COMP7/8118 M50

Data Mining

RNN: LSTM

Xiaofei Zhang

Slides compiled from Jiawei Han and Raymond C.W. Wong's work

THE UNIVERSITY OF MEMPHIS

Traditional LSTM (Long Short Term Memory)

- Disadvantage of Basic RNN
  - The basic RNN model is too “simple”.
  - It could not simulate our human brain well.
  - It is not easy for the basic RNN model to converge (i.e., it may take a very long time to train the RNN model)

Traditional LSTM

- Before we give the details of our brain, we want to emphasize that there is an internal state variable (i.e., variable s_t) to store our memory (i.e., a value)
- The next RNN to be described is called the **LSTM (Long Short-Term Memory)** model.

Traditional LSTM

- It could simulate the brain process.
- **Forget Feature**
  - It could “decide” to forget a portion of the internal state variable.
- **Input Feature**
  - It could “decide” to input a portion of the input variable for the model
  - It could “decide” the strength of the input for the model (i.e., the activation function) (called the “weight” of the input)

Traditional LSTM

- **Output Feature**
  - It could “decide” to output a portion of the output for the model
  - It could “decide” the strength of the output for the model (i.e., the activation function) (called the “weight” of the output)

Traditional LSTM

- Our brain includes the following steps.
  - Forget component
  - Input component
  - Input activation component
  - Internal state component
  - Output component
  - Final output component

Forget gate

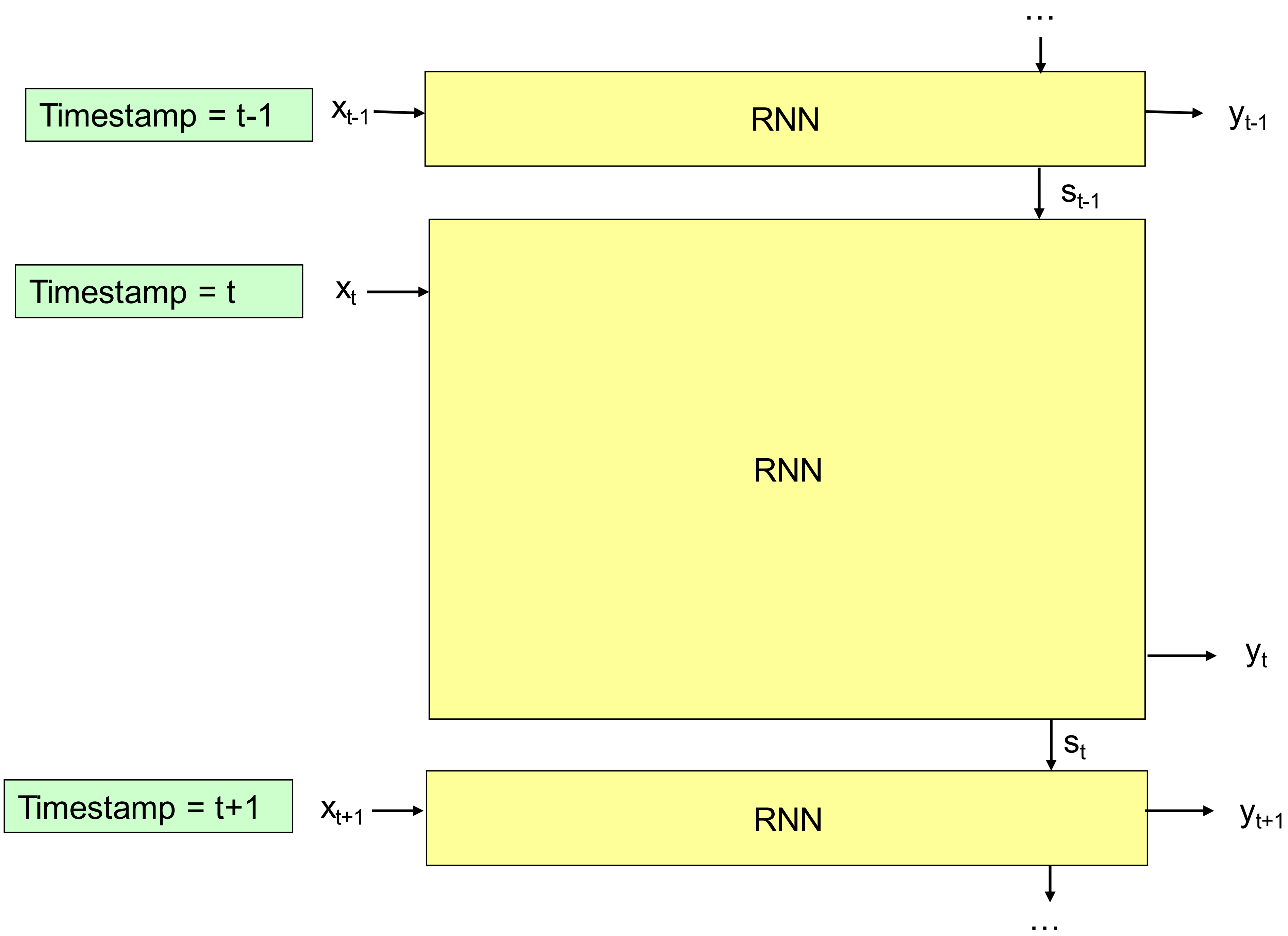
Input gate

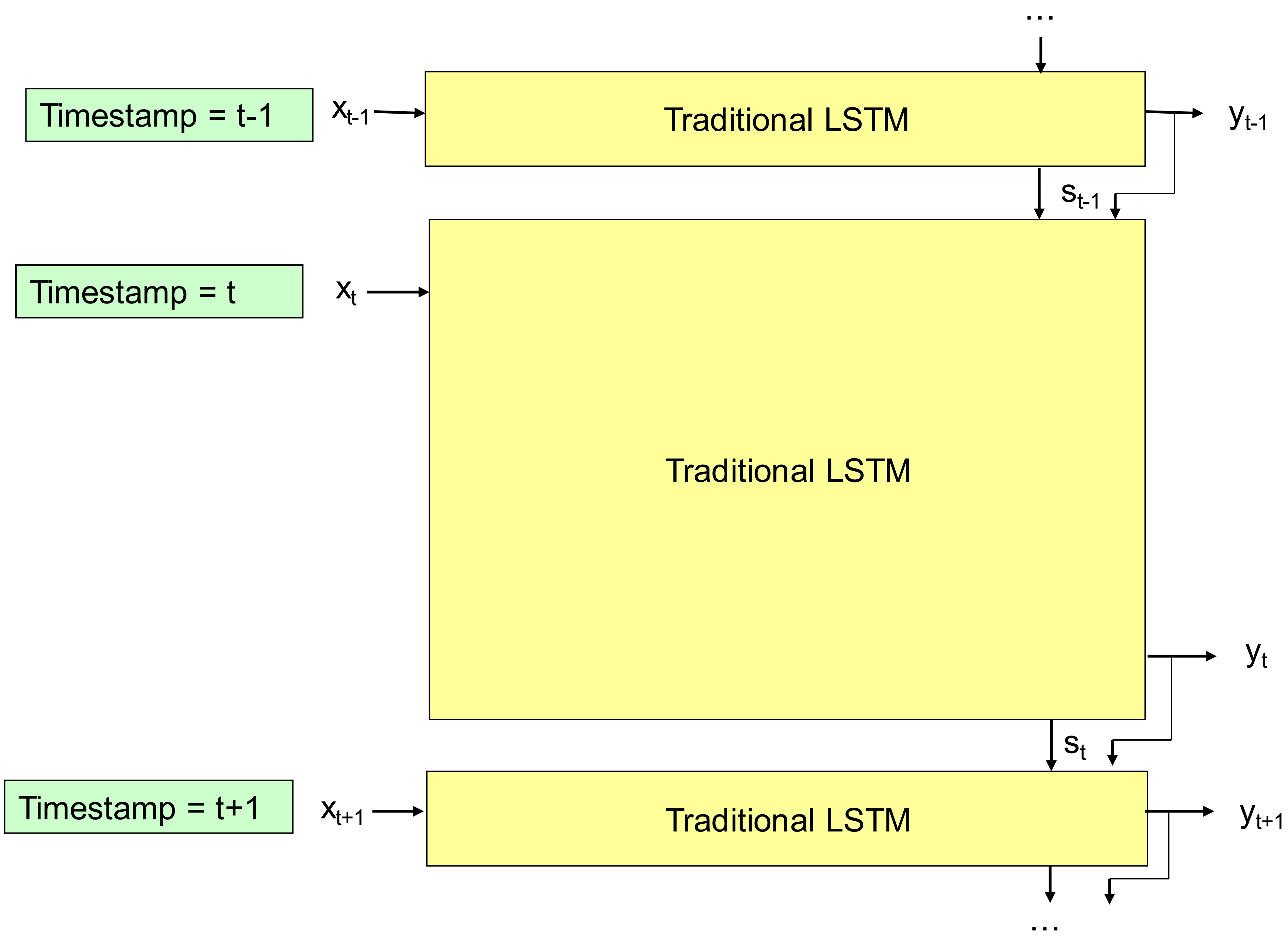
Input activation gate

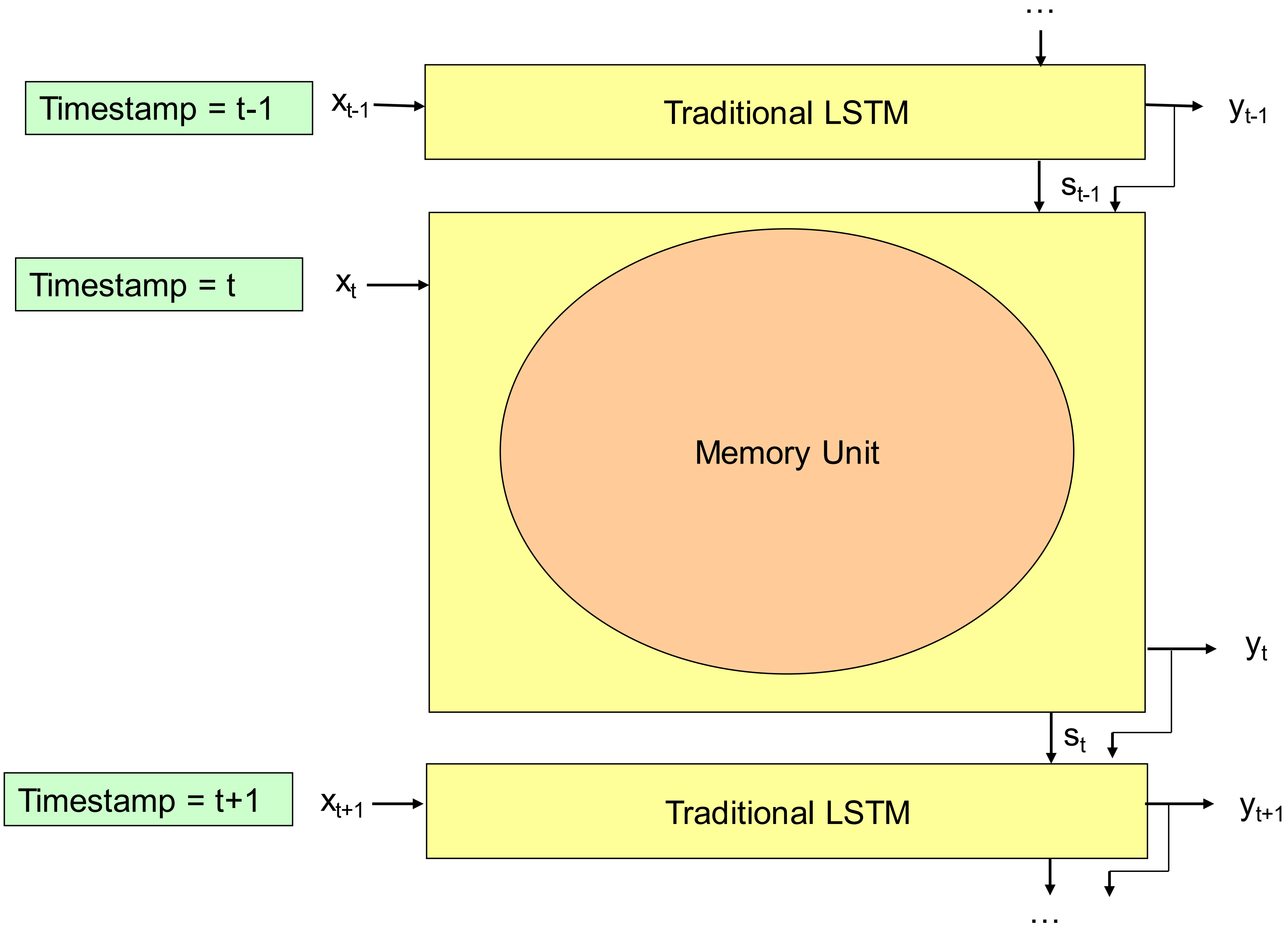
Input state gate

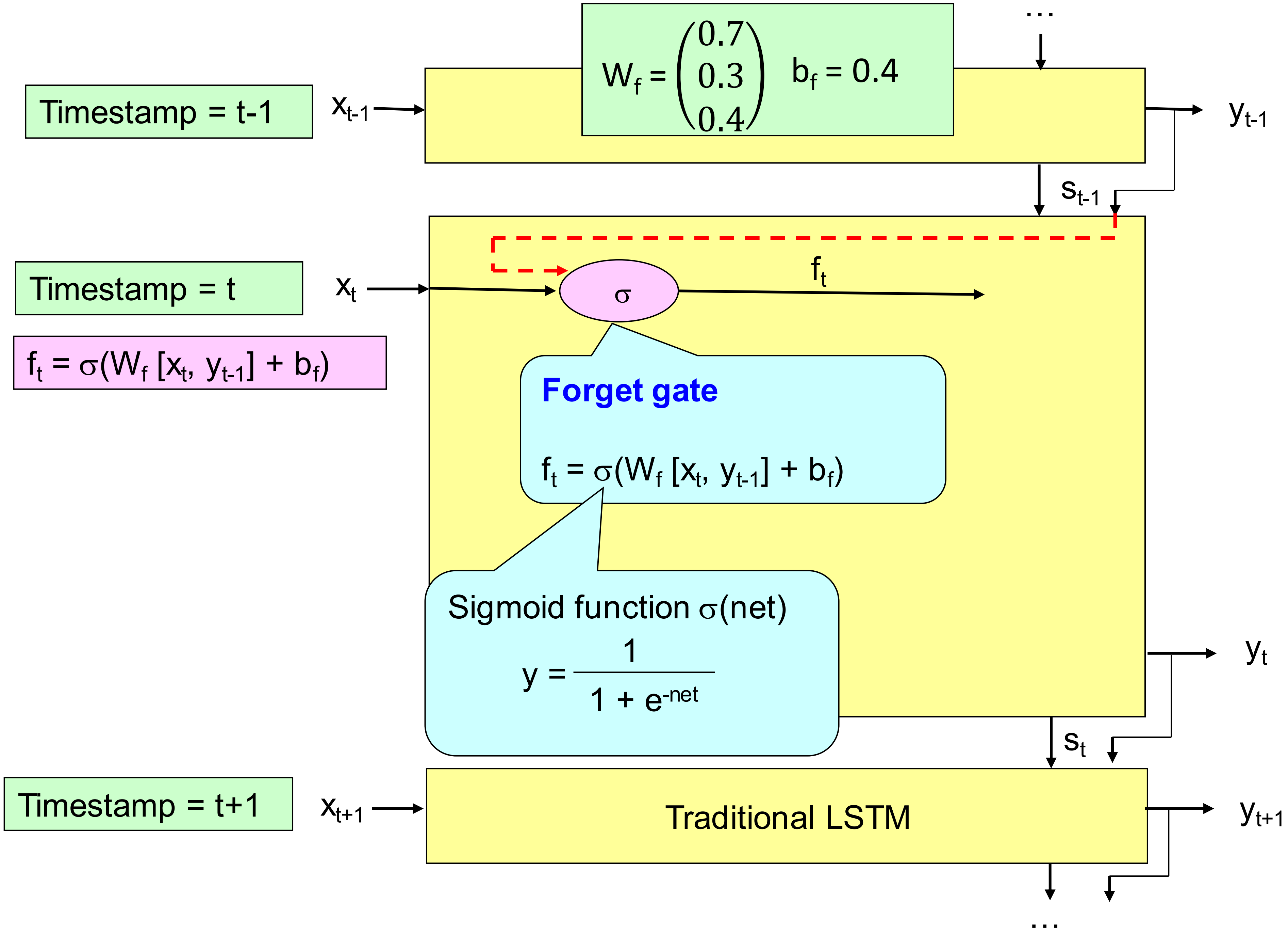
Output gate

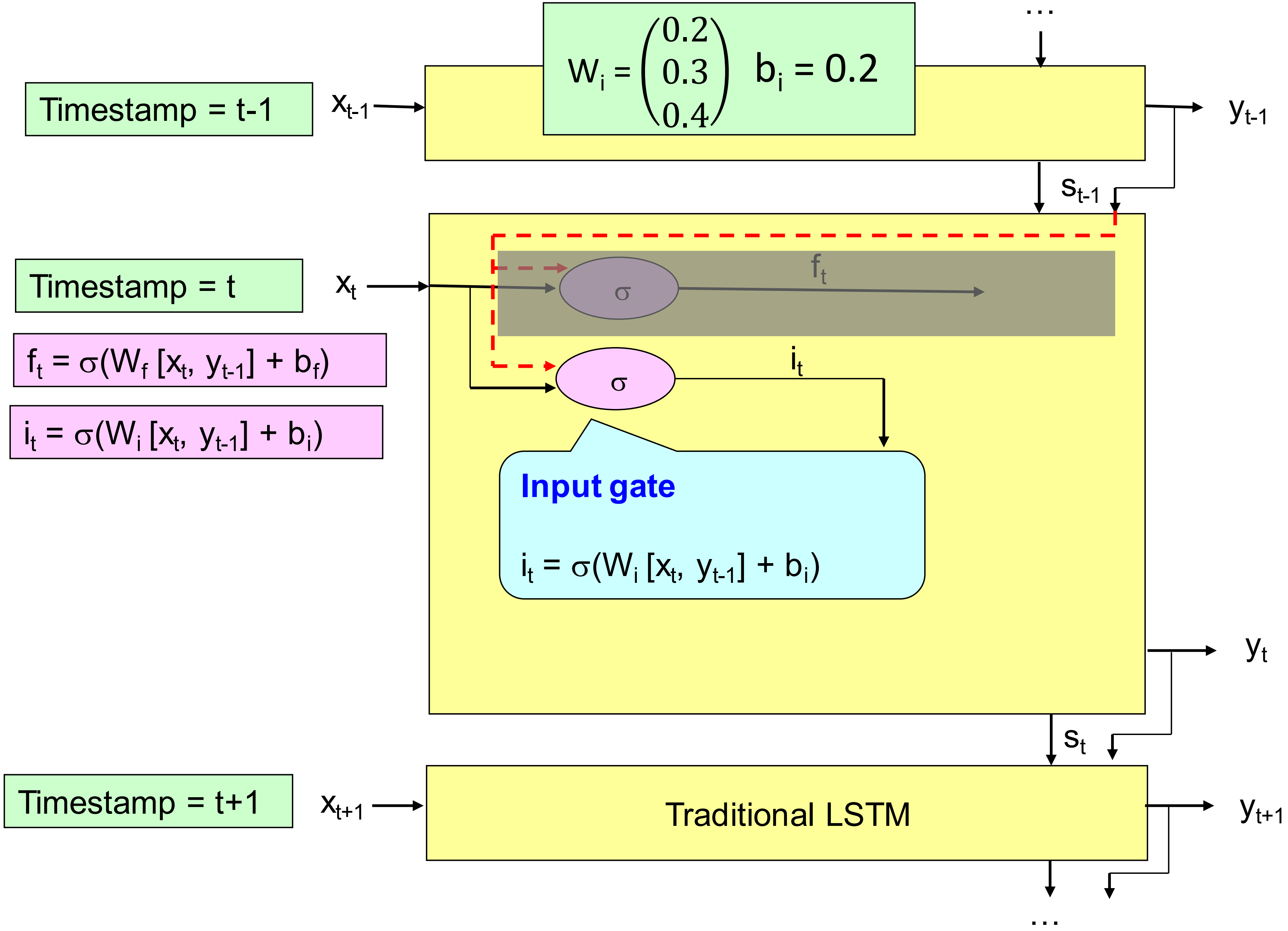
Final output gate


...
Timestamp = t-1
x_{t-1}
RNN
y_{t-1}
s_{t-1}
Timestamp = t
x_t
RNN
y_t
s_t
Timestamp = t+1
x_{t+1}
RNN
y_{t+1}
...


...
Timestamp = t-1
x_{t-1}
Traditional LSTM
y_{t-1}
s_{t-1}
Timestamp = t
x_t
Traditional LSTM
y_t
s_t
Timestamp = t+1
x_{t+1}
Traditional LSTM
y_{t+1}
...


Timestamp = t-1
Timestamp = t
Timestamp = t+1
x_{t-1}
x_t
x_{t+1}
Traditional LSTM
Memory Unit
Traditional LSTM
y_{t-1}
y_t
y_{t+1}
s_{t-1}
s_t
…
…


Timestamp = t-1
x_{t-1}
$W_f = \begin{pmatrix} 0.7 \\ 0.3 \\ 0.4 \end{pmatrix}$ $b_f = 0.4$
…
y_{t-1}
s_{t-1}
Timestamp = t
x_t
σ
f_t
$f_t = \sigma(W_f\,[x_t, y_{t-1}] + b_f)$
Forget gate
$f_t = \sigma(W_f\,[x_t, y_{t-1}] + b_f)$
Sigmoid function σ(net)
$y = \dfrac{1}{1 + e^{-net}}$
y_t
s_t
Timestamp = t+1
x_{t+1}
Traditional LSTM
y_{t+1}
…



Timestamp = t-1
x_{t-1}
$W_i = \begin{pmatrix} 0.2 \\ 0.3 \\ 0.4 \end{pmatrix}$ $b_i = 0.2$
...
y_{t-1}
s_{t-1}
Timestamp = t
x_t
$f_t = \sigma(W_f [x_t, y_{t-1}] + b_f)$
$i_t = \sigma(W_i [x_t, y_{t-1}] + b_i)$
σ
f_t
σ
i_t
Input gate
$i_t = \sigma(W_i [x_t, y_{t-1}] + b_i)$
y_t
s_t
Timestamp = t+1
x_{t+1}
Traditional LSTM
y_{t+1}
...

Timestamp = t-1

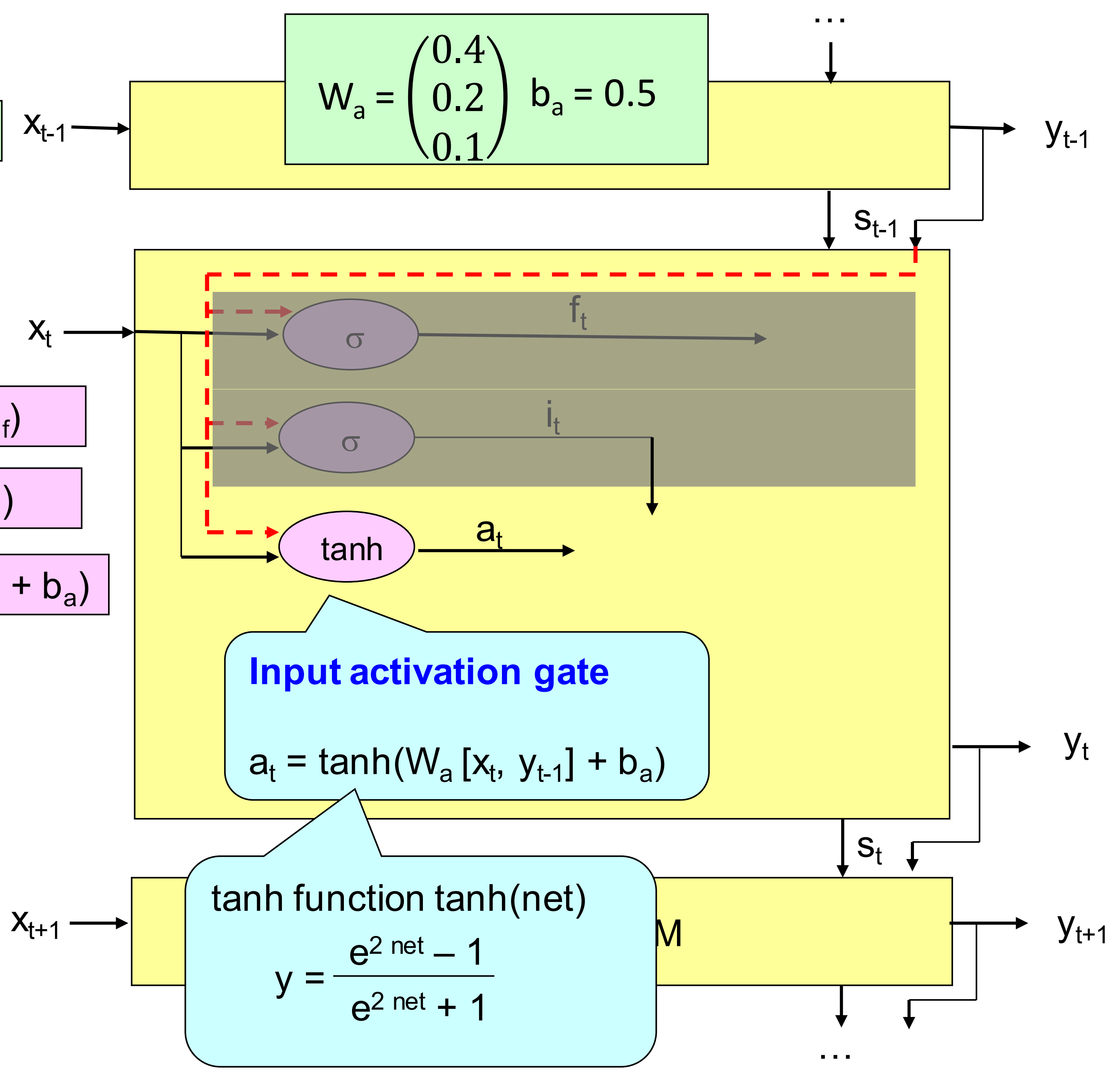
Timestamp = t

$f_t = \sigma(W_f\,[x_t,\ y_{t-1}] + b_f)$

$i_t = \sigma(W_i\,[x_t,\ y_{t-1}] + b_i)$

$a_t = \tanh(W_a\,[x_t,\ y_{t-1}] + b_a)$

Timestamp = t+1



Timestamp = t-1

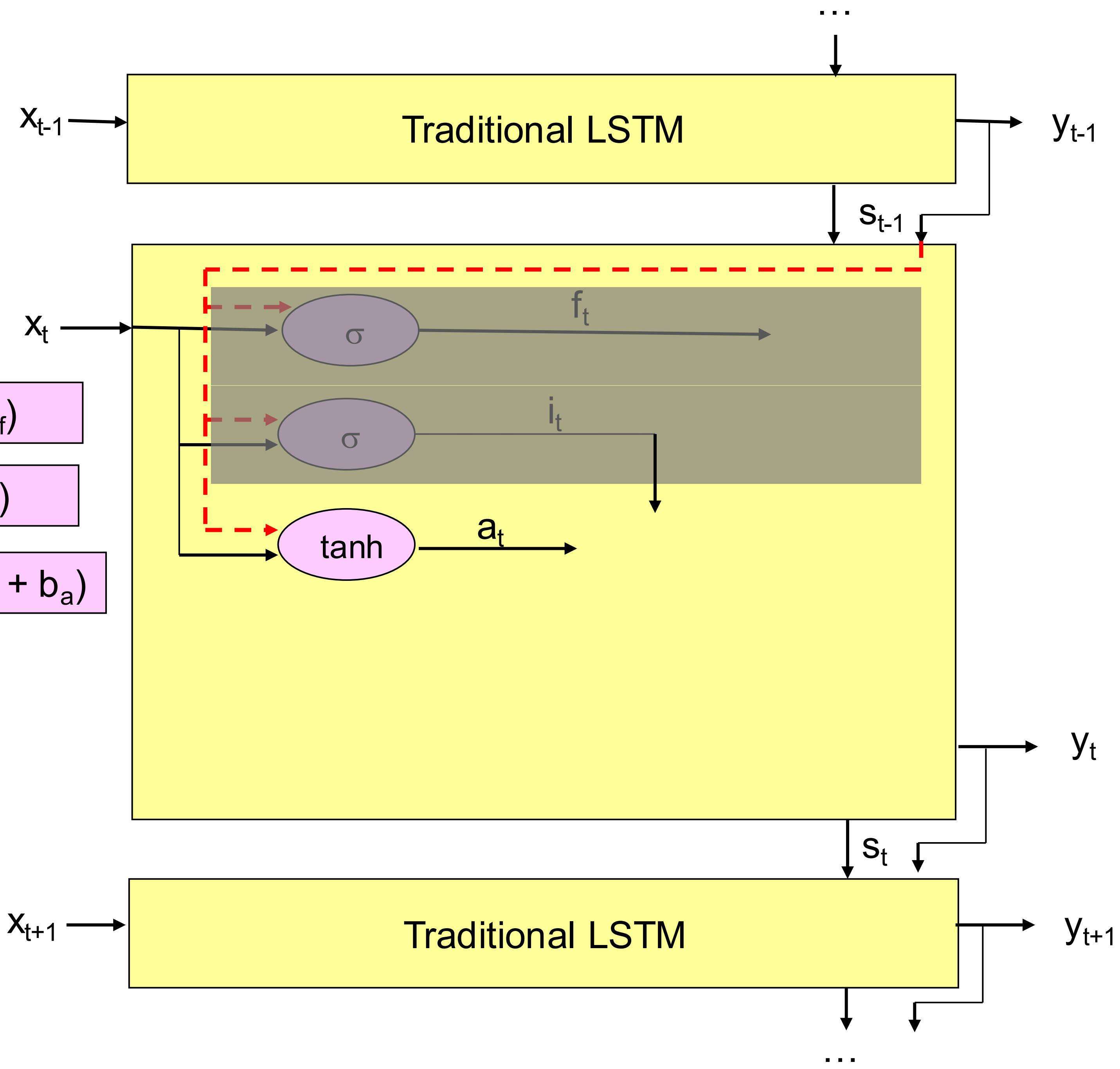
Timestamp = t

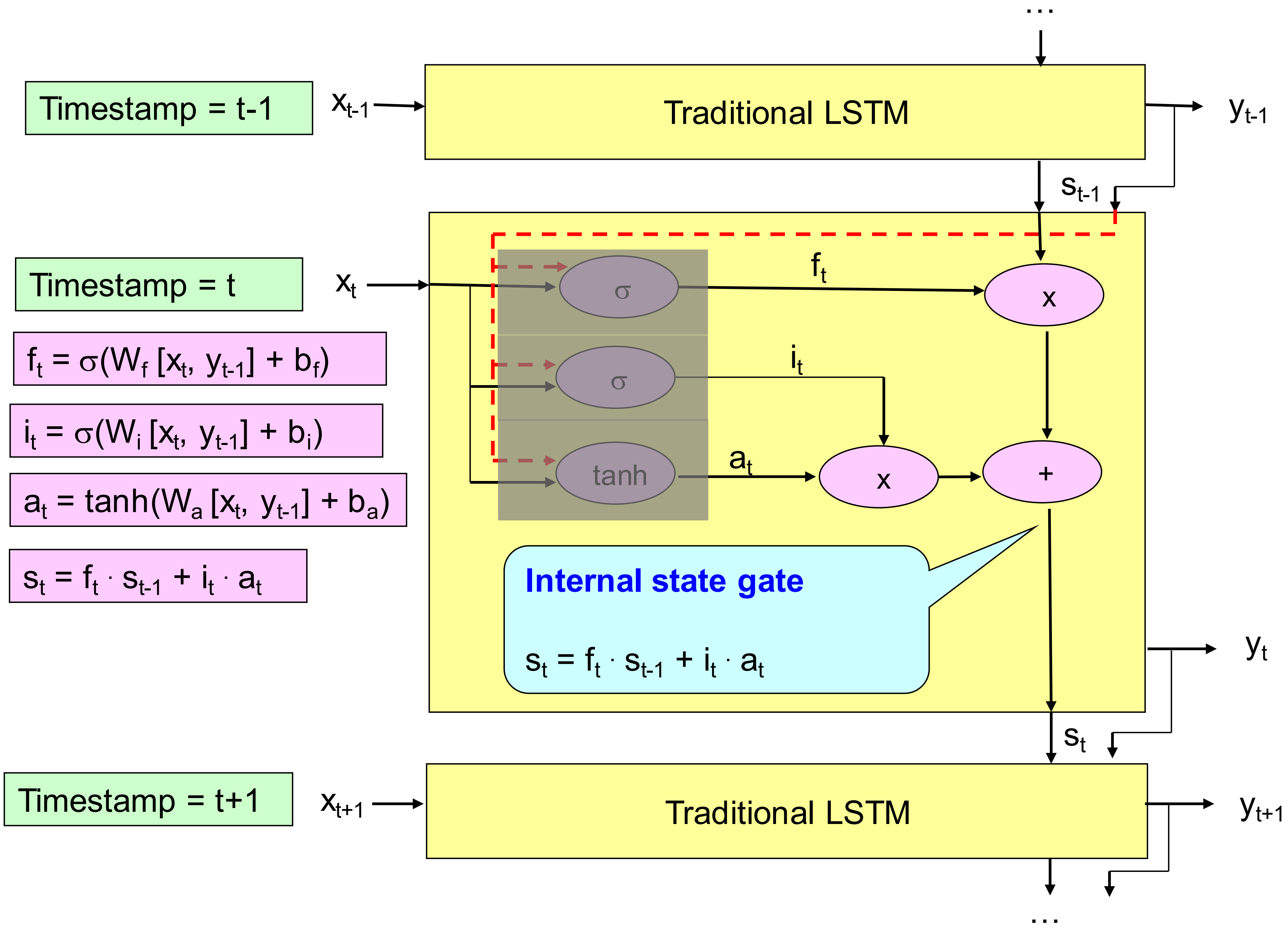
$f_t = \sigma(W_f\,[x_t,\ y_{t-1}] + b_f)$

$i_t = \sigma(W_i\,[x_t,\ y_{t-1}] + b_i)$

$a_t = \tanh(W_a\,[x_t,\ y_{t-1}] + b_a)$

Timestamp = t+1




Timestamp = t-1
x_{t-1}
Traditional LSTM
y_{t-1}
s_{t-1}
Timestamp = t
x_t
$f_t = \sigma(W_f [x_t, y_{t-1}] + b_f)$
$i_t = \sigma(W_i [x_t, y_{t-1}] + b_i)$
$a_t = \tanh(W_a [x_t, y_{t-1}] + b_a)$
$s_t = f_t \cdot s_{t-1} + i_t \cdot a_t$
σ
σ
tanh
f_t
i_t
a_t
x
x
+
Internal state gate
$s_t = f_t \cdot s_{t-1} + i_t \cdot a_t$
y_t
s_t
Timestamp = t+1
x_{t+1}
Traditional LSTM
y_{t+1}
…
…

Timestamp = t-1

Timestamp = t

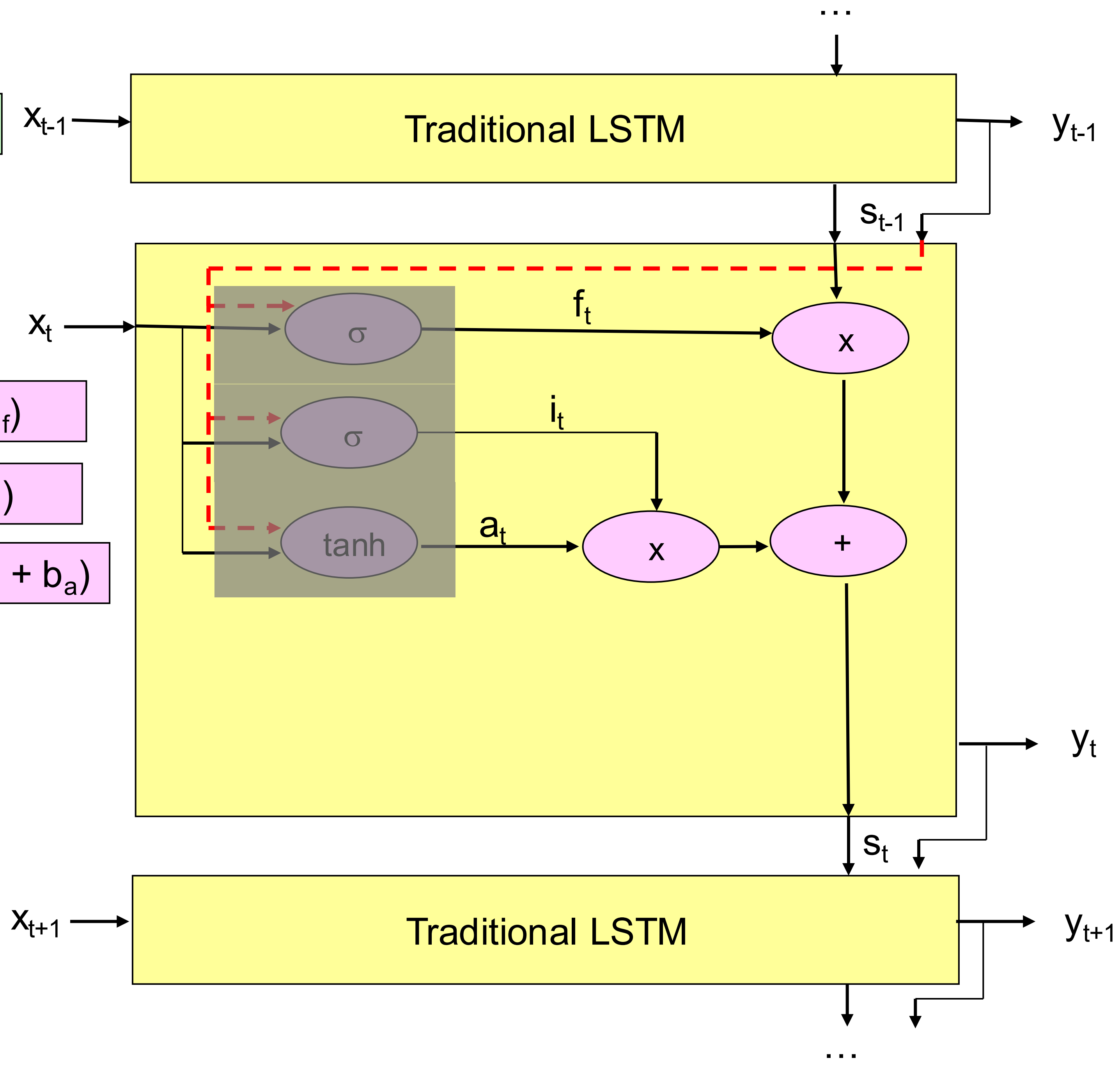
$f_t = \sigma(W_f\,[x_t,\ y_{t-1}] + b_f)$

$i_t = \sigma(W_i\,[x_t,\ y_{t-1}] + b_i)$

$a_t = \tanh(W_a\,[x_t,\ y_{t-1}] + b_a)$

$s_t = f_t \cdot s_{t-1} + i_t \cdot a_t$

Timestamp = t+1



Timestamp = t-1

Timestamp = t

$f_t = \sigma(W_f\,[x_t,\ y_{t-1}] + b_f)$

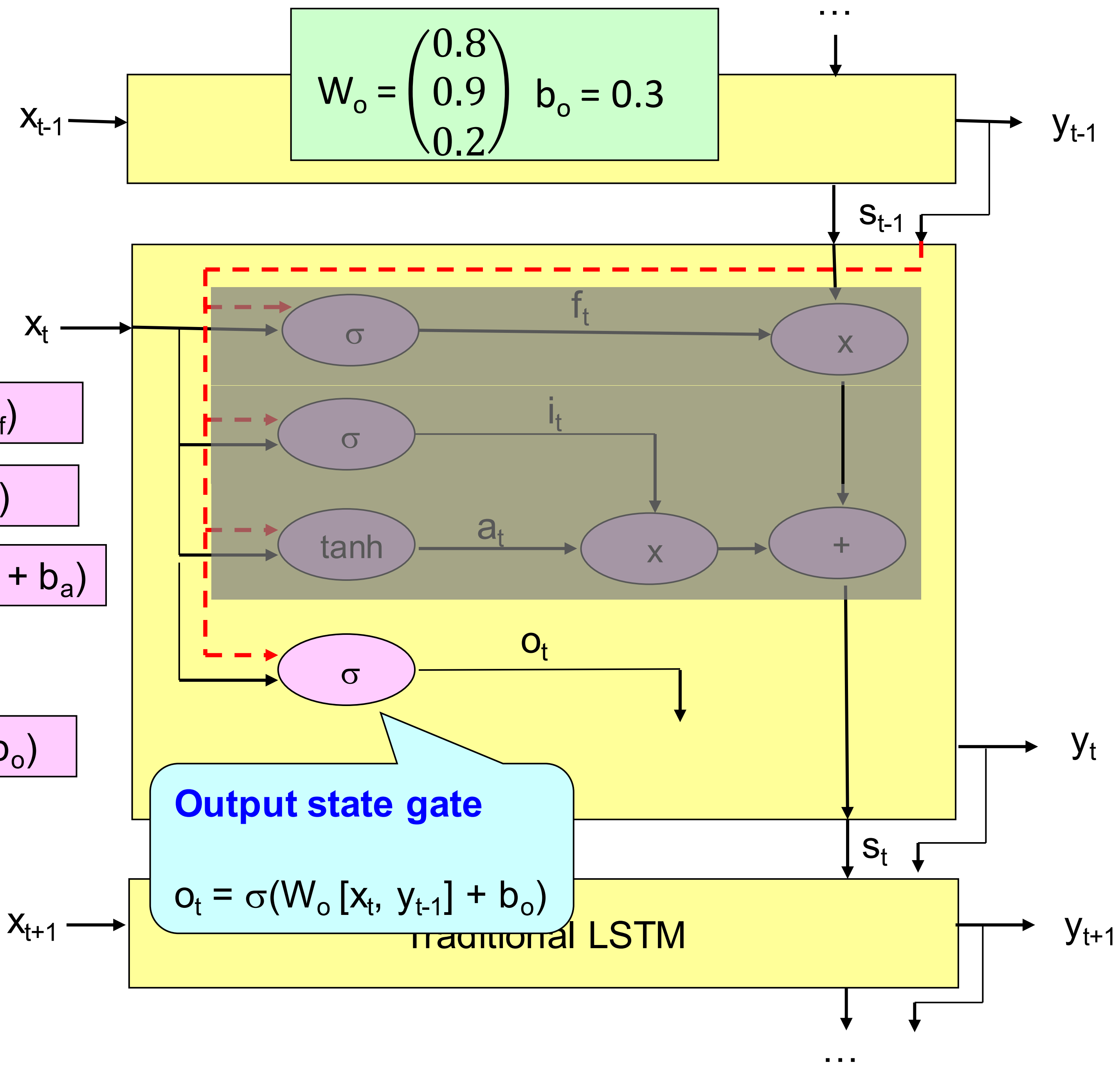
$i_t = \sigma(W_i\,[x_t,\ y_{t-1}] + b_i)$

$a_t = \tanh(W_a\,[x_t,\ y_{t-1}] + b_a)$

$s_t = f_t \cdot s_{t-1} + i_t \cdot a_t$

$o_t = \sigma(W_o\,[x_t,\ y_{t-1}] + b_o)$

Timestamp = t+1



Timestamp = t-1

Timestamp = t

$f_t = \sigma(W_f\,[x_t,\ y_{t-1}] + b_f)$

$i_t = \sigma(W_i\,[x_t,\ y_{t-1}] + b_i)$

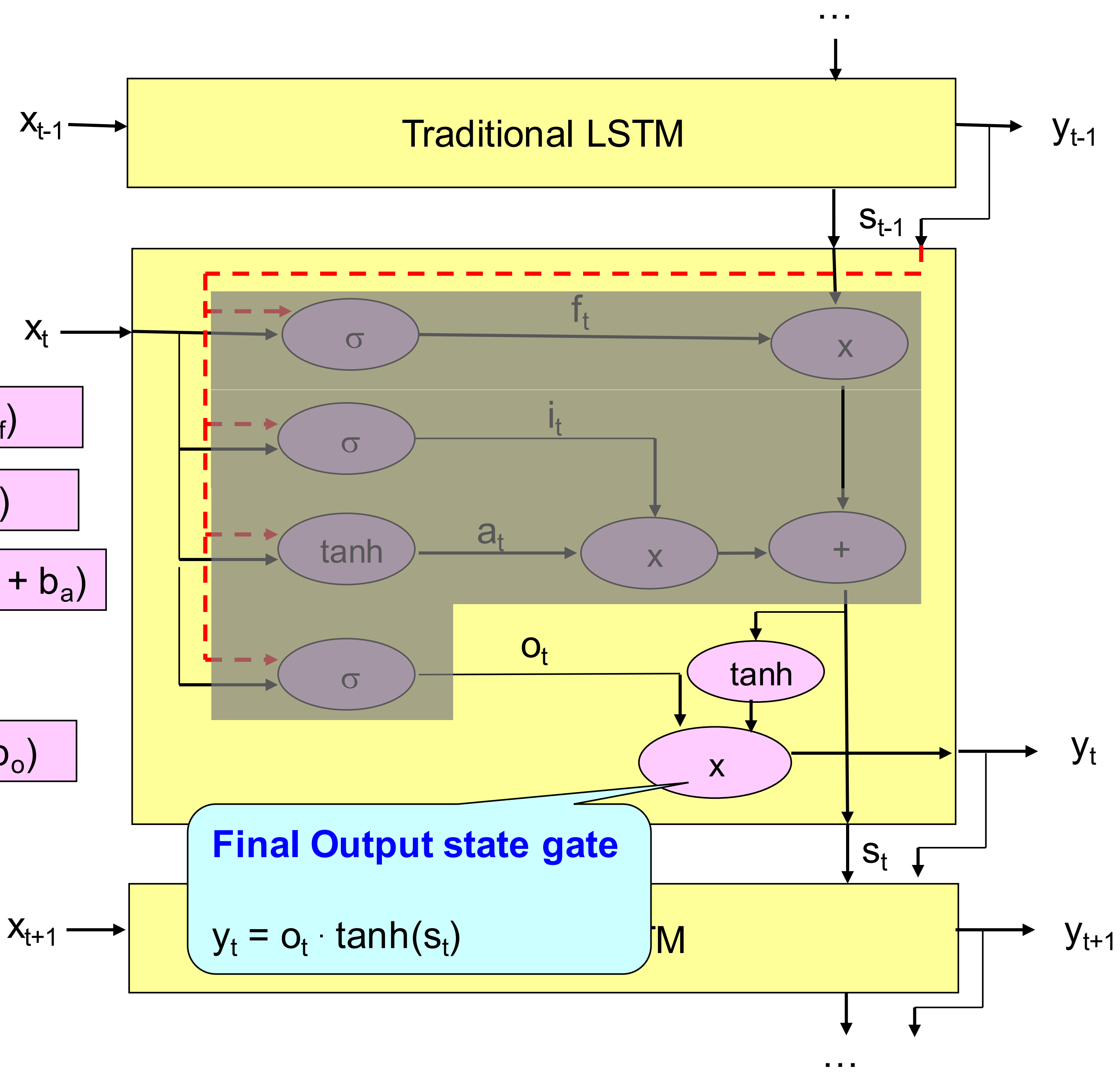
$a_t = \tanh(W_a\,[x_t,\ y_{t-1}] + b_a)$

$s_t = f_t \cdot s_{t-1} + i_t \cdot a_t$

$o_t = \sigma(W_o\,[x_t,\ y_{t-1}] + b_o)$

$y_t = o_t \cdot \tanh(s_t)$

Timestamp = t+1



Timestamp = t-1

Timestamp = t

$f_t = \sigma(W_f\,[x_t,\ y_{t-1}] + b_f)$

$i_t = \sigma(W_i\,[x_t,\ y_{t-1}] + b_i)$

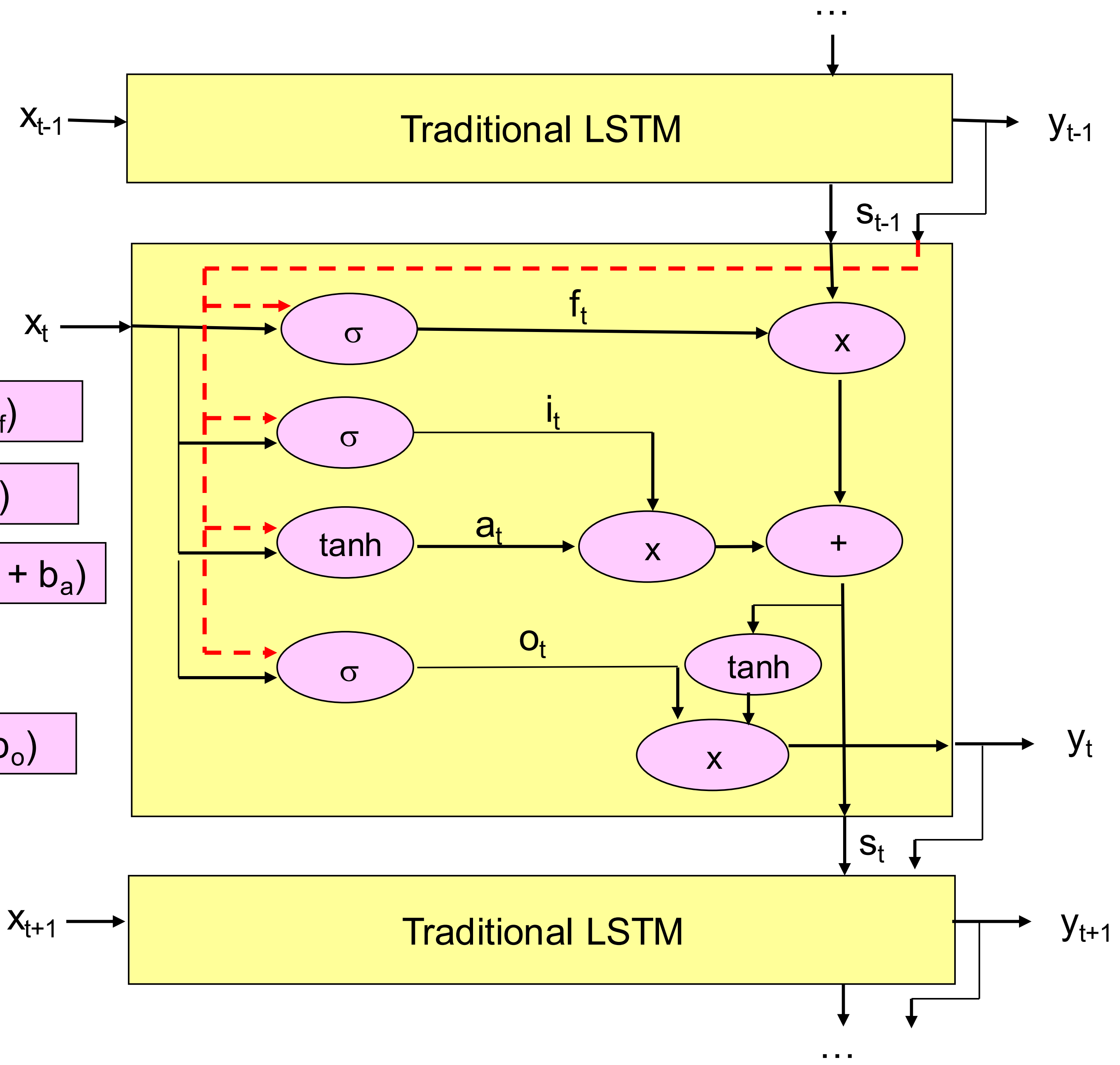
$a_t = \tanh(W_a\,[x_t,\ y_{t-1}] + b_a)$

$s_t = f_t \cdot s_{t-1} + i_t \cdot a_t$

$o_t = \sigma(W_o\,[x_t,\ y_{t-1}] + b_o)$

$y_t = o_t \cdot \tanh(s_t)$

Timestamp = t+1

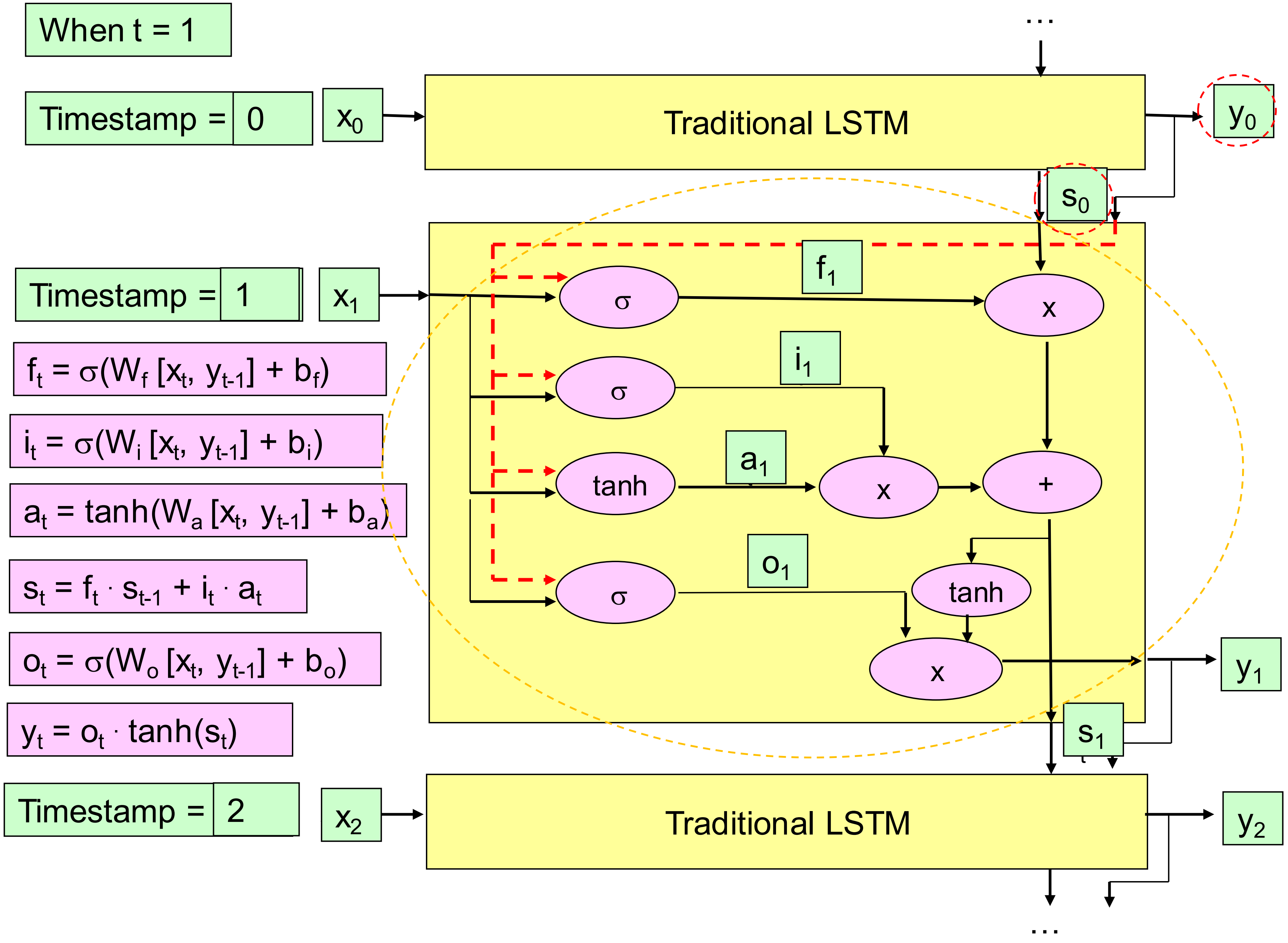


Traditional LSTM

- In the following, we want to compute (weight) values in the traditional LSTM.
- Similar to the neural network, the traditional LSTM model has two steps.
  - **Step 1** (Input Forward Propagation)
  - **Step 2** (Error Backward Propagation)
- In the following, we focus on “Input Forward Propagation”.
- In the traditional LSTM, “Error Backward Propagation” could be solved by an existing optimization tool (like “Neural Network”).

Time	$x_{t,1}$	$x_{t,2}$	y
t=1	0.1	0.4	0.3
t=2	0.7	0.9	0.5

- Consider this example with two timestamps.
  - When t = 1
  - When t = 2
- We use the traditional LSTM to do the training.


When t = 1
Timestamp = 0
x_0
Traditional LSTM
y_0
s_0
Timestamp = 1
x_1
$f_t = \sigma(W_f [x_t, y_{t-1}] + b_f)$
$i_t = \sigma(W_i [x_t, y_{t-1}] + b_i)$
$a_t = \tanh(W_a [x_t, y_{t-1}] + b_a)$
$s_t = f_t \cdot s_{t-1} + i_t \cdot a_t$
$o_t = \sigma(W_o [x_t, y_{t-1}] + b_o)$
$y_t = o_t \cdot \tanh(s_t)$
f_1
i_1
a_1
o_1
σ
tanh
x
+
y_1
s_1
Timestamp = 2
x_2
Traditional LSTM
y_2
…

$f_t = \sigma(W_f\,[x_t,\ y_{t-1}] + b_f)$

$s_t = f_t \cdot s_{t-1} + i_t \cdot a_t$

$i_t = \sigma(W_i\,[x_t,\ y_{t-1}] + b_i)$

$o_t = \sigma(W_o\,[x_t,\ y_{t-1}] + b_o)$

$a_t = \tanh(W_a\,[x_t,\ y_{t-1}] + b_a)$

$y_t = o_t \cdot \tanh(s_t)$

Time	$x_{t,1}$	$x_{t,2}$	y
t=1	0.1	0.4	0.3
t=2	0.7	0.9	0.5

y_0	s_0
0	0

Step 1 (Input Forward Propagation)

$$f_1 = \sigma(W_f\,[x_1,\ y_0] + b_f)$$

$$= \sigma(\begin{pmatrix}0.7\\0.3\\0.4\end{pmatrix}\begin{pmatrix}0.1\\0.4\\0\end{pmatrix} + 0.4)$$

$$= \sigma(0.7 \cdot 0.1 + 0.3 \cdot 0.4 + 0.4 \cdot 0 + 0.4)$$

$$= \sigma(0.59)$$

$$= 0.6434$$

$f_1 = 0.6434$

$W_f = \begin{pmatrix}0.7\\0.3\\0.4\end{pmatrix}$ $b_f = 0.4$

$W_i = \begin{pmatrix}0.2\\0.3\\0.4\end{pmatrix}$ $b_i = 0.2$

$W_a = \begin{pmatrix}0.4\\0.2\\0.1\end{pmatrix}$ $b_a = 0.5$

$W_o = \begin{pmatrix}0.8\\0.9\\0.2\end{pmatrix}$ $b_o = 0.3$

$f_t = \sigma(W_f\,[x_t,\ y_{t-1}] + b_f)$

$i_t = \sigma(W_i\,[x_t,\ y_{t-1}] + b_i)$

$a_t = \tanh(W_a\,[x_t,\ y_{t-1}] + b_a)$

$s_t = f_t \cdot s_{t-1} + i_t \cdot a_t$

$o_t = \sigma(W_o\,[x_t,\ y_{t-1}] + b_o)$

$y_t = o_t \cdot \tanh(s_t)$

Time	$x_{t,1}$	$x_{t,2}$	y
t=1	0.1	0.4	0.3
t=2	0.7	0.9	0.5

y_0	s_0
0	0

Step 1 (Input Forward Propagation)

$i_1 = \sigma(W_i\,[x_1,\ y_0] + b_i)$

$= \sigma(\begin{pmatrix}0.2\\0.3\\0.4\end{pmatrix}\begin{pmatrix}0.1\\0.4\\0\end{pmatrix} + 0.2)$

$= \sigma(0.2 \cdot 0.1 + 0.3 \cdot 0.4 + 0.4 \cdot 0 + 0.2)$

$= \sigma(0.34)$

$= 0.5842$

$f_1 = 0.6434$

$i_1 \ = 0.5842$

$W_f = \begin{pmatrix}0.7\\0.3\\0.4\end{pmatrix}$ $b_f = 0.4$

$W_i = \begin{pmatrix}0.2\\0.3\\0.4\end{pmatrix}$ $b_i = 0.2$

$W_a = \begin{pmatrix}0.4\\0.2\\0.1\end{pmatrix}$ $b_a = 0.5$

$W_o = \begin{pmatrix}0.8\\0.9\\0.2\end{pmatrix}$ $b_o = 0.3$

$f_t = \sigma(W_f\,[x_t,\ y_{t-1}] + b_f)$

$s_t = f_t \cdot s_{t-1} + i_t \cdot a_t$

$i_t = \sigma(W_i\,[x_t,\ y_{t-1}] + b_i)$

$o_t = \sigma(W_o\,[x_t,\ y_{t-1}] + b_o)$

$a_t = \tanh(W_a\,[x_t,\ y_{t-1}] + b_a)$

$y_t = o_t \cdot \tanh(s_t)$

Time	$x_{t,1}$	$x_{t,2}$	y
t=1	0.1	0.4	0.3
t=2	0.7	0.9	0.5

y_0	s_0
0	0

Step 1 (Input Forward Propagation)

$a_1 = \tanh(W_a\,[x_1,\ y_0] + b_a)$

$= \tanh(\begin{pmatrix}0.4\\0.2\\0.1\end{pmatrix}\begin{pmatrix}0.1\\0.4\\0\end{pmatrix} + 0.5)$

$= \tanh(0.4 \cdot 0.1 + 0.2 \cdot 0.4 + 0.1 \cdot 0 + 0.5)$

$= \tanh(0.62)$

$= 0.5511$

$f_1 = 0.6434$

$i_1 = 0.5842$

$a_1 = 0.5511$

$W_f = \begin{pmatrix}0.7\\0.3\\0.4\end{pmatrix}$ $b_f = 0.4$

$W_i = \begin{pmatrix}0.2\\0.3\\0.4\end{pmatrix}$ $b_i = 0.2$

$W_a = \begin{pmatrix}0.4\\0.2\\0.1\end{pmatrix}$ $b_a = 0.5$

$W_o = \begin{pmatrix}0.8\\0.9\\0.2\end{pmatrix}$ $b_o = 0.3$

$f_t = \sigma(W_f\,[x_t,\, y_{t-1}] + b_f)$

$s_t = f_t \cdot s_{t-1} + i_t \cdot a_t$

$i_t = \sigma(W_i\,[x_t,\, y_{t-1}] + b_i)$

$o_t = \sigma(W_o\,[x_t,\, y_{t-1}] + b_o)$

$a_t = \tanh(W_a\,[x_t,\, y_{t-1}] + b_a)$

$y_t = o_t \cdot \tanh(s_t)$

Time	$x_{t,1}$	$x_{t,2}$	y
t=1	0.1	0.4	0.3
t=2	0.7	0.9	0.5

y_0	s_0
0	0

Step 1 (Input Forward Propagation)

$s_1 = f_1 \cdot s_0 + i_1 \cdot a_1$

$= 0.6434 \cdot 0 + 0.5842 \cdot 0.5511$

$= 0.3220$

$f_1 = 0.6434$

$i_1 = 0.5842$

$a_1 = 0.5511$

$s_1 = 0.3220$

$W_f = \begin{pmatrix} 0.7 \\ 0.3 \\ 0.4 \end{pmatrix}$ $b_f = 0.4$

$W_i = \begin{pmatrix} 0.2 \\ 0.3 \\ 0.4 \end{pmatrix}$ $b_i = 0.2$

$W_a = \begin{pmatrix} 0.4 \\ 0.2 \\ 0.1 \end{pmatrix}$ $b_a = 0.5$

$W_o = \begin{pmatrix} 0.8 \\ 0.9 \\ 0.2 \end{pmatrix}$ $b_o = 0.3$

$f_t = \sigma(W_f\,[x_t,\ y_{t-1}] + b_f)$

$s_t = f_t \cdot s_{t-1} + i_t \cdot a_t$

$i_t = \sigma(W_i\,[x_t,\ y_{t-1}] + b_i)$

$o_t = \sigma(W_o\,[x_t,\ y_{t-1}] + b_o)$

$a_t = \tanh(W_a\,[x_t,\ y_{t-1}] + b_a)$

$y_t = o_t \cdot \tanh(s_t)$

Time	$x_{t,1}$	$x_{t,2}$	y
t=1	0.1	0.4	0.3
t=2	0.7	0.9	0.5

y_0	s_0
0	0

Step 1 (Input Forward Propagation)

$o_1 = \sigma(W_o\,[x_1,\ y_0] + b_o)$

$= \sigma(\begin{pmatrix}0.8\\0.9\\0.2\end{pmatrix}\begin{pmatrix}0.1\\0.4\\0\end{pmatrix} + 0.3)$

$= \sigma(0.8 \cdot 0.1 + 0.9 \cdot 0.4 + 0.2 \cdot 0 + 0.3)$

$= \sigma(0.74)$

$= 0.6770$

$W_f = \begin{pmatrix}0.7\\0.3\\0.4\end{pmatrix}$ $b_f = 0.4$

$W_i = \begin{pmatrix}0.2\\0.3\\0.4\end{pmatrix}$ $b_i = 0.2$

$W_a = \begin{pmatrix}0.4\\0.2\\0.1\end{pmatrix}$ $b_a = 0.5$

$W_o = \begin{pmatrix}0.8\\0.9\\0.2\end{pmatrix}$ $b_o = 0.3$

$f_1 = 0.6434$

$i_1 = 0.5842$

$a_1 = 0.5511$

$s_1 = 0.3220$

$o_1 = 0.6770$

$f_t = \sigma(W_f\,[x_t,\, y_{t-1}] + b_f)$

$s_t = f_t \cdot s_{t-1} + i_t \cdot a_t$

$i_t = \sigma(W_i\,[x_t,\, y_{t-1}] + b_i)$

$o_t = \sigma(W_o\,[x_t,\, y_{t-1}] + b_o)$

$a_t = \tanh(W_a\,[x_t,\, y_{t-1}] + b_a)$

$y_t = o_t \cdot \tanh(s_t)$

Time	$x_{t,1}$	$x_{t,2}$	y
t=1	0.1	0.4	0.3
t=2	0.7	0.9	0.5

y_0	s_0
0	0

Step 1 (Input Forward Propagation)

$y_1 = o_1 \cdot \tanh(s_1)$

$= 0.6770 \cdot \tanh(0.3220)$

$= 0.2107$

$W_f = \begin{pmatrix} 0.7 \\ 0.3 \\ 0.4 \end{pmatrix}$ $b_f = 0.4$

$W_i = \begin{pmatrix} 0.2 \\ 0.3 \\ 0.4 \end{pmatrix}$ $b_i = 0.2$

$W_a = \begin{pmatrix} 0.4 \\ 0.2 \\ 0.1 \end{pmatrix}$ $b_a = 0.5$

$W_o = \begin{pmatrix} 0.8 \\ 0.9 \\ 0.2 \end{pmatrix}$ $b_o = 0.3$

$f_1 = 0.6434$

$i_1 = 0.5842$

$a_1 = 0.5511$

$s_1 = 0.3220$

$o_1 = 0.6770$

$y_1 = 0.2107$

$f_t = \sigma(W_f [x_t, y_{t-1}] + b_f)$

$s_t = f_t \cdot s_{t-1} + i_t \cdot a_t$

$i_t = \sigma(W_i [x_t, y_{t-1}] + b_i)$

$o_t = \sigma(W_o [x_t, y_{t-1}] + b_o)$

$a_t = \tanh(W_a [x_t, y_{t-1}] + b_a)$

$y_t = o_t \cdot \tanh(s_t)$

Time	$x_{t,1}$	$x_{t,2}$	y
t=1	0.1	0.4	0.3
t=2	0.7	0.9	0.5

y_0	s_0
0	0

Step 1 (Input Forward Propagation)

Error = y_1 - y

= 0.2107 – 0.3

= -0.0893

$$W_f = \begin{pmatrix} 0.7 \\ 0.3 \\ 0.4 \end{pmatrix} \quad b_f = 0.4$$

$$W_i = \begin{pmatrix} 0.2 \\ 0.3 \\ 0.4 \end{pmatrix} \quad b_i = 0.2$$

$$W_a = \begin{pmatrix} 0.4 \\ 0.2 \\ 0.1 \end{pmatrix} \quad b_a = 0.5$$

$$W_o = \begin{pmatrix} 0.8 \\ 0.9 \\ 0.2 \end{pmatrix} \quad b_o = 0.3$$

$f_1 = 0.6434$

$i_1 = 0.5842$

$a_1 = 0.5511$

$s_1 = 0.3220$

$o_1 = 0.6770$

$y_1 = 0.2107$

$f_t = \sigma(W_f\,[x_t,\ y_{t-1}] + b_f)$

$s_t = f_t \cdot s_{t-1} + i_t \cdot a_t$

$i_t = \sigma(W_i\,[x_t,\ y_{t-1}] + b_i)$

$o_t = \sigma(W_o\,[x_t,\ y_{t-1}] + b_o)$

$a_t = \tanh(W_a\,[x_t,\ y_{t-1}] + b_a)$

$y_t = o_t \cdot \tanh(s_t)$

Time	$x_{t,1}$	$x_{t,2}$	y
t=1	0.1	0.4	0.3
t=2	0.7	0.9	0.5

Step 1 (Input Forward Propagation)

y_1	s_1
0.2107	0.3220

$$W_f = \begin{pmatrix}0.7\\0.3\\0.4\end{pmatrix} \quad b_f = 0.4$$

$$W_i = \begin{pmatrix}0.2\\0.3\\0.4\end{pmatrix} \quad b_i = 0.2$$

$$W_a = \begin{pmatrix}0.4\\0.2\\0.1\end{pmatrix} \quad b_a = 0.5$$

$$W_o = \begin{pmatrix}0.8\\0.9\\0.2\end{pmatrix} \quad b_o = 0.3$$

$f_1 = 0.6434$

$i_1 = 0.5842$

$a_1 = 0.5511$

$s_1 = 0.3220$

$o_1 = 0.6770$

$y_1 = 0.2107$

$f_t = \sigma(W_f [x_t, y_{t-1}] + b_f)$

$s_t = f_t \cdot s_{t-1} + i_t \cdot a_t$

$i_t = \sigma(W_i [x_t, y_{t-1}] + b_i)$

$o_t = \sigma(W_o [x_t, y_{t-1}] + b_o)$

$a_t = \tanh(W_a [x_t, y_{t-1}] + b_a)$

$y_t = o_t \cdot \tanh(s_t)$

Time	$x_{t,1}$	$x_{t,2}$	y
t=1	0.1	0.4	0.3
t=2	0.7	0.9	0.5

y_1	s_1
0.2107	0.3220

Step 1 (Input Forward Propagation)

$f_2 = \sigma(W_f [x_2, y_1] + b_f)$

$$= \sigma(\begin{pmatrix}0.7\\0.3\\0.4\end{pmatrix}\begin{pmatrix}0.7\\0.9\\0.2107\end{pmatrix} + 0.4)$$

$$= \sigma(0.7 \cdot 0.7 + 0.3 \cdot 0.9 + 0.4 \cdot 0.2107 + 0.4)$$

$$= \sigma(1.2443)$$

$$= 0.7763$$

$f_2 = 0.7763$

$W_f = \begin{pmatrix}0.7\\0.3\\0.4\end{pmatrix}$ $b_f = 0.4$

$W_i = \begin{pmatrix}0.2\\0.3\\0.4\end{pmatrix}$ $b_i = 0.2$

$W_a = \begin{pmatrix}0.4\\0.2\\0.1\end{pmatrix}$ $b_a = 0.5$

$W_o = \begin{pmatrix}0.8\\0.9\\0.2\end{pmatrix}$ $b_o = 0.3$

$f_t = \sigma(W_f\,[x_t,\ y_{t-1}] + b_f)$

$s_t = f_t \cdot s_{t-1} + i_t \cdot a_t$

$i_t = \sigma(W_i\,[x_t,\ y_{t-1}] + b_i)$

$o_t = \sigma(W_o\,[x_t,\ y_{t-1}] + b_o)$

$a_t = \tanh(W_a\,[x_t,\ y_{t-1}] + b_a)$

$y_t = o_t \cdot \tanh(s_t)$

Time	$x_{t,\,1}$	$x_{t,\,2}$	y
t=1	0.1	0.4	0.3
t=2	0.7	0.9	0.5

y_1	s_1
0.2107	0.3220

$$W_f = \begin{pmatrix}0.7\\0.3\\0.4\end{pmatrix} \quad b_f = 0.4$$

$$W_i = \begin{pmatrix}0.2\\0.3\\0.4\end{pmatrix} \quad b_i = 0.2$$

$$W_a = \begin{pmatrix}0.4\\0.2\\0.1\end{pmatrix} \quad b_a = 0.5$$

$$W_o = \begin{pmatrix}0.8\\0.9\\0.2\end{pmatrix} \quad b_o = 0.3$$

Step 1 (Input Forward Propagation)

$i_2 = \sigma(W_i\,[x_2,\ y_1] + b_i)$

$$= \sigma(\begin{pmatrix}0.2\\0.3\\0.4\end{pmatrix}\begin{pmatrix}0.7\\0.9\\0.2107\end{pmatrix} + 0.2)$$

$$= \sigma(0.2 \cdot 0.7 + 0.3 \cdot 0.9 + 0.4 \cdot 0.2107 + 0.2)$$

$$= \sigma(0.6943)$$

$$= 0.6669$$

$f_2 = 0.7763$

$i_2 = 0.6669$

$f_t = \sigma(W_f\,[x_t, y_{t-1}] + b_f)$

$s_t = f_t \cdot s_{t-1} + i_t \cdot a_t$

$i_t = \sigma(W_i\,[x_t, y_{t-1}] + b_i)$

$o_t = \sigma(W_o\,[x_t, y_{t-1}] + b_o)$

$a_t = \tanh(W_a\,[x_t, y_{t-1}] + b_a)$

$y_t = o_t \cdot \tanh(s_t)$

Time	$x_{t,1}$	$x_{t,2}$	y
t=1	0.1	0.4	0.3
t=2	0.7	0.9	0.5

y_1	s_1
0.2107	0.3220

Step 1 (Input Forward Propagation)

$a_2 = \tanh(W_a\,[x_2, y_1] + b_a)$

$= \tanh(\begin{pmatrix}0.4\\0.2\\0.1\end{pmatrix}\begin{pmatrix}0.7\\0.9\\0.2107\end{pmatrix} + 0.5)$

$= \tanh(0.4 \cdot 0.7 + 0.2 \cdot 0.9 + 0.1 \cdot 0.2107 + 0.5)$

$= \tanh(0.9811)$

$= 0.7535$

$f_2 = 0.7763$

$i_2 = 0.6669$

$a_2 = 0.7535$

$W_f = \begin{pmatrix}0.7\\0.3\\0.4\end{pmatrix}$ $b_f = 0.4$

$W_i = \begin{pmatrix}0.2\\0.3\\0.4\end{pmatrix}$ $b_i = 0.2$

$W_a = \begin{pmatrix}0.4\\0.2\\0.1\end{pmatrix}$ $b_a = 0.5$

$W_o = \begin{pmatrix}0.8\\0.9\\0.2\end{pmatrix}$ $b_o = 0.3$

$f_t = \sigma(W_f [x_t, y_{t-1}] + b_f)$

$s_t = f_t \cdot s_{t-1} + i_t \cdot a_t$

$i_t = \sigma(W_i [x_t, y_{t-1}] + b_i)$

$o_t = \sigma(W_o [x_t, y_{t-1}] + b_o)$

$a_t = \tanh(W_a [x_t, y_{t-1}] + b_a)$

$y_t = o_t \cdot \tanh(s_t)$

Time	$x_{t,1}$	$x_{t,2}$	y
t=1	0.1	0.4	0.3
t=2	0.7	0.9	0.5

y_1	s_1
0.2107	0.3220

Step 1 (Input Forward Propagation)

$s_2 = f_2 \cdot s_1 + i_2 \cdot a_2$

$= 0.7763 \cdot 0.3220 + 0.6669 \cdot 0.7535$

$= 0.7525$

$f_2 = 0.7763$

$i_2 = 0.6669$

$a_2 = 0.7535$

$s_2 = 0.7525$

$W_f = \begin{pmatrix} 0.7 \\ 0.3 \\ 0.4 \end{pmatrix}$ $b_f = 0.4$

$W_i = \begin{pmatrix} 0.2 \\ 0.3 \\ 0.4 \end{pmatrix}$ $b_i = 0.2$

$W_a = \begin{pmatrix} 0.4 \\ 0.2 \\ 0.1 \end{pmatrix}$ $b_a = 0.5$

$W_o = \begin{pmatrix} 0.8 \\ 0.9 \\ 0.2 \end{pmatrix}$ $b_o = 0.3$

$f_t = \sigma(W_f\,[x_t, y_{t-1}] + b_f)$

$s_t = f_t \cdot s_{t-1} + i_t \cdot a_t$

$i_t = \sigma(W_i\,[x_t, y_{t-1}] + b_i)$

$o_t = \sigma(W_o\,[x_t, y_{t-1}] + b_o)$

$a_t = \tanh(W_a\,[x_t, y_{t-1}] + b_a)$

$y_t = o_t \cdot \tanh(s_t)$

Time	$x_{t,1}$	$x_{t,2}$	y
t=1	0.1	0.4	0.3
t=2	0.7	0.9	0.5

y_1	s_1
0.2107	0.3220

Step 1 (Input Forward Propagation)

$o_2 = \sigma(W_o\,[x_2, y_1] + b_o)$

$$= \sigma(\begin{pmatrix}0.8\\0.9\\0.2\end{pmatrix}\begin{pmatrix}0.7\\0.9\\0.2107\end{pmatrix} + 0.3)$$

$$= \sigma(0.8 \cdot 0.7 + 0.9 \cdot 0.9 + 0.2 \cdot 0.2107 + 0.3)$$

$$= \sigma(1.7121)$$

$$= 0.8471$$

$W_f = \begin{pmatrix}0.7\\0.3\\0.4\end{pmatrix}$ $b_f = 0.4$

$W_i = \begin{pmatrix}0.2\\0.3\\0.4\end{pmatrix}$ $b_i = 0.2$

$W_a = \begin{pmatrix}0.4\\0.2\\0.1\end{pmatrix}$ $b_a = 0.5$

$W_o = \begin{pmatrix}0.8\\0.9\\0.2\end{pmatrix}$ $b_o = 0.3$

$f_2 = 0.7763$

$i_2 = 0.6669$

$a_2 = 0.7535$

$s_2 = 0.7525$

$o_2 = 0.8471$

$f_t = \sigma(W_f\,[x_t, y_{t-1}] + b_f)$

$s_t = f_t \cdot s_{t-1} + i_t \cdot a_t$

$i_t = \sigma(W_i\,[x_t, y_{t-1}] + b_i)$

$o_t = \sigma(W_o\,[x_t, y_{t-1}] + b_o)$

$a_t = \tanh(W_a\,[x_t, y_{t-1}] + b_a)$

$y_t = o_t \cdot \tanh(s_t)$

Time	$x_{t,1}$	$x_{t,2}$	y
t=1	0.1	0.4	0.3
t=2	0.7	0.9	0.5

y_1	s_1
0.2107	0.3220

Step 1 (Input Forward Propagation)

$y_2 = o_2 \cdot \tanh(s_2)$

$= 0.8471 \cdot \tanh(0.7525)$

$= 0.5393$

$$W_f = \begin{pmatrix}0.7\\0.3\\0.4\end{pmatrix} \quad b_f = 0.4$$

$$W_i = \begin{pmatrix}0.2\\0.3\\0.4\end{pmatrix} \quad b_i = 0.2$$

$$W_a = \begin{pmatrix}0.4\\0.2\\0.1\end{pmatrix} \quad b_a = 0.5$$

$$W_o = \begin{pmatrix}0.8\\0.9\\0.2\end{pmatrix} \quad b_o = 0.3$$

$f_2 = 0.7763$

$i_2 = 0.6669$

$a_2 = 0.7535$

$s_2 = 0.7525$

$o_2 = 0.8471$

$y_2 = 0.5393$

$f_t = \sigma(W_f [x_t, y_{t-1}] + b_f)$

$s_t = f_t \cdot s_{t-1} + i_t \cdot a_t$

$i_t = \sigma(W_i [x_t, y_{t-1}] + b_i)$

$o_t = \sigma(W_o [x_t, y_{t-1}] + b_o)$

$a_t = \tanh(W_a [x_t, y_{t-1}] + b_a)$

$y_t = o_t \cdot \tanh(s_t)$

Time	$x_{t,1}$	$x_{t,2}$	y
t=1	0.1	0.4	0.3
t=2	0.7	0.9	0.5

y_1	s_1
0.2107	0.3220

Step 1 (Input Forward Propagation)

$\text{Error} = y_2 - y$

$= 0.5393 - 0.5$

$= 0.0393$

$W_f = \begin{pmatrix} 0.7 \\ 0.3 \\ 0.4 \end{pmatrix}$ $b_f = 0.4$

$W_i = \begin{pmatrix} 0.2 \\ 0.3 \\ 0.4 \end{pmatrix}$ $b_i = 0.2$

$W_a = \begin{pmatrix} 0.4 \\ 0.2 \\ 0.1 \end{pmatrix}$ $b_a = 0.5$

$W_o = \begin{pmatrix} 0.8 \\ 0.9 \\ 0.2 \end{pmatrix}$ $b_o = 0.3$

$f_2 = 0.7763$

$i_2 = 0.6669$

$a_2 = 0.7535$

$s_2 = 0.7525$

$o_2 = 0.8471$

$y_2 = 0.5393$

LSTM

- Similar to the “neural network”, the LSTM model (and the basic RNN model) could also have multiple layers and have multiple memory units in each layer.