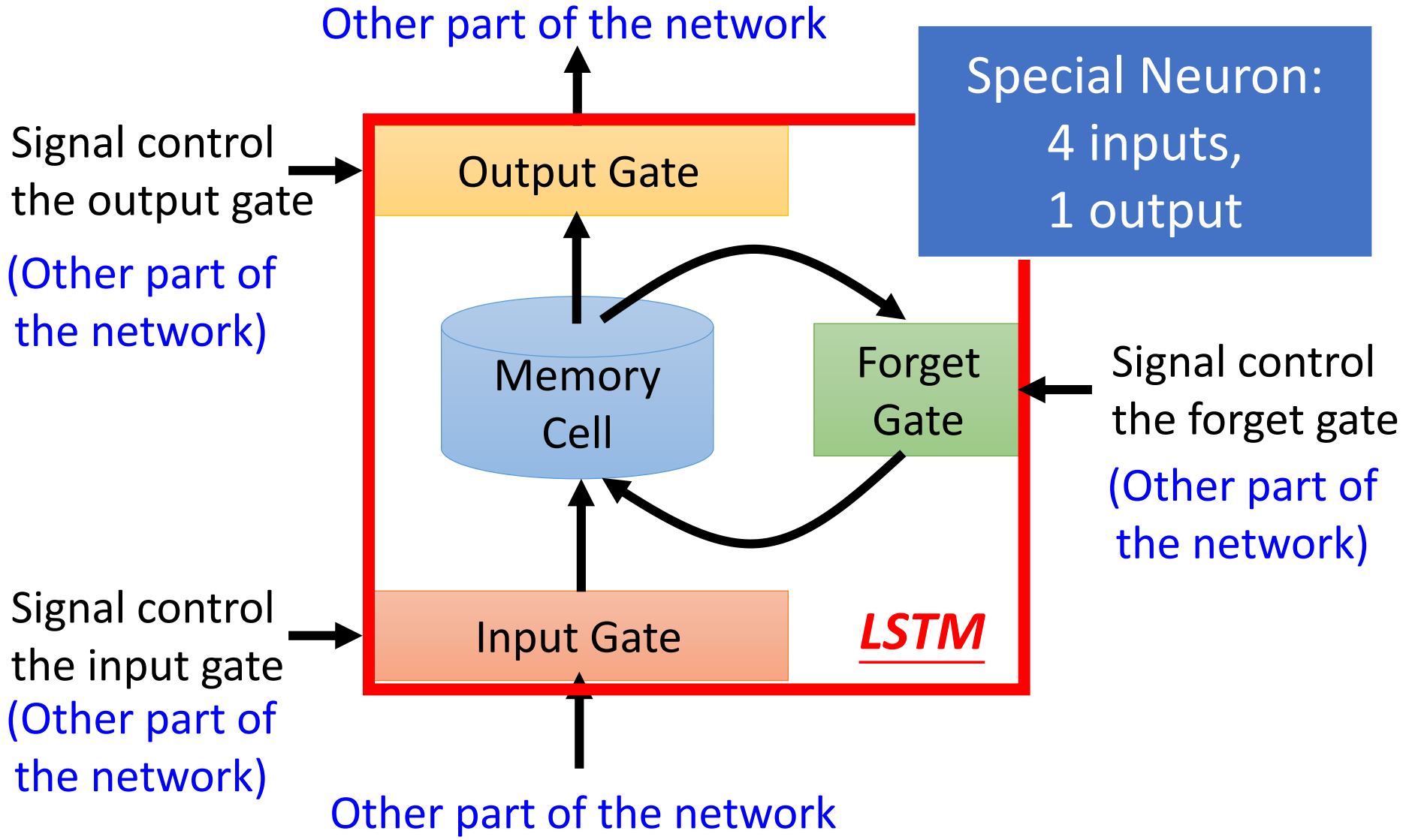
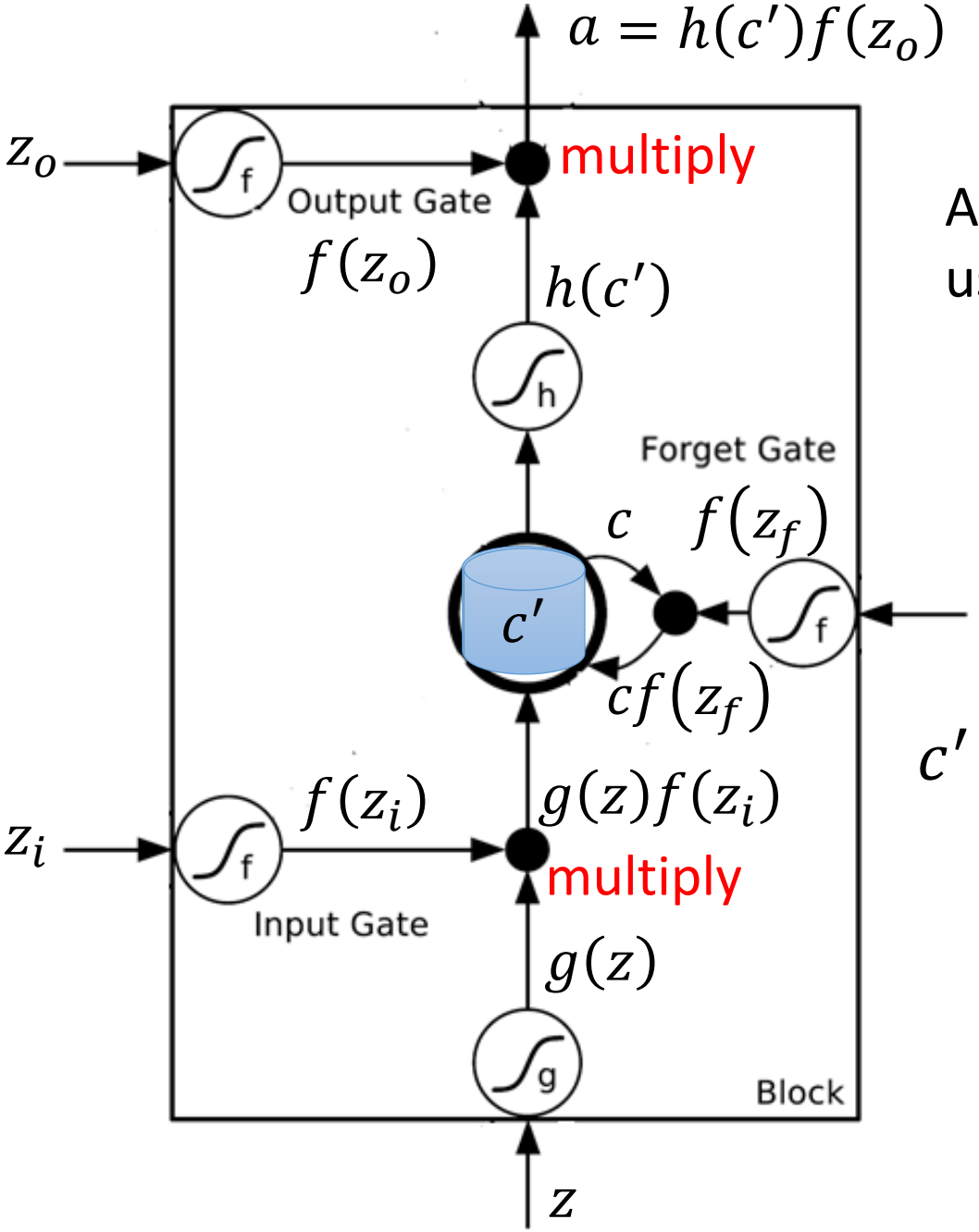
Long Short-term Memory (LSTM)





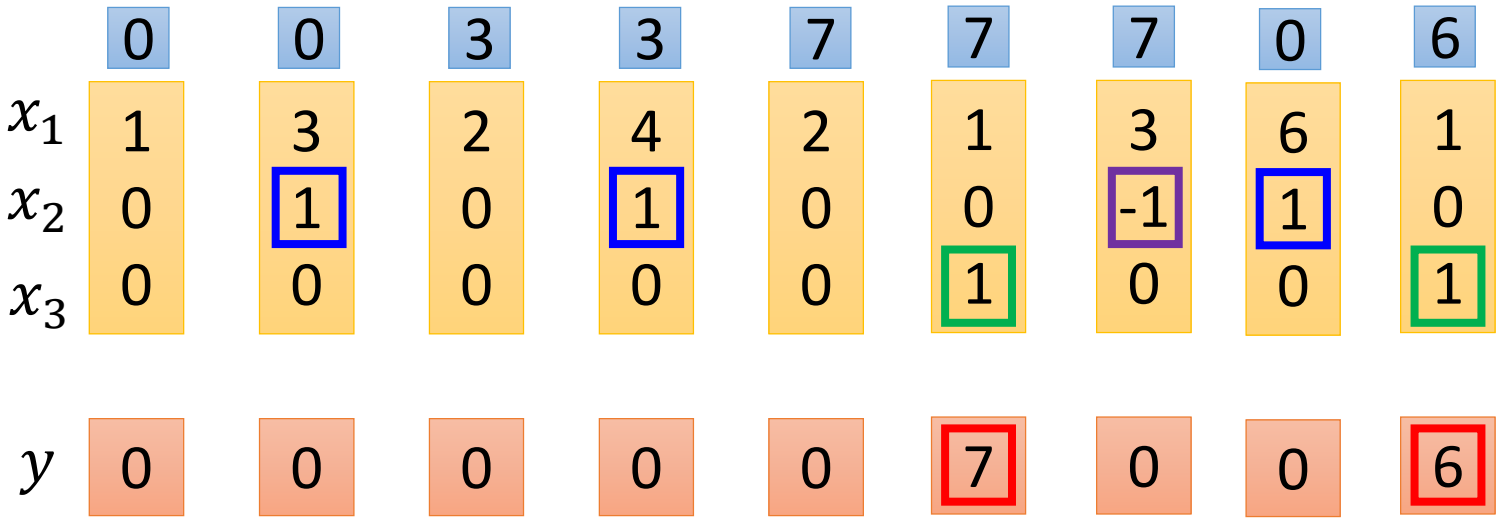
Activation function f is usually a sigmoid function

Between 0 and 1

Mimic open and close gate

$$c' = g(z)f(z_i) + cf(z_f)$$

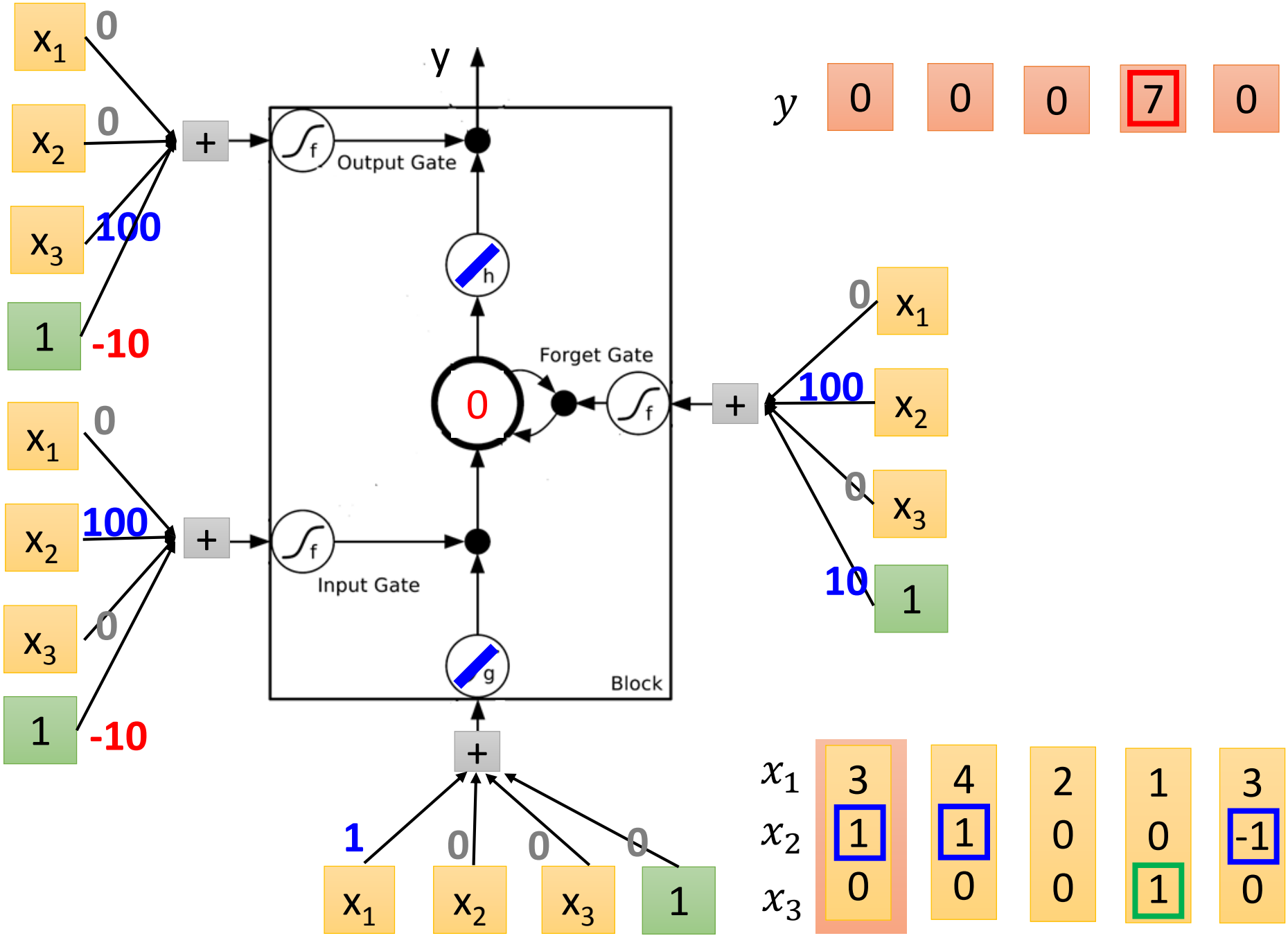
LSTM - Example

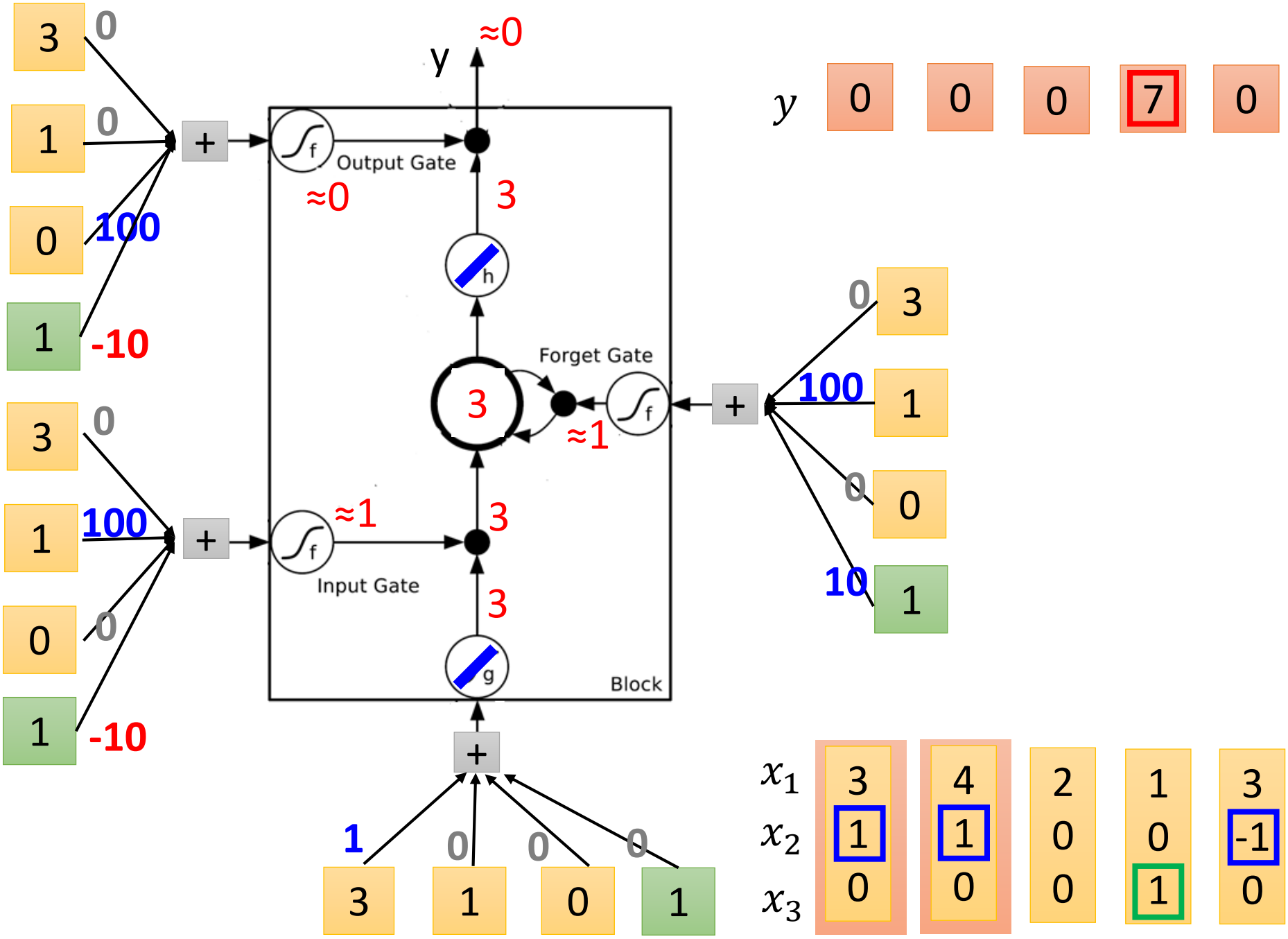


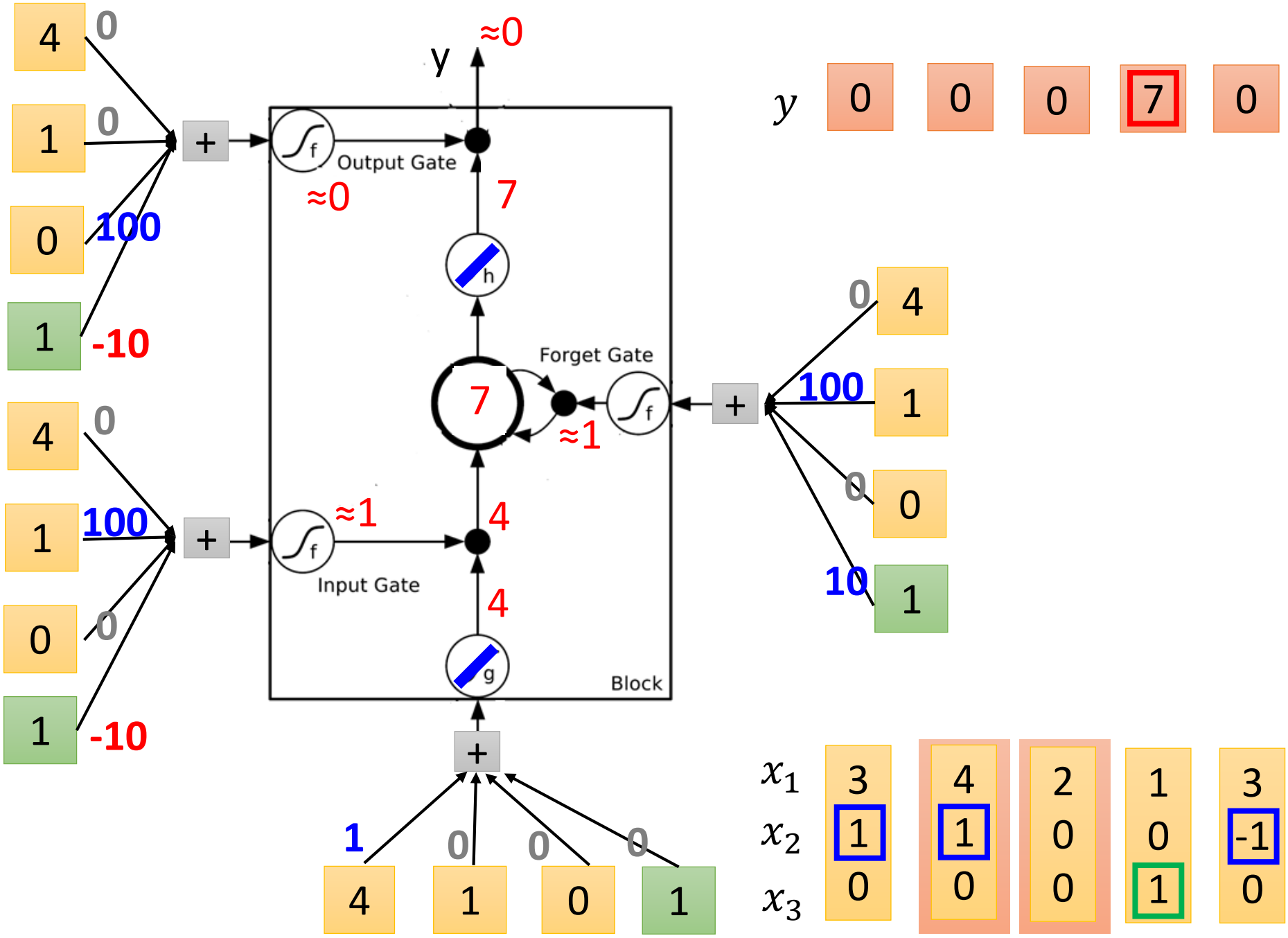
When x_2 = 1, add the numbers of x_1 into the memory

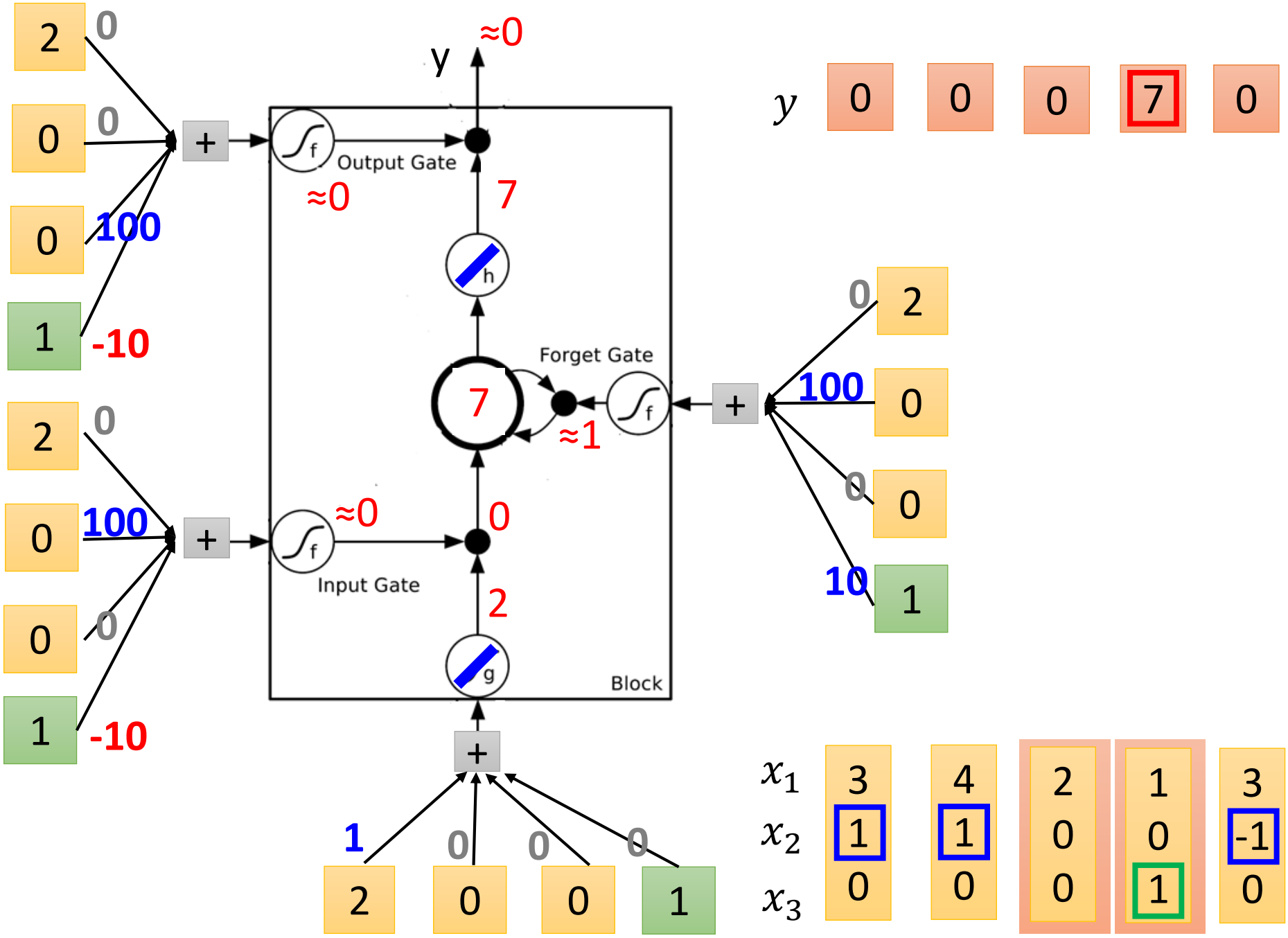
When x_2 = -1, reset the memory

When x_3 = 1, output the number in the memory.


x1
x2
x3
1
0
0
100
-10
+
f
Output Gate
y
h
Forget Gate
0
f
+
0
100
0
10
x1
x2
x3
1
x1
x2
x3
1
0
100
0
-10
+
f
Input Gate
g
Block
+
1
0
0
0
x1
x2
x3
1
y
0
0
0
7
0
x1
x2
x3
3
1
0
4
1
0
2
0
0
1
0
1
3
-1
0

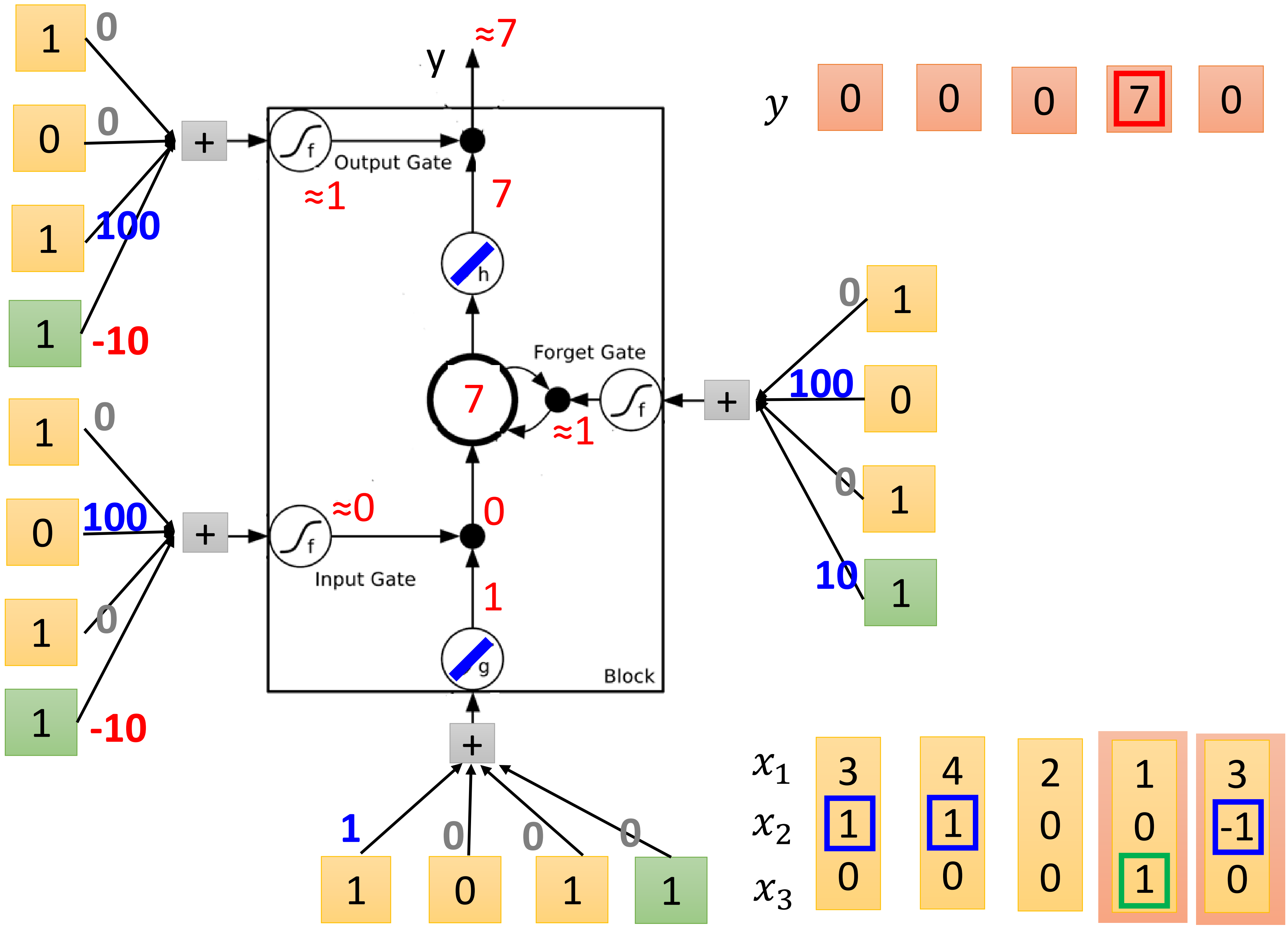

y
≈0
Output Gate
≈0
3
Forget Gate
3
≈1
Input Gate
≈1
3
3
Block
+
+
+
+
f
f
f
h
g
3
0
1
100
0
1
-10
3
0
1
100
0
1
-10
0
3
100
1
0
0
10
1
1
0
0
0
3
1
0
1
y
0
0
0
7
0
x_1
x_2
x_3
3
1
0
4
1
0
2
0
0
1
0
1
3
-1
0


y
≈0
Output Gate
≈0
7
Forget Gate
7
≈1
Input Gate
≈1
4
4
Block
Block
+
+
+
+
+
f
f
f
h
g
4
0
1
0
100
0
1
-10
4
0
1
100
0
0
1
-10
0
4
1
100
0
0
10
1
1
0
0
0
4
1
0
1
y
0
0
0
7
0
x_1
x_2
x_3
3
1
0
4
1
0
2
0
0
1
0
1
3
-1
0

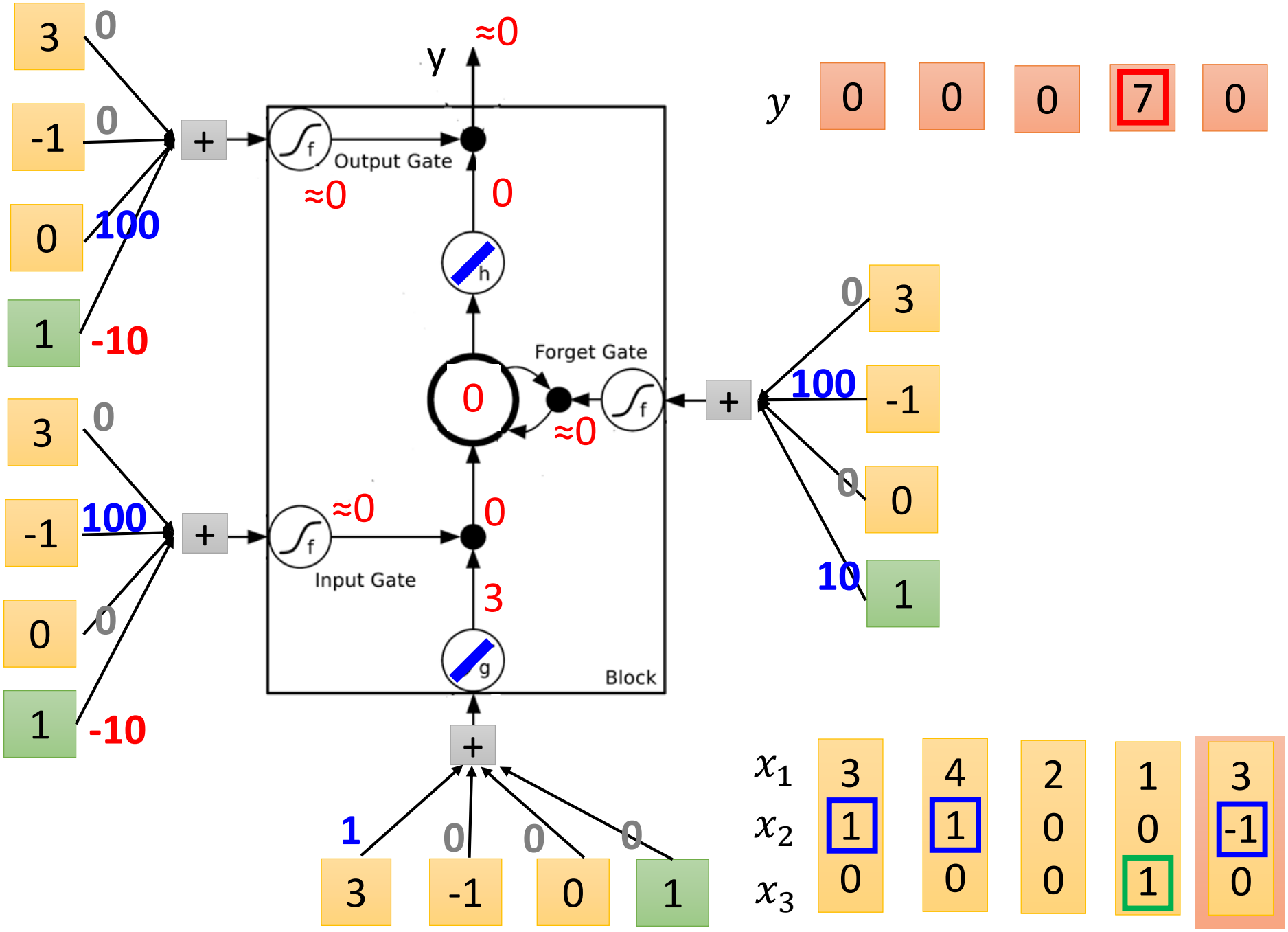


y 0 0 0 7 0

x_1	3	4	2	1	3
x_2	1	1	0	0	-1
x_3	0	0	0	1	0

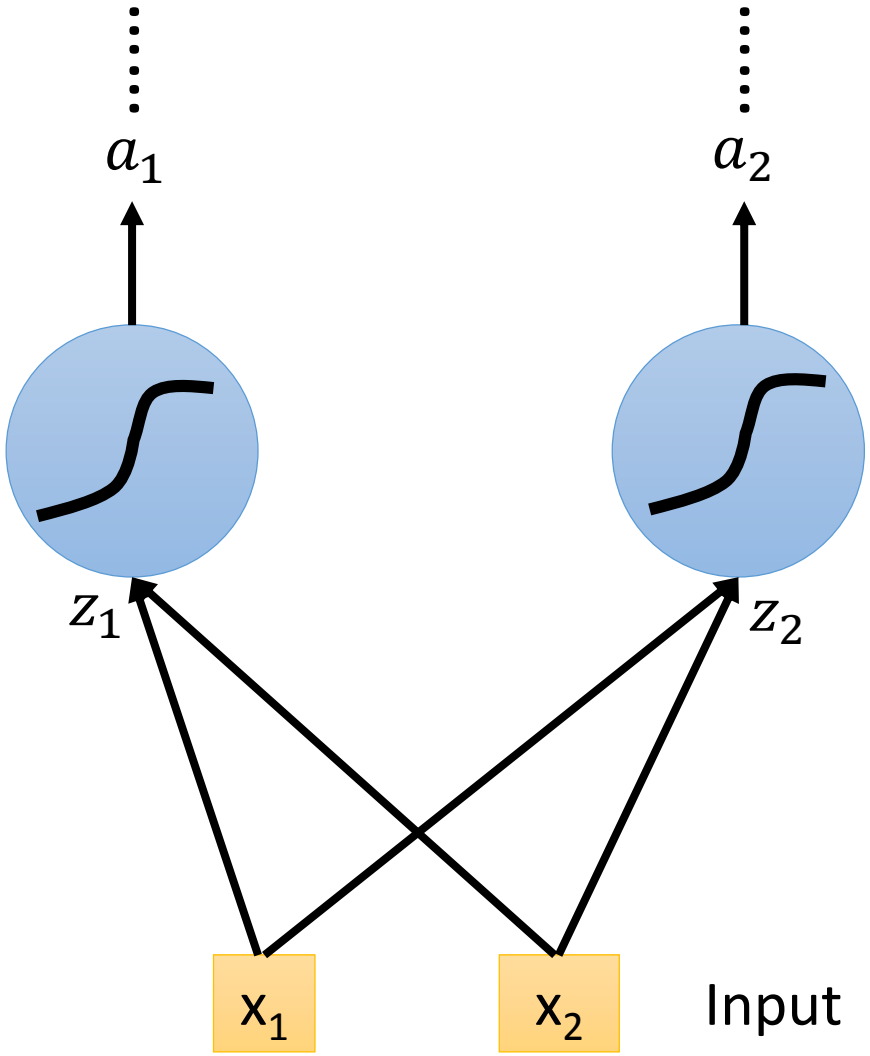


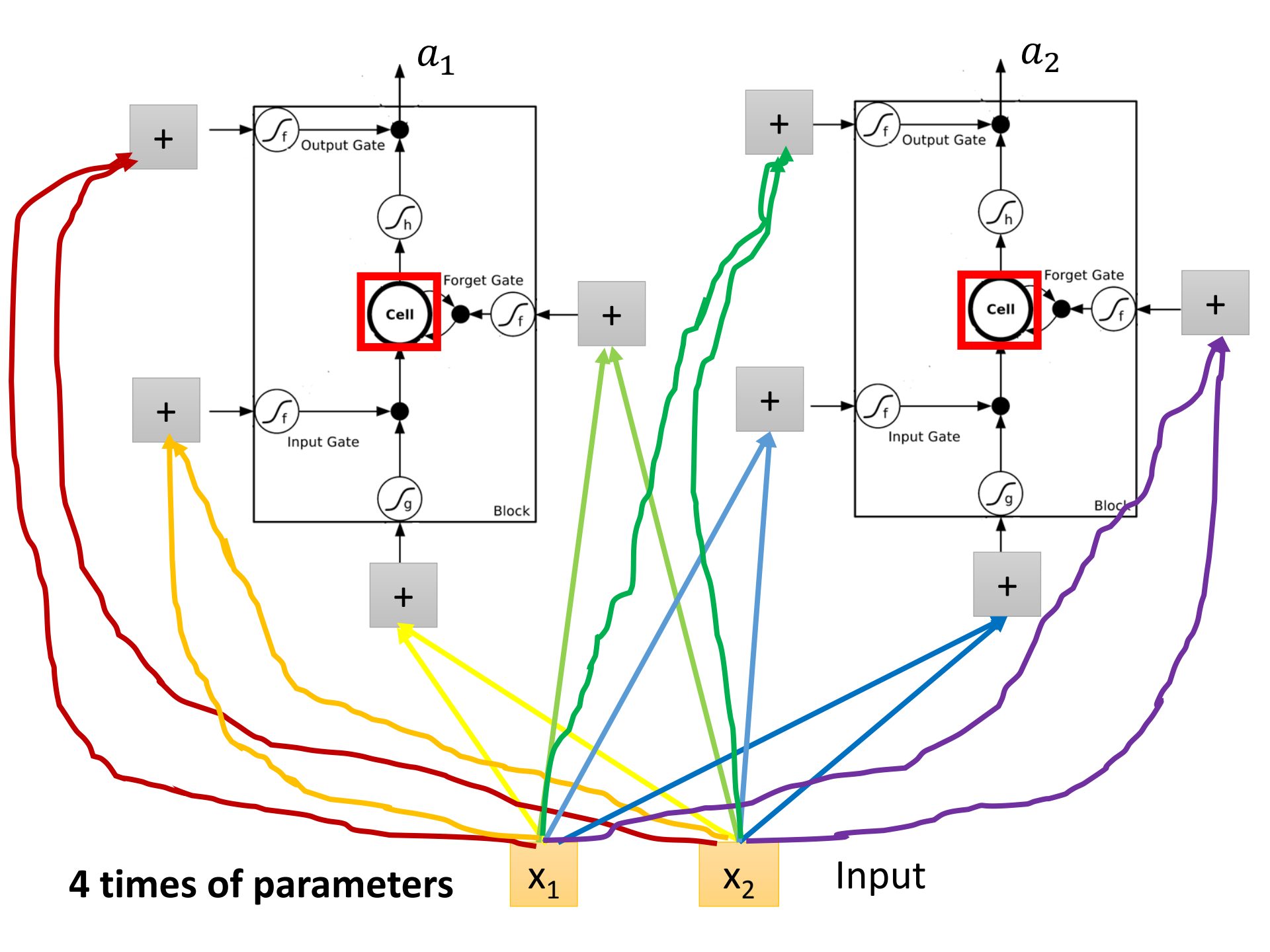
y
≈7
Output Gate
≈1
7
h
Forget Gate
≈1
f
≈0
0
1
Input Gate
g
Block
100
-10
10
y 0 0 0 7 0
x_1 3 4 2 1 3
x_2 1 1 0 0 -1
x_3 0 0 0 1 0


y
≈0
Output Gate
≈0
0
h
Forget Gate
0
≈0
Input Gate
≈0
0
3
Block
g
+
f
3
-1
0
1
0
100
-10
0
3
-1
0
1
100
-10
0
3
100
-1
0
10
1
1
0
0
0
3
-1
0
1
y
0
0
0
7
0
x_1
x_2
x_3
3
1
0
4
1
0
2
0
0
1
0
1
3
-1
0

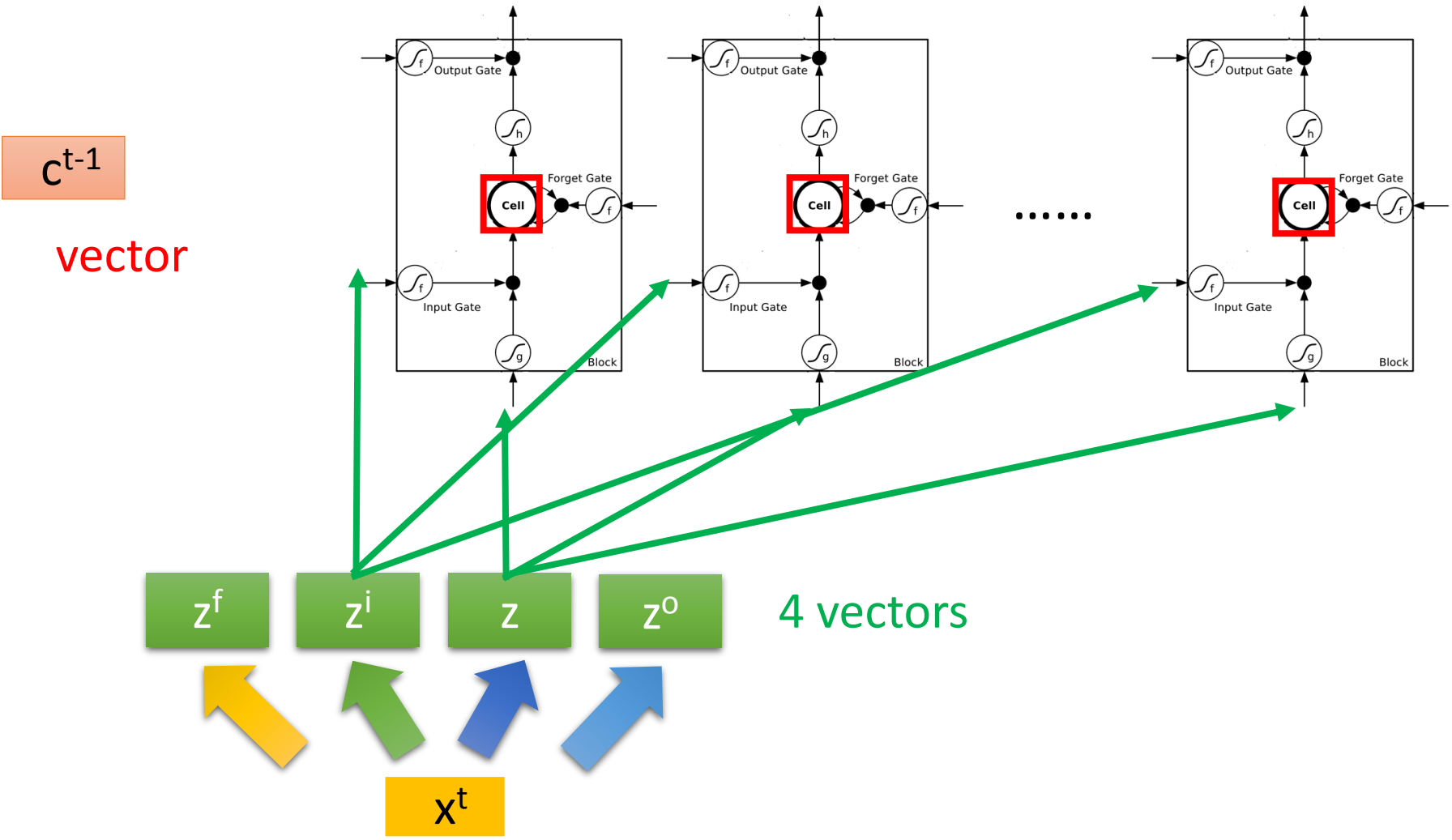
Original Network:

- Simply replace the neurons with LSTM

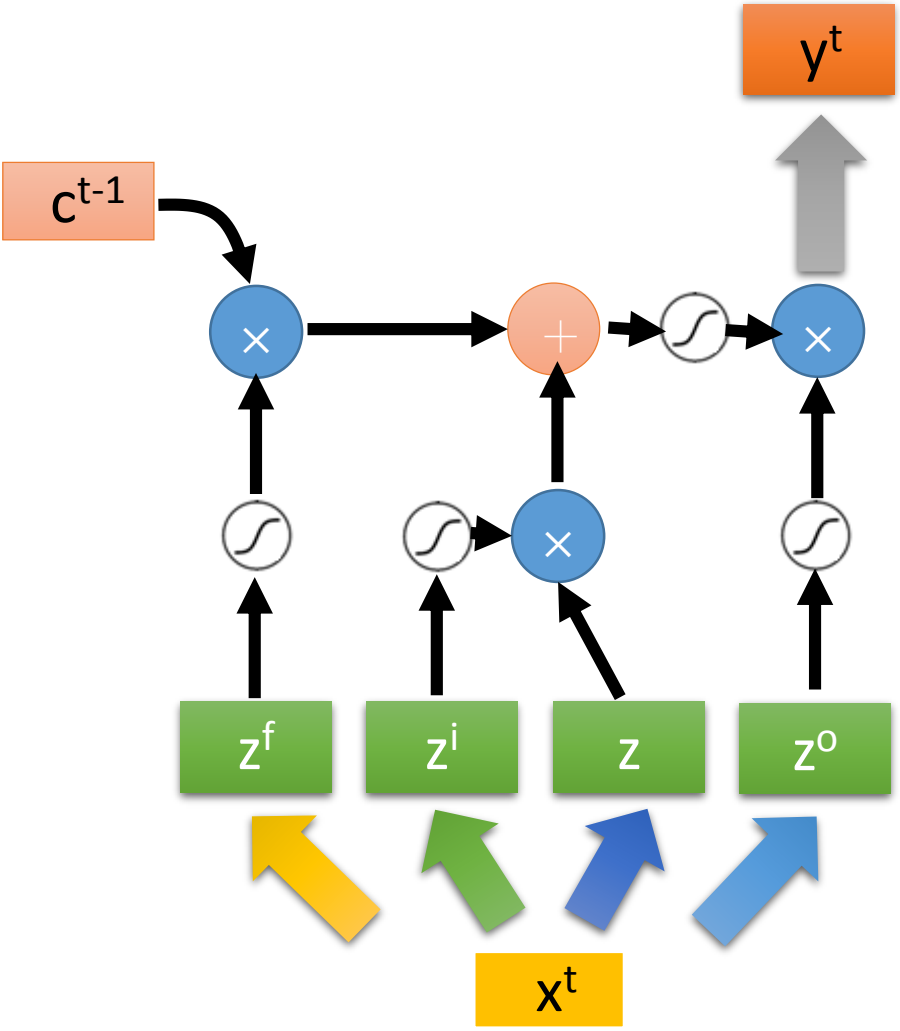


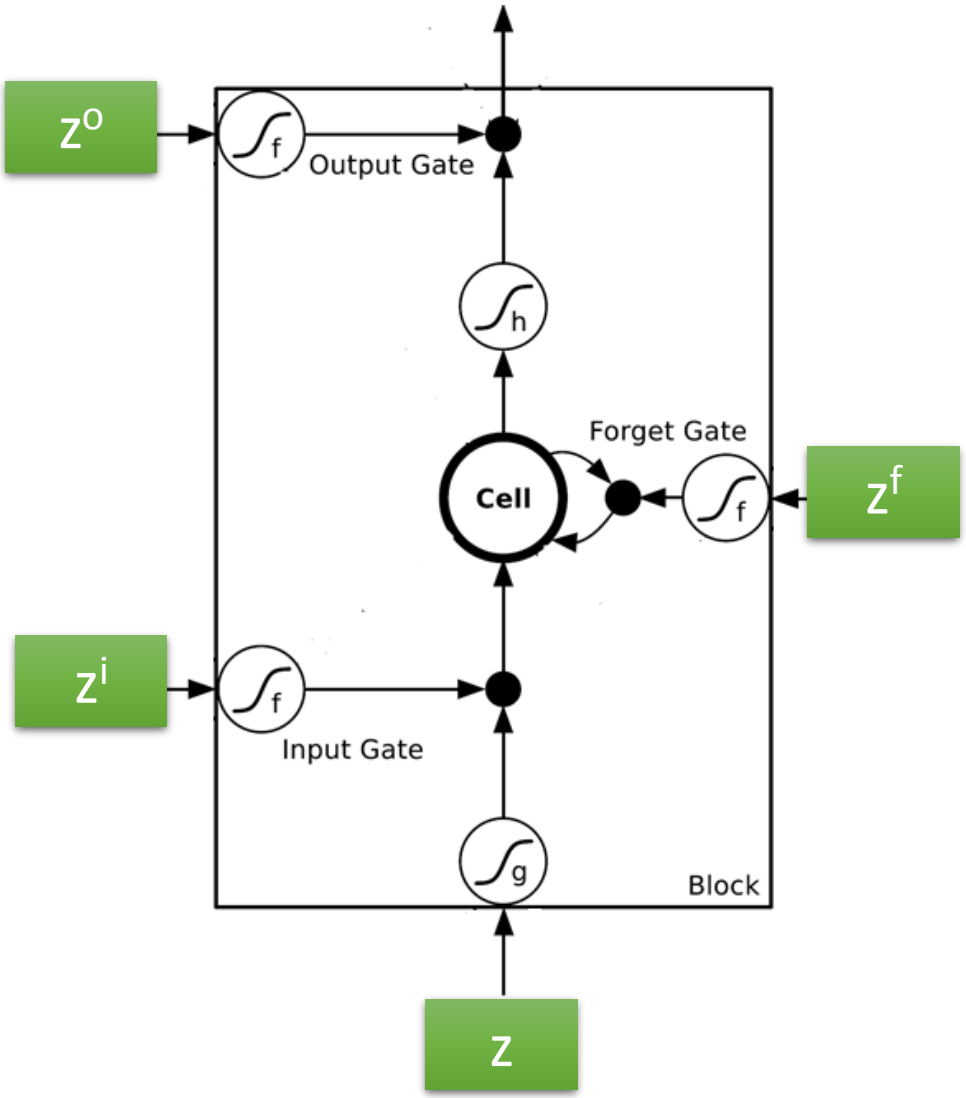

a_1
a_2
Output Gate
Forget Gate
Input Gate
Cell
Block
f
h
g
+
4 times of parameters
x_1
x_2
Input

LSTM


c^{t-1}
vector
Output Gate
Forget Gate
Cell
Input Gate
Block
.....
z^f
z^i
z
z^o
4 vectors
x^t

LSTM


y^t
c^{t-1}
z^f
z^i
z
z^o
x^t


z^o
Output Gate
h
Forget Gate
Cell
z^f
z^i
Input Gate
g
Block
z

LSTM

Extension: “peephole”