bmatrix} \quad W_{hf} = \begin{bmatrix} 0.00 & 0.00 \\ 0.00 & -30.00 \end{bmatrix} \quad b_{hf} = \begin{bmatrix} -30.00 \\ 0.00 \end{bmatrix}$$
$$x^{(7)} = [1.00, 0.00]^\top \quad h^{(6)} = [0.00, 0.76]^\top$$

$$f^{(7)} = \sigma(W_{if}x^{(7)} + b_{if} + W_{hf}h^{(6)} + b_{hf}) \quad (88)$$

$$= \sigma([-30.00, -22.85]^\top) \quad (89)$$

$$= [0.00, 0.00]^\top \quad (90)$$

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

$$W_{io} = \begin{bmatrix} 0.00 & 0.00 \\ 0.00 & 0.00 \end{bmatrix} \quad b_{io} = \begin{bmatrix} 30.00 \\ 30.00 \end{bmatrix}$$
$$x^{(7)} = [1.00, 0.00]^\top$$

$$W_{ho} = \begin{bmatrix} 0.00 & 0.00 \\ 0.00 & 0.00 \end{bmatrix} \quad b_{ho} = \begin{bmatrix} 0.00 \\ 0.00 \end{bmatrix}$$
$$h^{(6)} = [0.00, 0.76]^\top$$

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

$$W_{io} = \begin{bmatrix} 0.00 & 0.00 \\ 0.00 & 0.00 \end{bmatrix} \quad b_{io} = \begin{bmatrix} 30.00 \\ 30.00 \end{bmatrix} \quad W_{ho} = \begin{bmatrix} 0.00 & 0.00 \\ 0.00 & 0.00 \end{bmatrix} \quad b_{ho} = \begin{bmatrix} 0.00 \\ 0.00 \end{bmatrix}$$
$$x^{(7)} = [1.00, 0.00]^\top \quad h^{(6)} = [0.00, 0.76]^\top$$

$$o^{(7)} = \sigma(W_{io}x^{(7)} + b_{io} + W_{ho}h^{(6)} + b_{ho}) \quad (91)$$

$$= \sigma([30.00, 30.00]^\top) \quad (92)$$

$$= [1.00, 1.00]^\top \quad (93)$$

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

$$W_{i\tilde{c}} = \begin{bmatrix} 30.00 & 0.00 \\ 0.00 & 30.00 \end{bmatrix} \quad b_{i\tilde{c}} = \begin{bmatrix} 0.00 \\ 0.00 \end{bmatrix}$$
$$x^{(7)} = [1.00, 0.00]^\top$$

$$W_{h\tilde{c}} = \begin{bmatrix} 0.00 & 0.00 \\ 0.00 & 0.00 \end{bmatrix} \quad b_{h\tilde{c}} = \begin{bmatrix} 0.00 \\ 0.00 \end{bmatrix}$$
$$h^{(6)} = [0.00, 0.76]^\top$$

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

$$W_{i\tilde{c}} = \begin{bmatrix} 30.00 & 0.00 \\ 0.00 & 30.00 \end{bmatrix} \quad b_{i\tilde{c}} = \begin{bmatrix} 0.00 \\ 0.00 \end{bmatrix} \quad W_{h\tilde{c}} = \begin{bmatrix} 0.00 & 0.00 \\ 0.00 & 0.00 \end{bmatrix} \quad b_{h\tilde{c}} = \begin{bmatrix} 0.00 \\ 0.00 \end{bmatrix}$$
$$x^{(7)} = [1.00, 0.00]^\top \quad h^{(6)} = [0.00, 0.76]^\top$$

$$\tilde{c}^{(7)} = \tanh(W_{i\tilde{c}}x^{(7)} + b_{i\tilde{c}} + W_{h\tilde{c}}h^{(6)} + b_{h\tilde{c}}) \quad (94)$$

$$= \tanh([30.00, 0.00]^\top) \quad (95)$$

$$= [1.00, 0.00]^\top \quad (96)$$

Forward message at time step 7

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

- Message forward (c_7)

$$c_7 = f_7 \circ c_6 + i_7 \circ \tilde{c}_7 \quad (97)$$

$$(98)$$

Forward message at time step 7

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

■ Message forward (c_7)

$$c_7 = f_7 \circ c_6 + i_7 \circ \tilde{c}_7 \quad (97)$$

$$= [0.00, 0.00]^\top \circ [0.00, 1.00]^\top + [1.00, 0.00]^\top \circ [1.00, 0.00]^\top \quad (98)$$

$$(99)$$

Forward message at time step 7

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

■ Message forward (c_7)

$$c_7 = f_7 \circ c_6 + i_7 \circ \tilde{c}_7 \quad (97)$$

$$= [0.00, 0.00]^\top \circ [0.00, 1.00]^\top + [1.00, 0.00]^\top \circ [1.00, 0.00]^\top \quad (98)$$

$$= [1.00, 0.00]^\top \quad (99)$$

Forward message at time step 7

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

- Message forward (c_7)

$$c_7 = [1.00, 0.00]^\top \quad (97)$$

- New hidden (h_7)

$$h_7 \quad (98)$$

Forward message at time step 7

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

- Message forward (c_7)

$$c_7 = [1.00, 0.00]^\top \quad (97)$$

- New hidden (h_7)

$$h_7 = o_7 \circ \tanh(c_7) \quad (98)$$

$$(99)$$

Forward message at time step 7

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

- Message forward (c_7)

$$c_7 = [1.00, 0.00]^\top \quad (97)$$

- New hidden (h_7)

$$h_7 = o_7 \circ \tanh(c_7) \quad (98)$$

$$= [1.00, 1.00]^\top \circ \tanh([1.00, 0.00]^\top) \quad (99)$$

$$(100)$$

Forward message at time step 7

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

- Message forward (c_7)

$$c_7 = [1.00, 0.00]^\top \quad (97)$$

- New hidden (h_7)

$$h_7 = o_7 \circ \tanh(c_7) \quad (98)$$

$$= [1.00, 1.00]^\top \circ \tanh([1.00, 0.00]^\top) \quad (99)$$

$$= [0.76, 0.00]^\top \quad (100)$$

Forward message at time step 7

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

- Message forward (c_7)

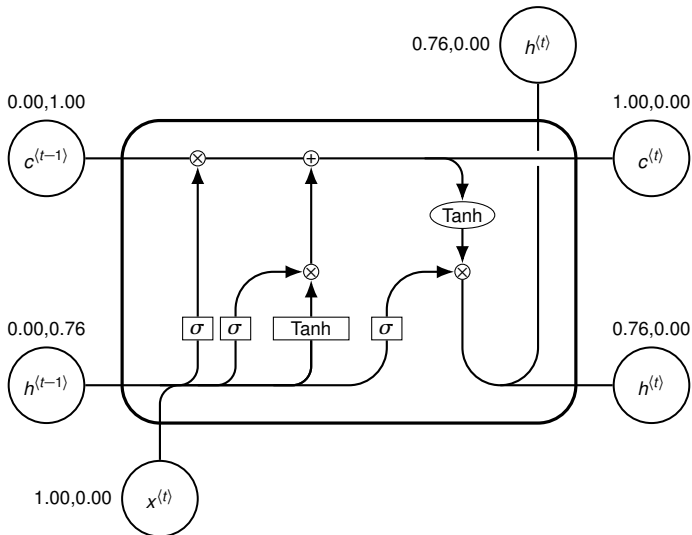
$$c_7 = [1.00, 0.00]^\top \quad (97)$$

- New hidden (h_7)

$$h_7 = [0.76, 0.00]^\top \quad (98)$$

- Prediction $y_7 = \text{softmax}(h_7) = 0$

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?