Natural Language Processing

第五周 循环神经网络

庞彦

yanpang@gzhu.edu.cn

Spring 2023

Overview

CONTENTS

01 循环神经网络

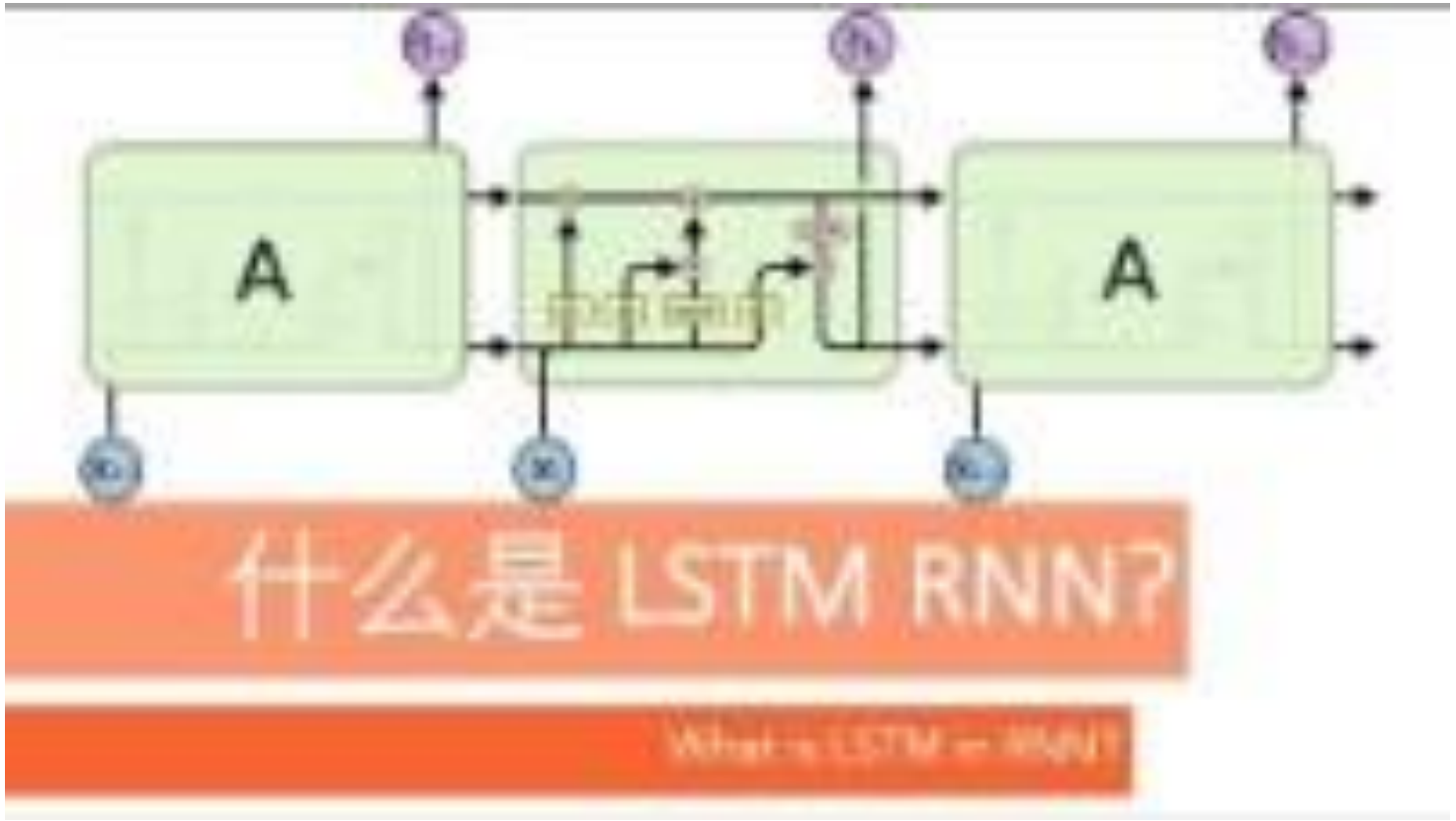
02 长短期记忆网络

01 Recurrent Neural Network

循环神经网络

Spring 2023

Introduction Video



Example Application

填空

I would like to arrive Beijing tomorrow.

ticket booking system

Slot
空

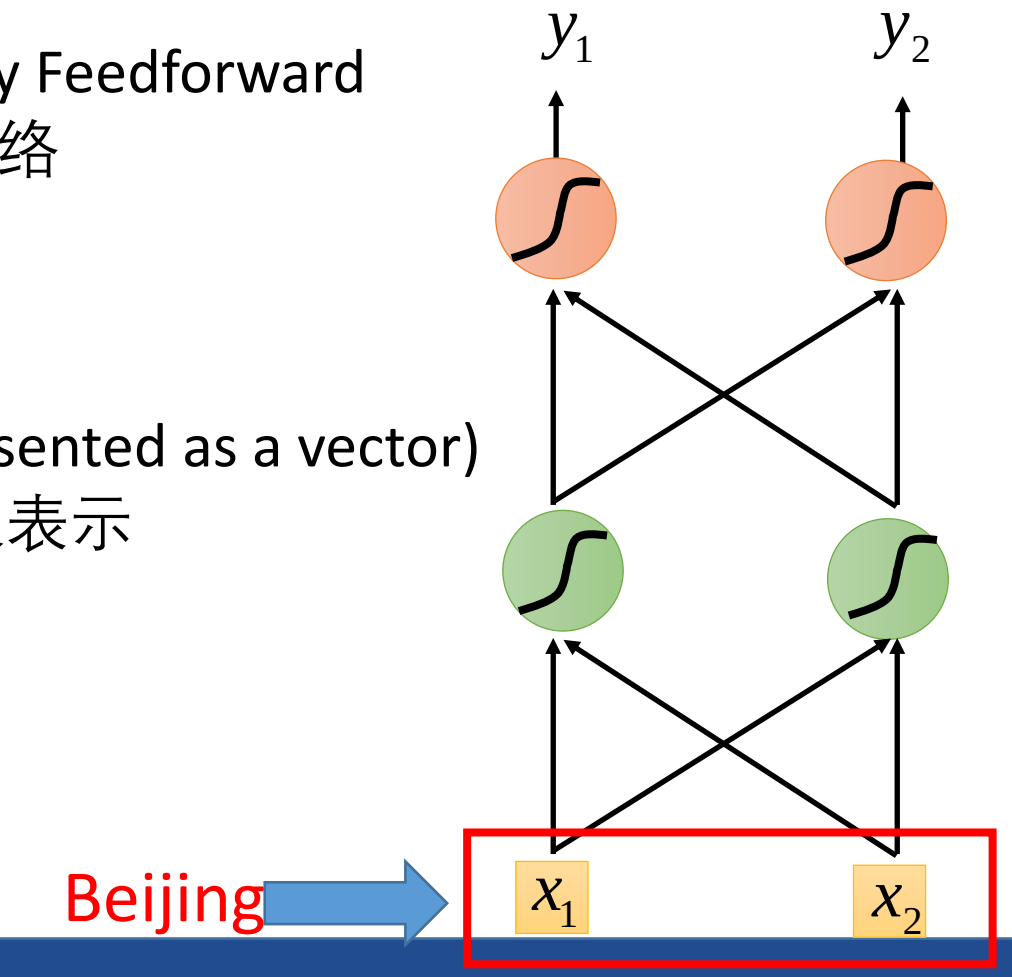
Destination:	Beijing
time of arrival:	September 27th

Example Application

Solving slot filling by Feedforward network?正反馈网络

Input: a word
输入： 单词

(Each word is represented as a vector)
单词可以用矢量来表示



1-of-N encoding

How to represent each word as a vector?
如何将单词表达成为一个矢量?

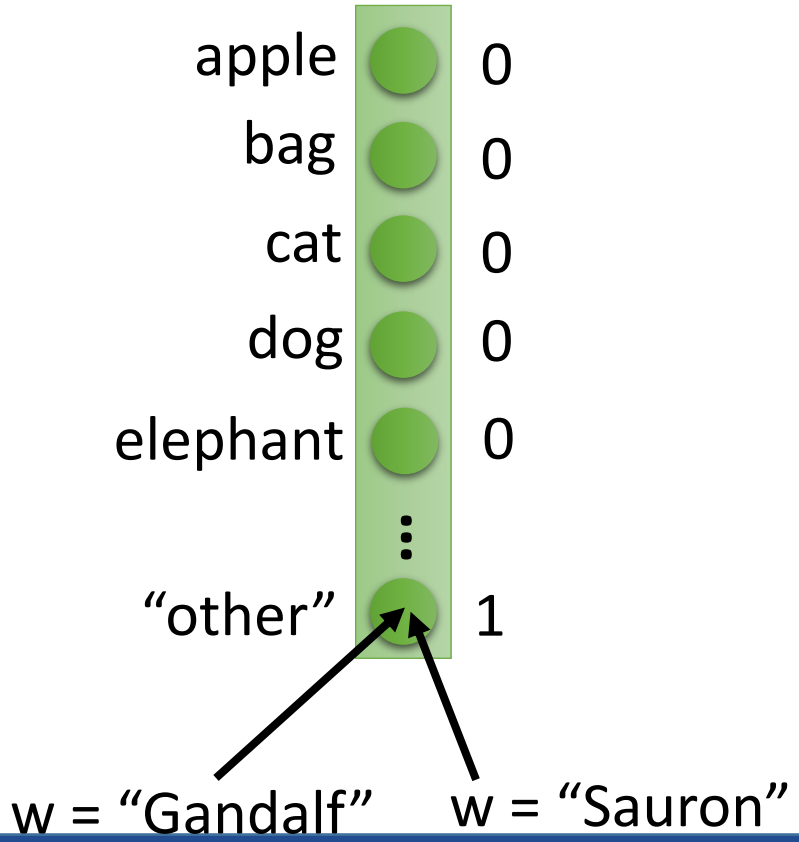
1-of-N Encoding lexicon 词典= {apple, bag, cat, dog, elephant}

单一矢量长度为词典长度
每个维度都有固定的含义
数值只能0或1

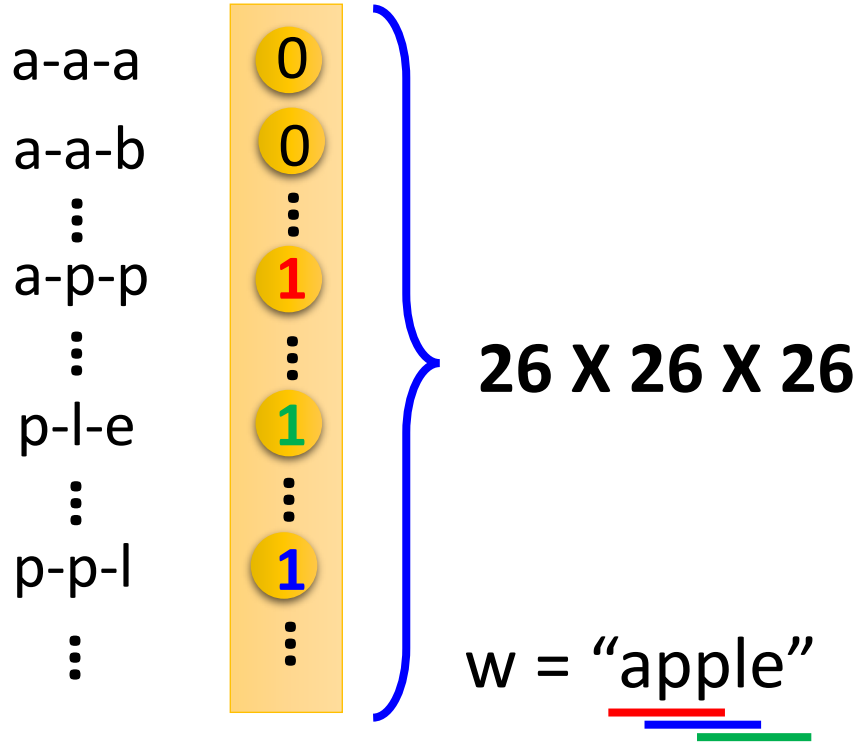
apple = [1 0 0 0 0]
bag = [0 1 0 0 0]
cat = [0 0 1 0 0]
dog = [0 0 0 1 0]
elephant = [0 0 0 0 1]

Beyond 1-of-N encoding

Dimension for "Other"



Word hashing



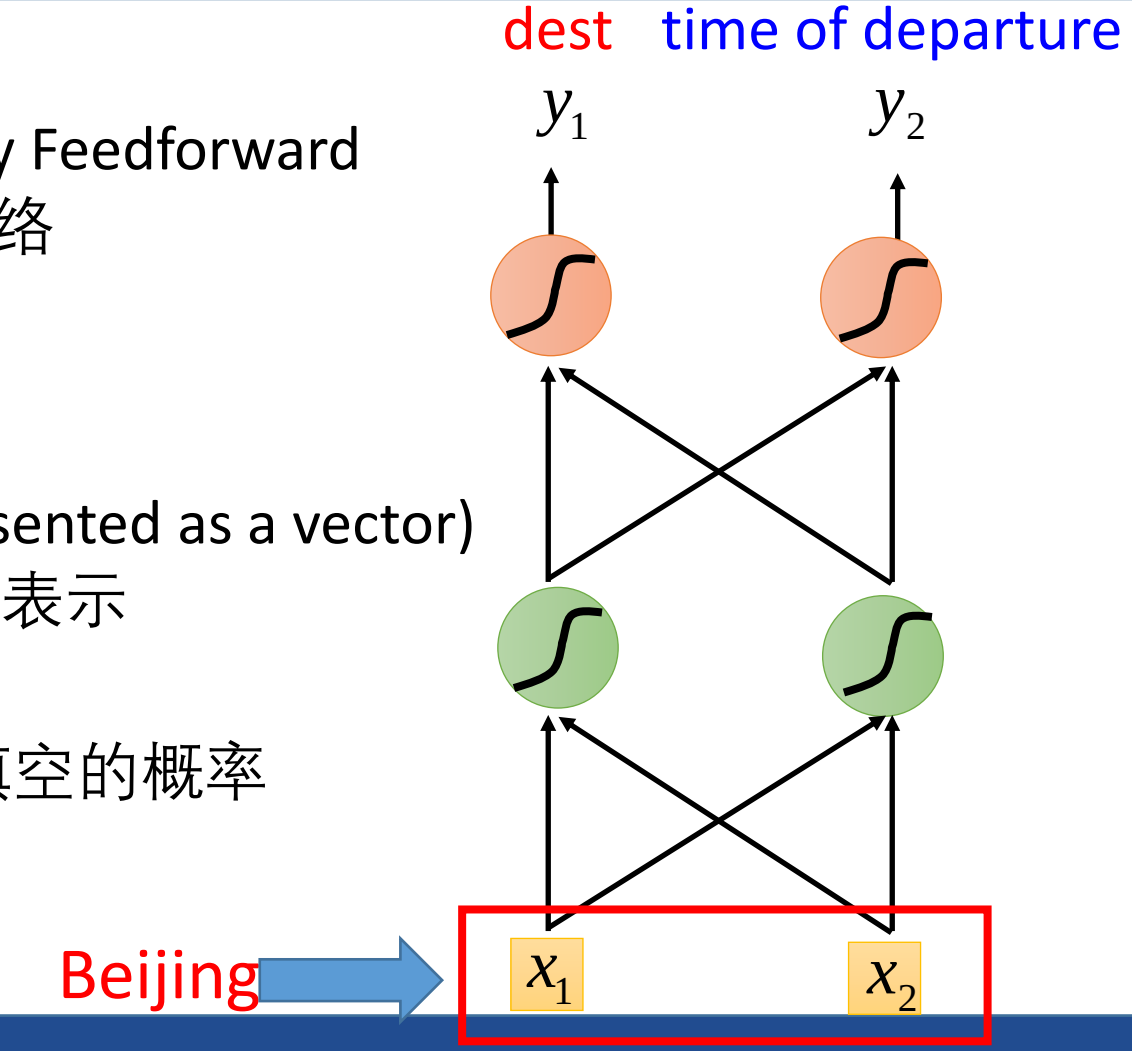
Example Application

Solving slot filling by Feedforward network?正反馈网络

Input: a word
输入： 单词

(Each word is represented as a vector)
单词可以用矢量来表示

Output:输出
输入单词是正确填空的概率



Example Application

arrive Beijing on September 27th

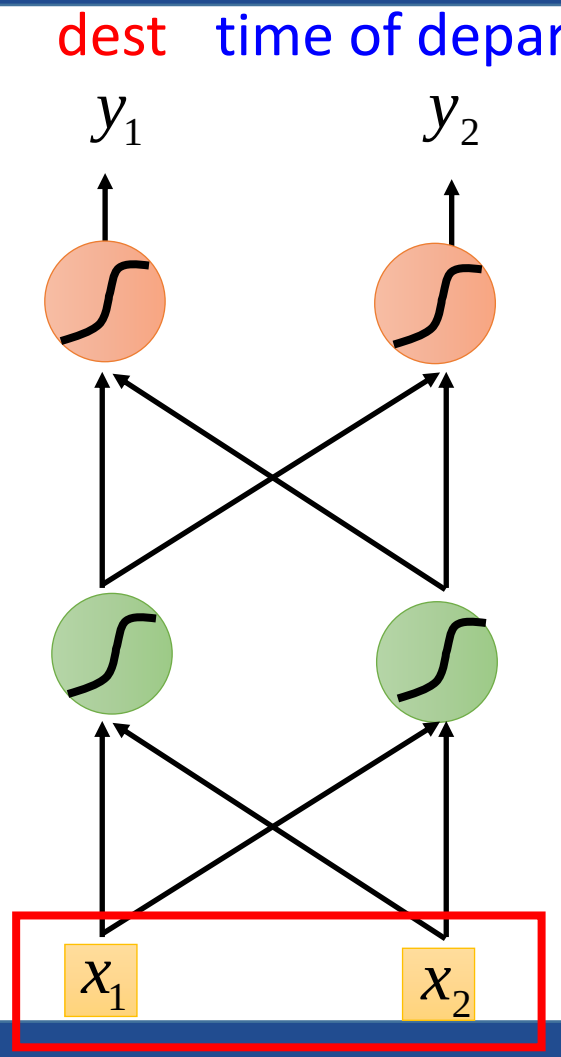
other dest other time time

Problem?

leave Beijing on September 27th

place of departure

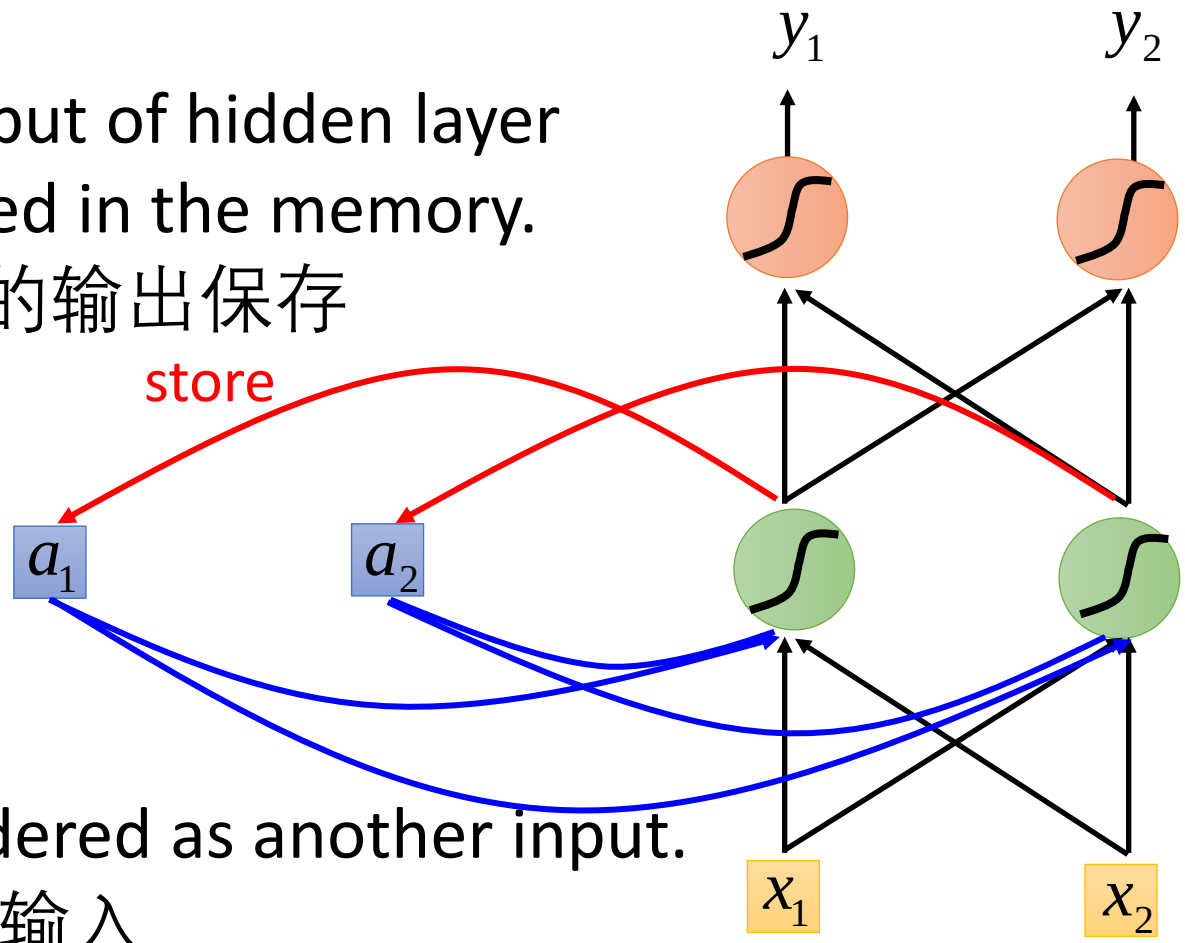
Neural network needs memory!



Beijing

Recurrent Neural Network (RNN)

The output of hidden layer are stored in the memory.
隐藏层的输出保存



Memory can be considered as another input.
存储值可视为额外的输入

Example

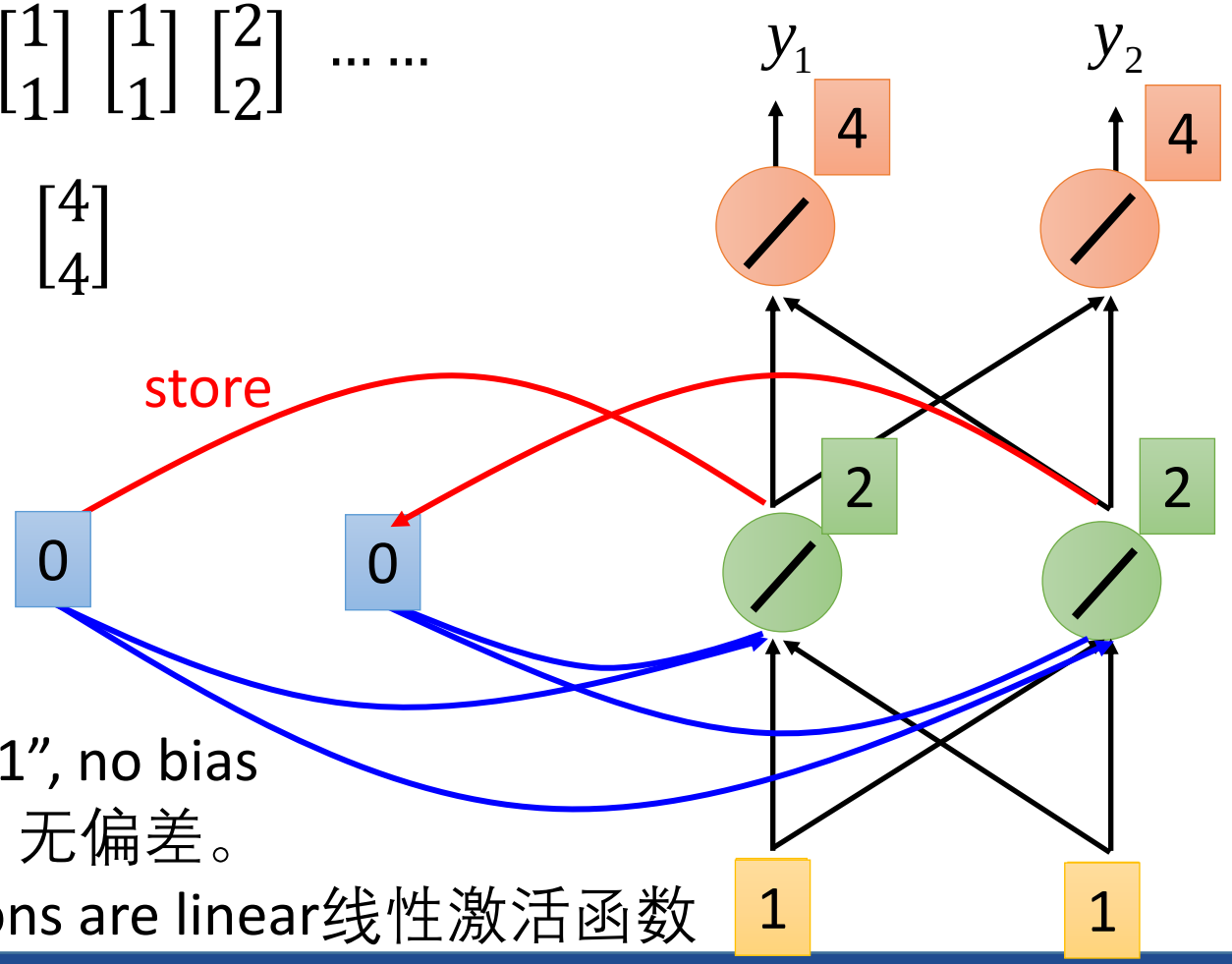
Input sequence:
输入序列

$\begin{bmatrix}1\\1\end{bmatrix}$ $\begin{bmatrix}1\\1\end{bmatrix}$ $\begin{bmatrix}2\\2\end{bmatrix}$ ……

output sequence:
输出序列

$\begin{bmatrix}4\\4\end{bmatrix}$

given Initial values
初始化值

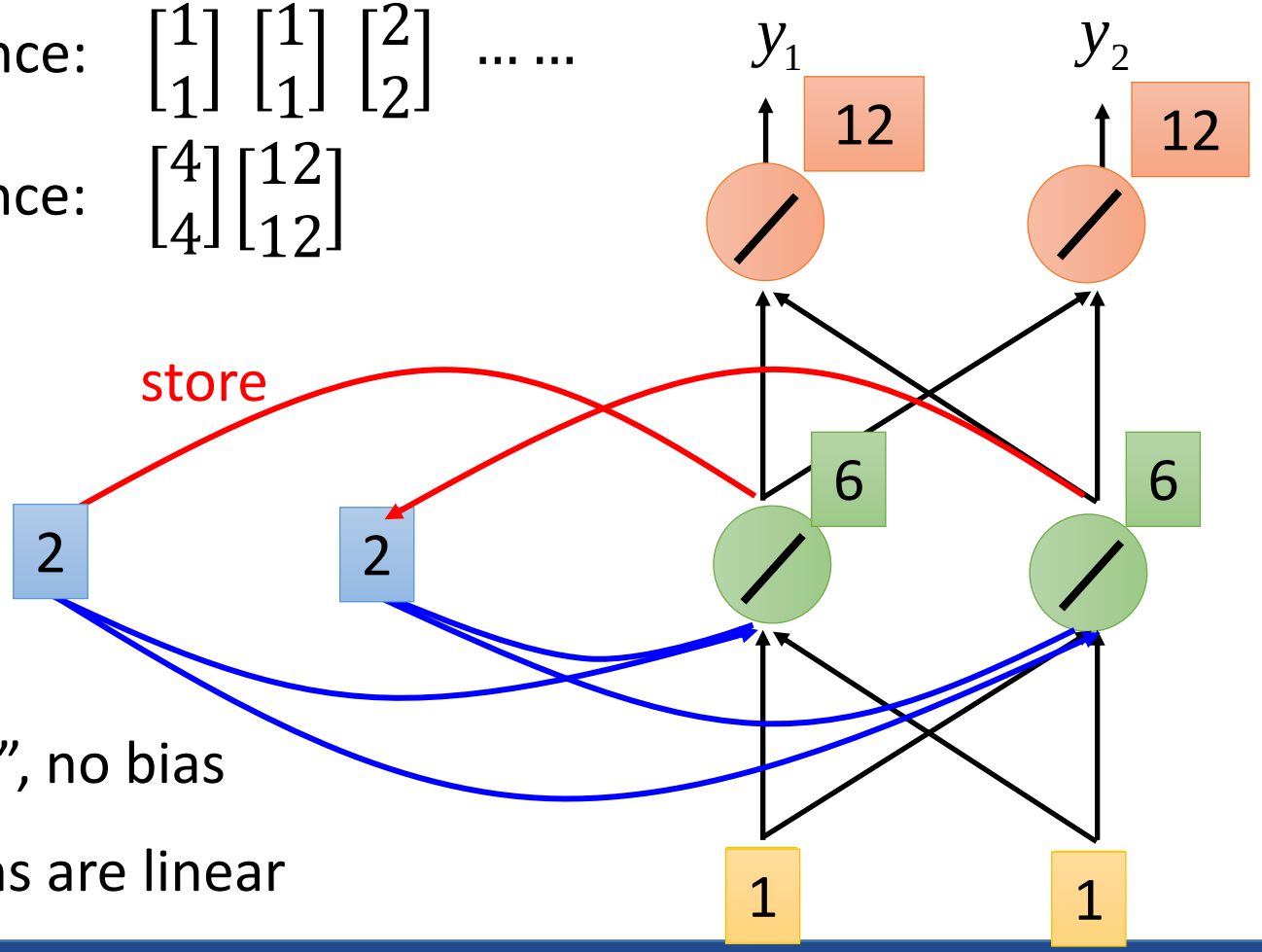


All the weights are “1”, no bias
所有的权重都为1，无偏差。
All activation functions are linear线性激活函数

Example

Input sequence: $\begin{bmatrix}1\\1\end{bmatrix}$ $\begin{bmatrix}1\\1\end{bmatrix}$ $\begin{bmatrix}2\\2\end{bmatrix}$ ……

output sequence: $\begin{bmatrix}4\\4\end{bmatrix}\begin{bmatrix}12\\12\end{bmatrix}$



All the weights are “1”, no bias

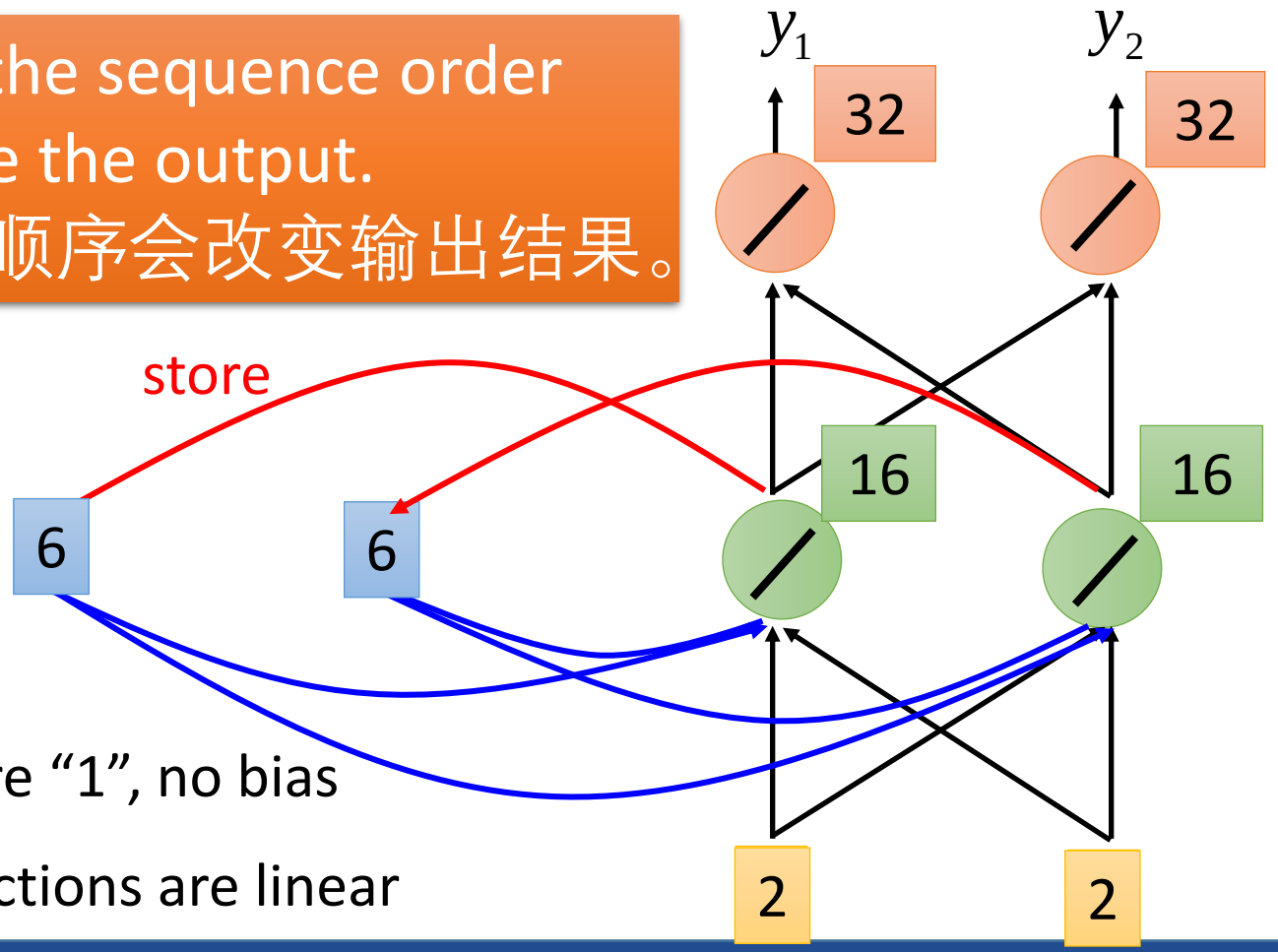
All activation functions are linear

Example

Input sequence: $\begin{bmatrix}1\\1\end{bmatrix}\ \begin{bmatrix}1\\1\end{bmatrix}\ \begin{bmatrix}2\\2\end{bmatrix}\ \cdots\cdots$

output sequence: $\begin{bmatrix}4\\4\end{bmatrix}\begin{bmatrix}12\\12\end{bmatrix}\begin{bmatrix}32\\32\end{bmatrix}$

Changing the sequence order will change the output.
改变序列顺序会改变输出结果。



All the weights are “1”, no bias

All activation functions are linear

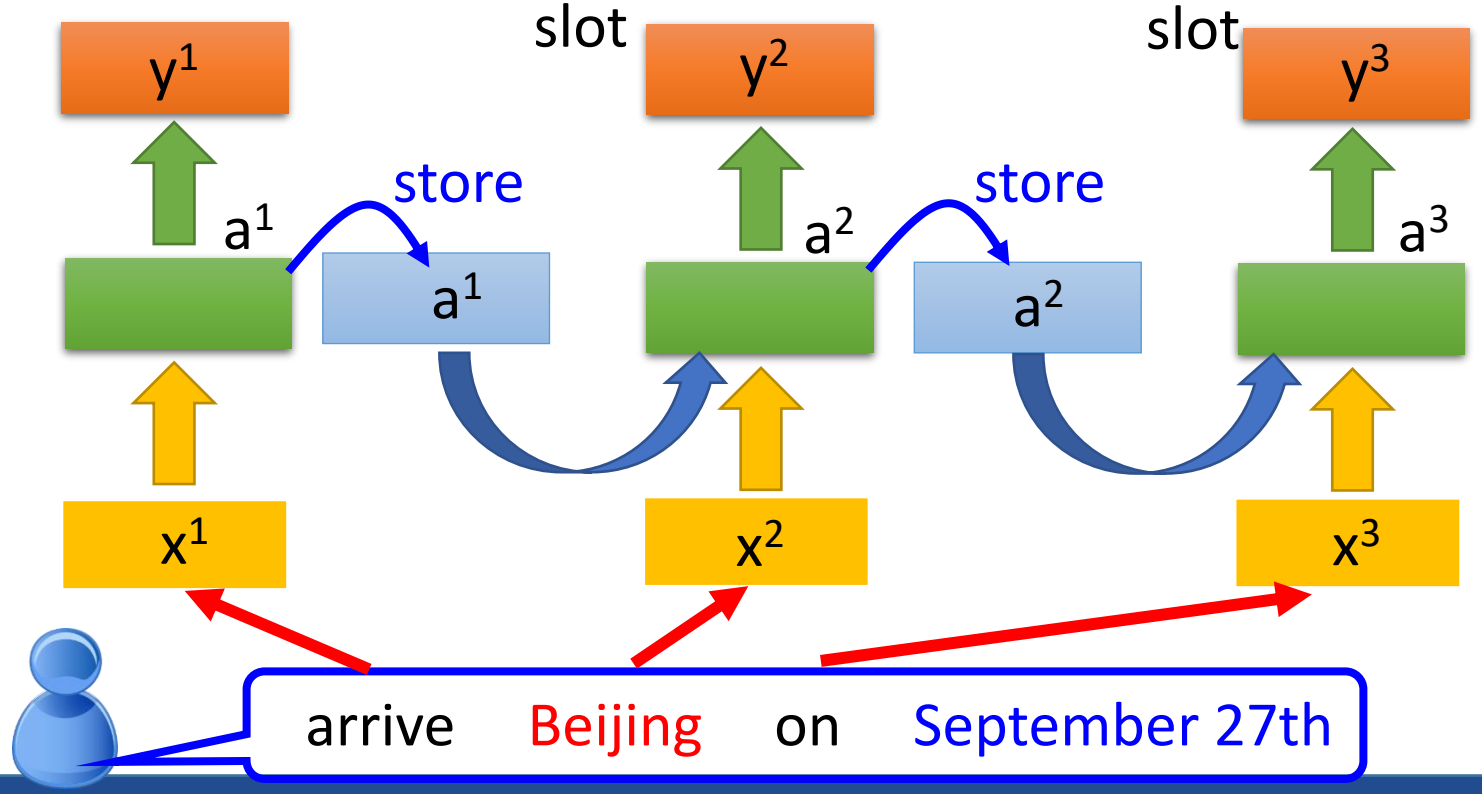
RNN

同样的网络重复计算

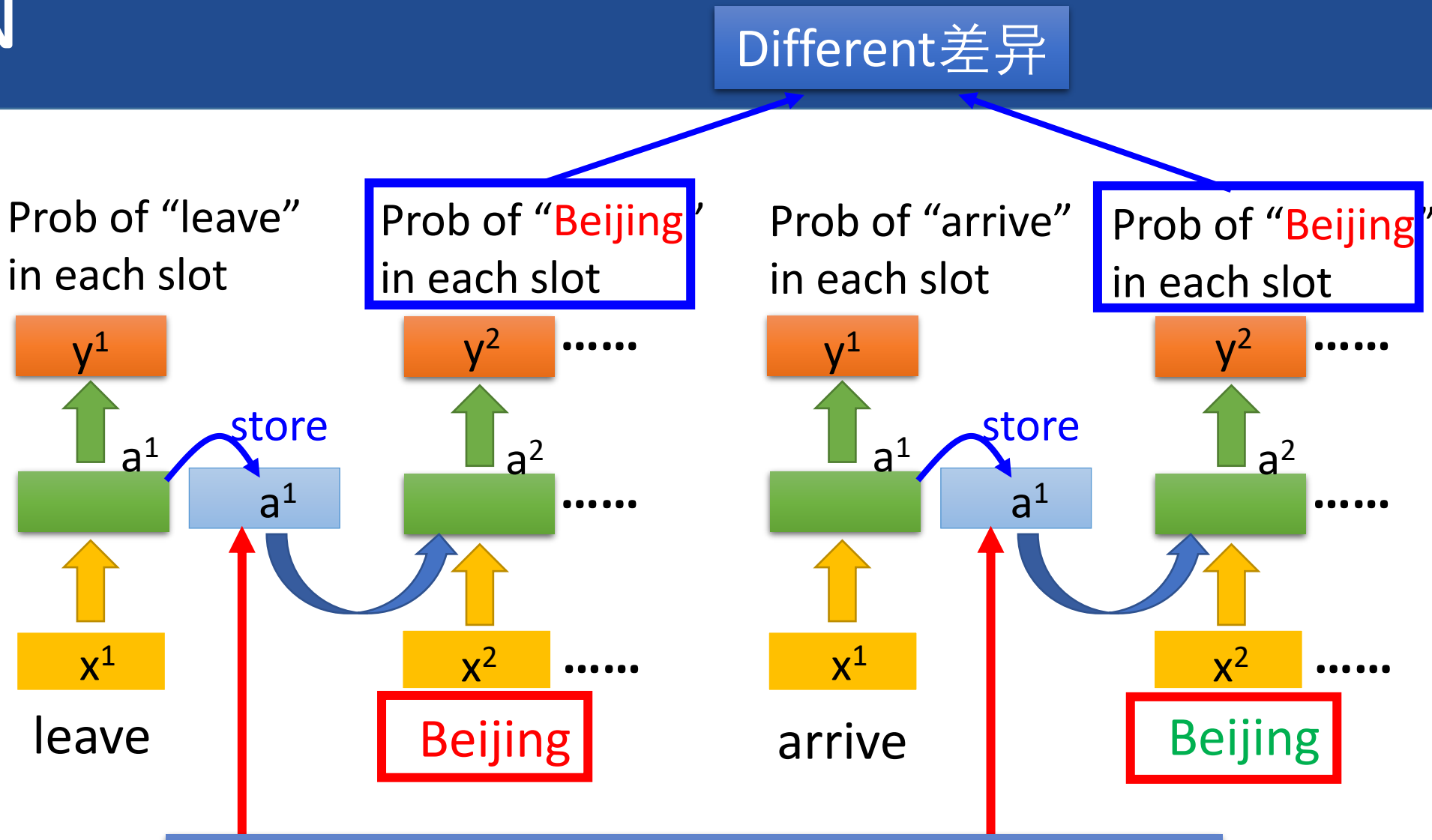
概率Probability of “arrive” in each slot

概率Probability of “Beijing” in each slot

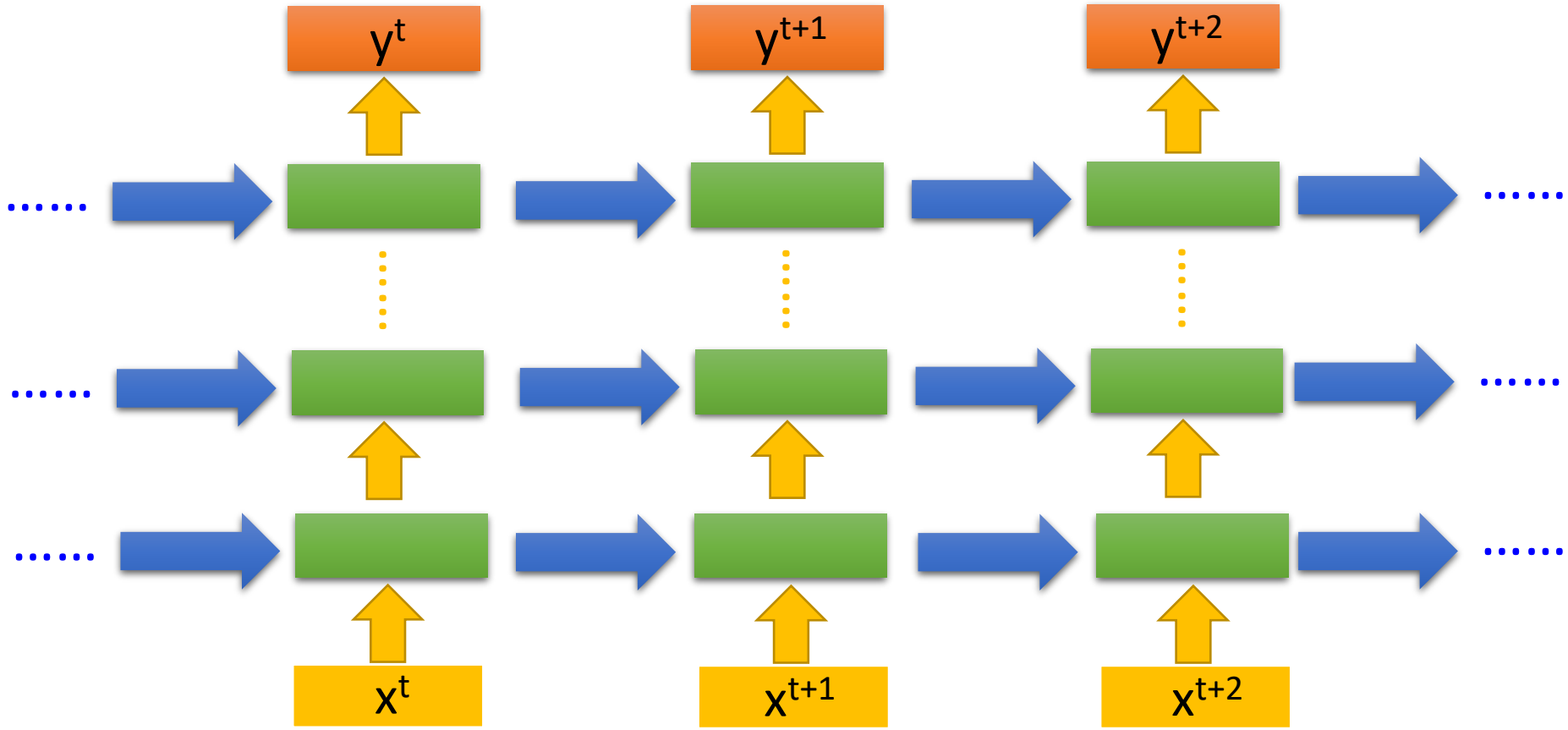
概率Probability of “on” in each slot



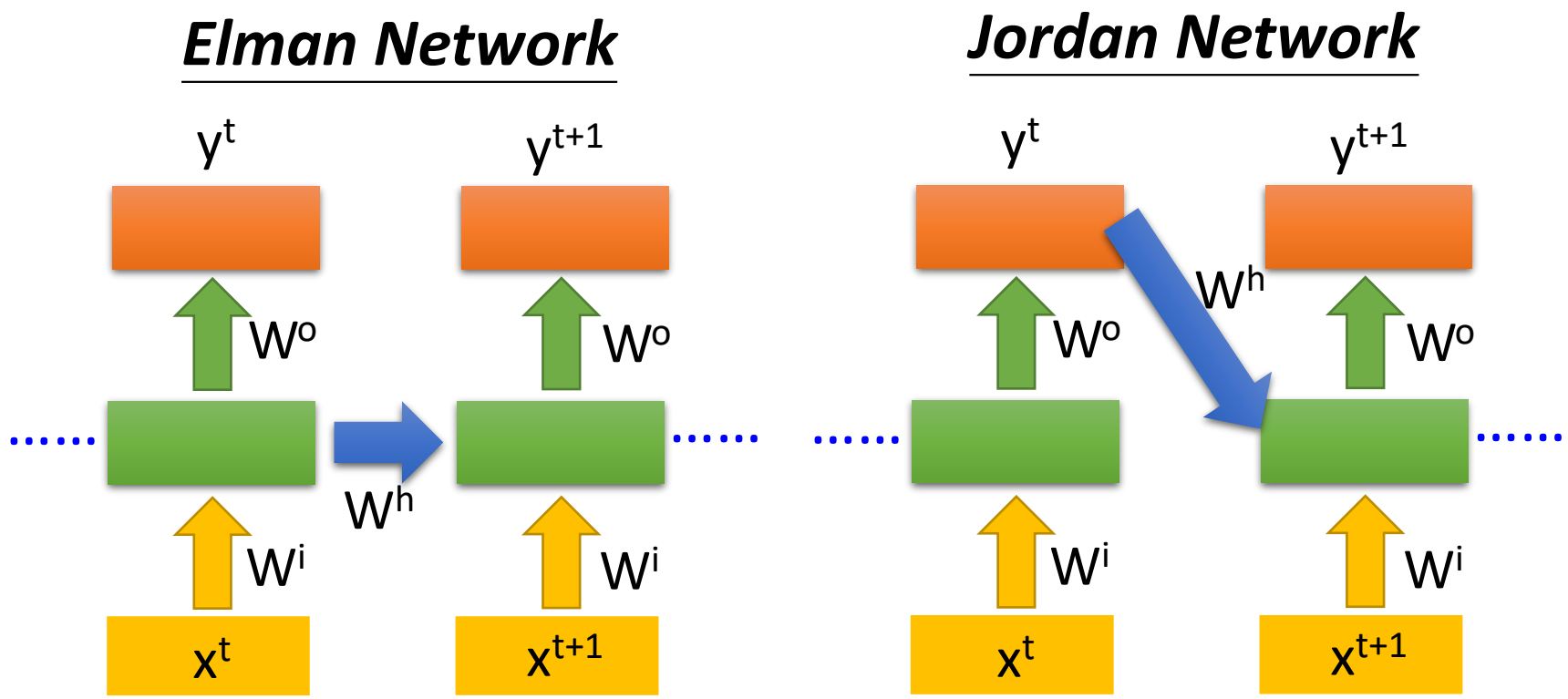
RNN


Different差异
Prob of "leave" in each slot
Prob of "Beijing" in each slot
Prob of "arrive" in each slot
Prob of "Beijing" in each slot
y1
y2
a1
a2
store
x1
x2
leave
Beijing
arrive
Beijing
The values stored in the memory is different.
存储值不同

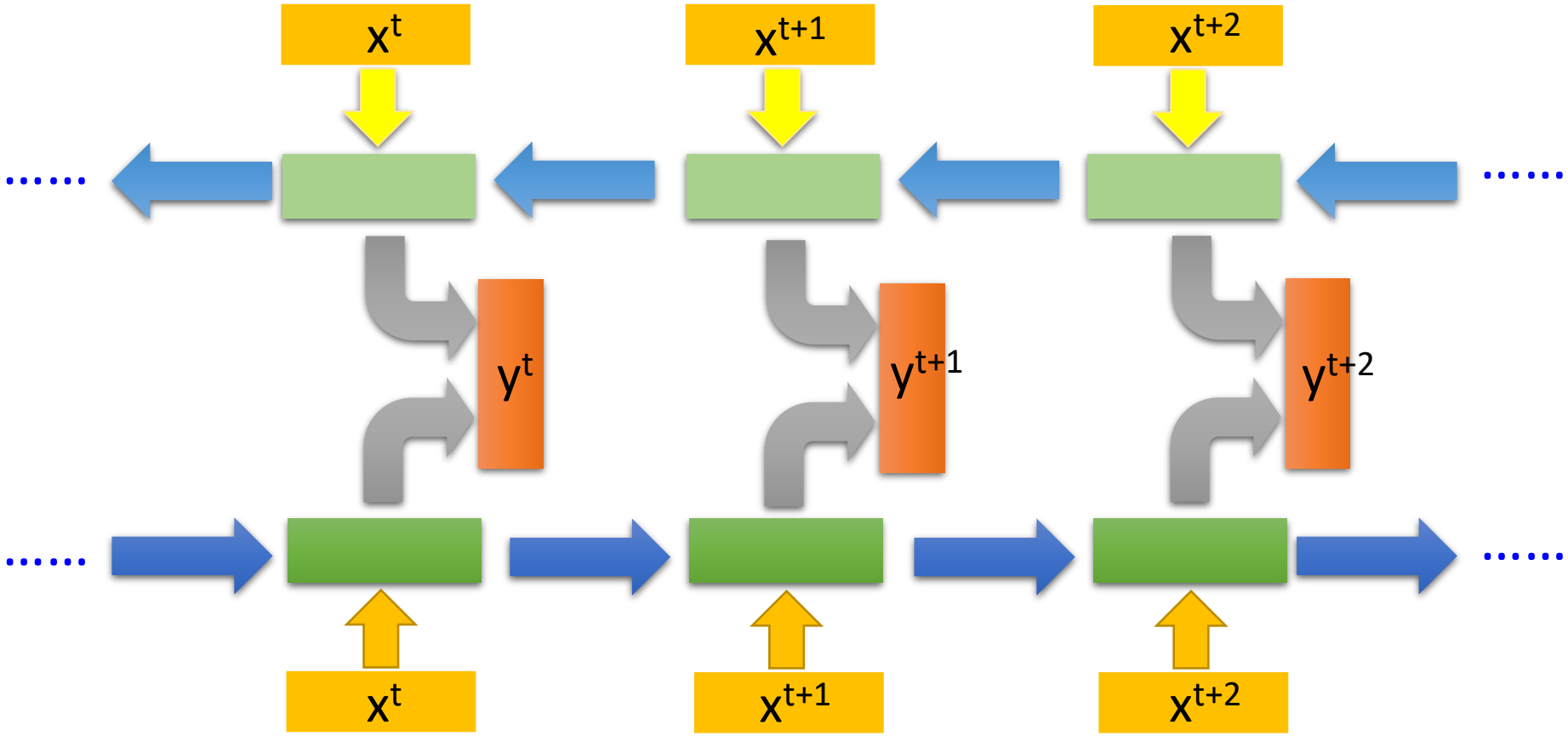
Of course it can be deep …

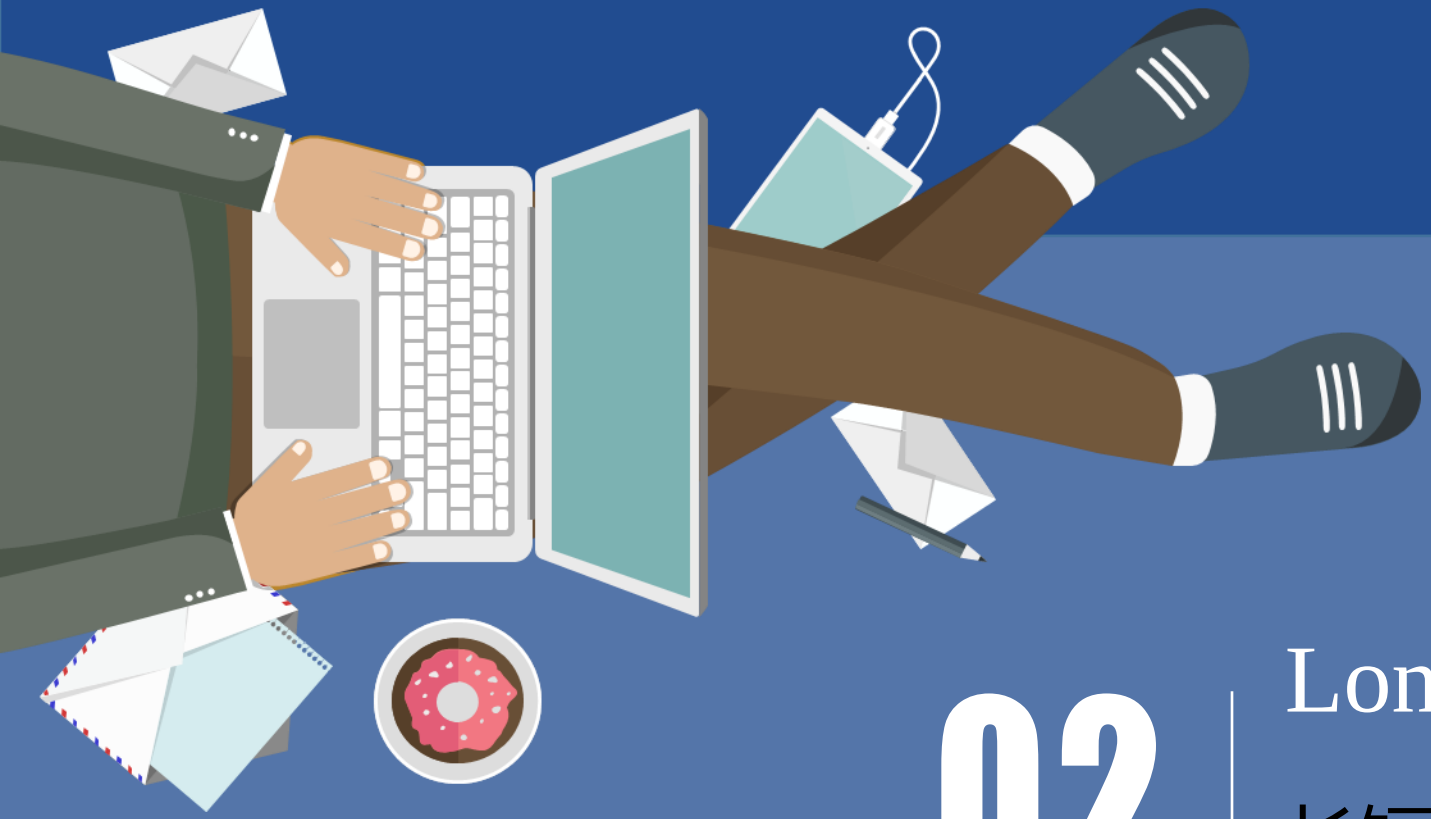


Elman Network & Jordan Network



Bidirectional RNN





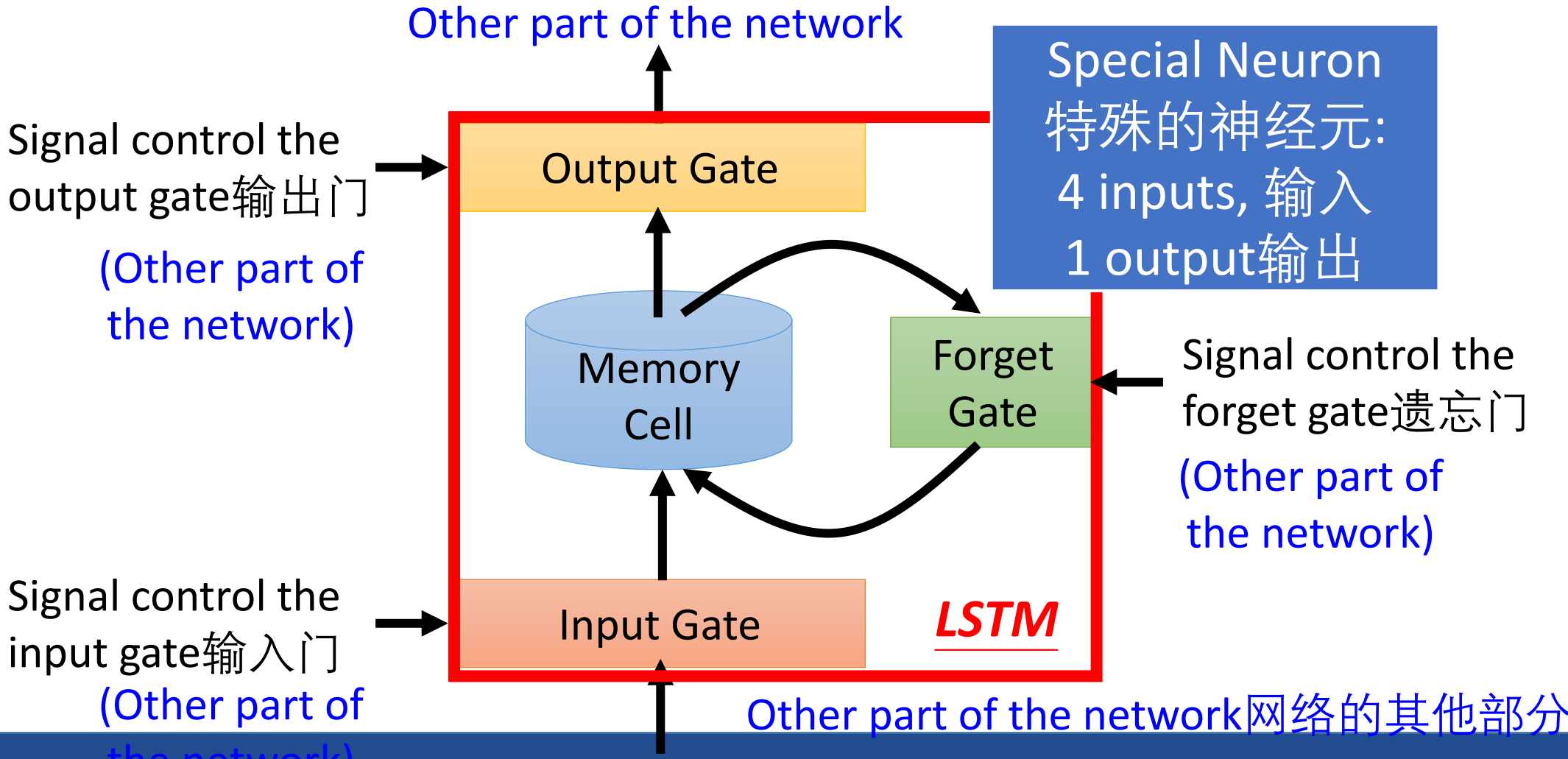
02 Long Short-term Memory

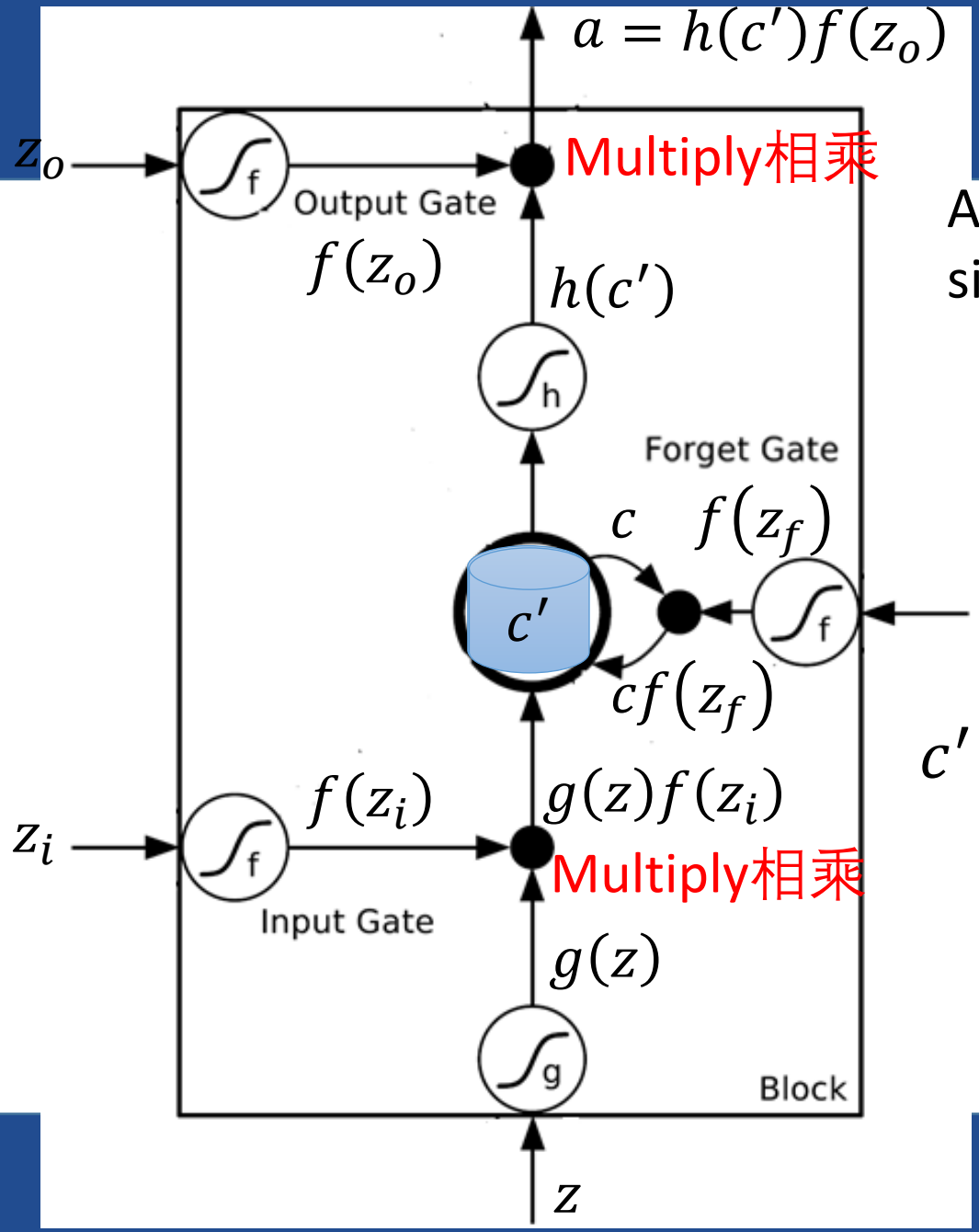
长短期记忆网络

Spring 2023

Introduction Video

Long Short-term Memory (LSTM)





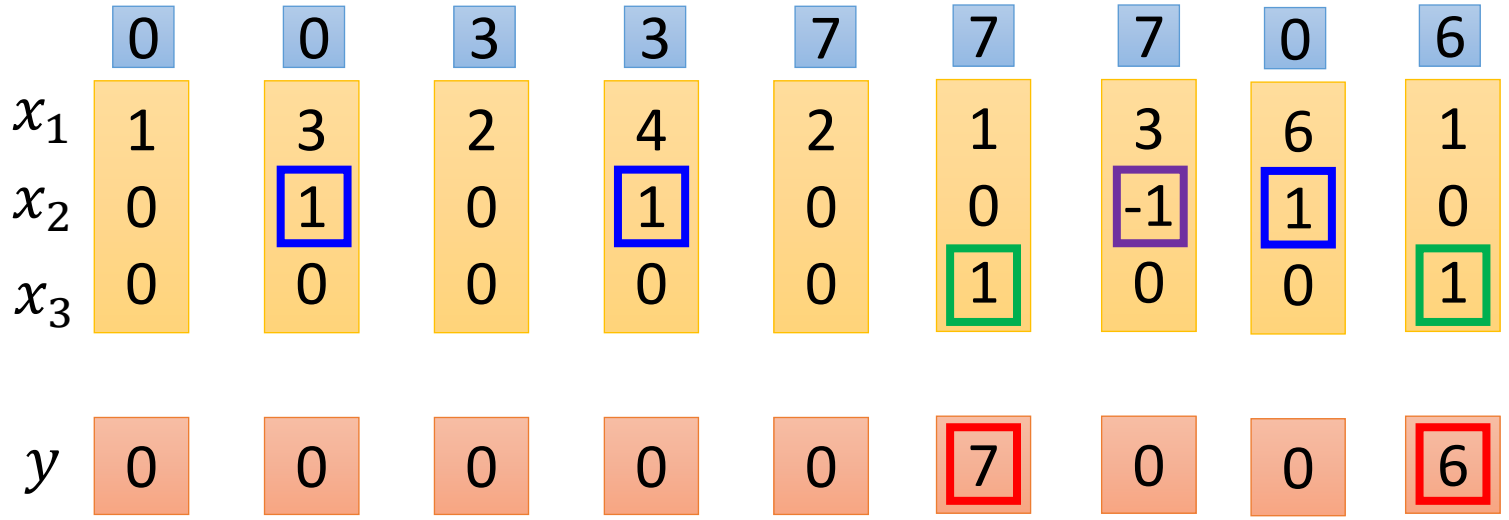
Activation function f is usually a sigmoid function激活函数

Between 0 and 1

Mimic open and close gate
模仿门的开关

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

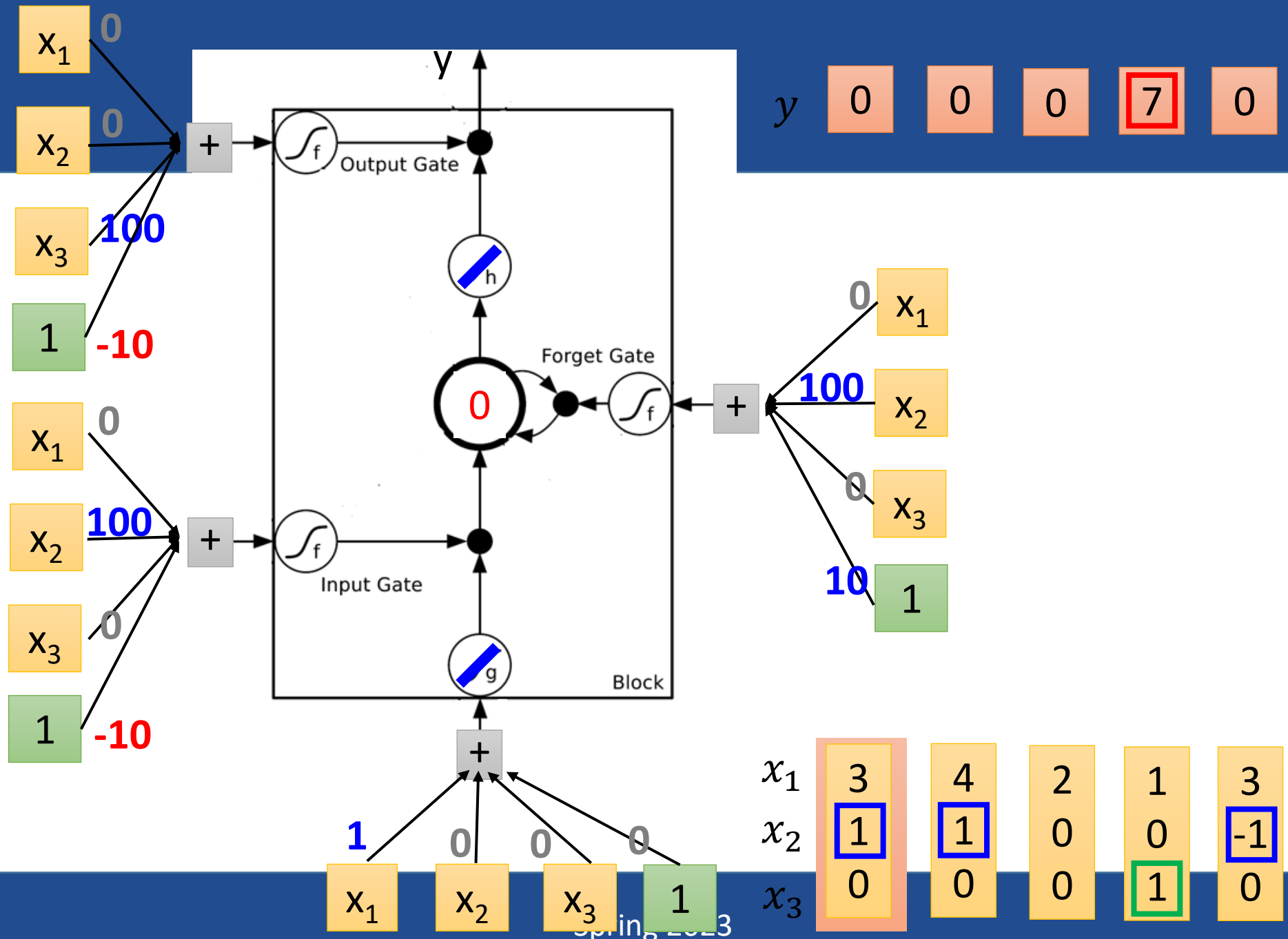
LSTM - Example



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

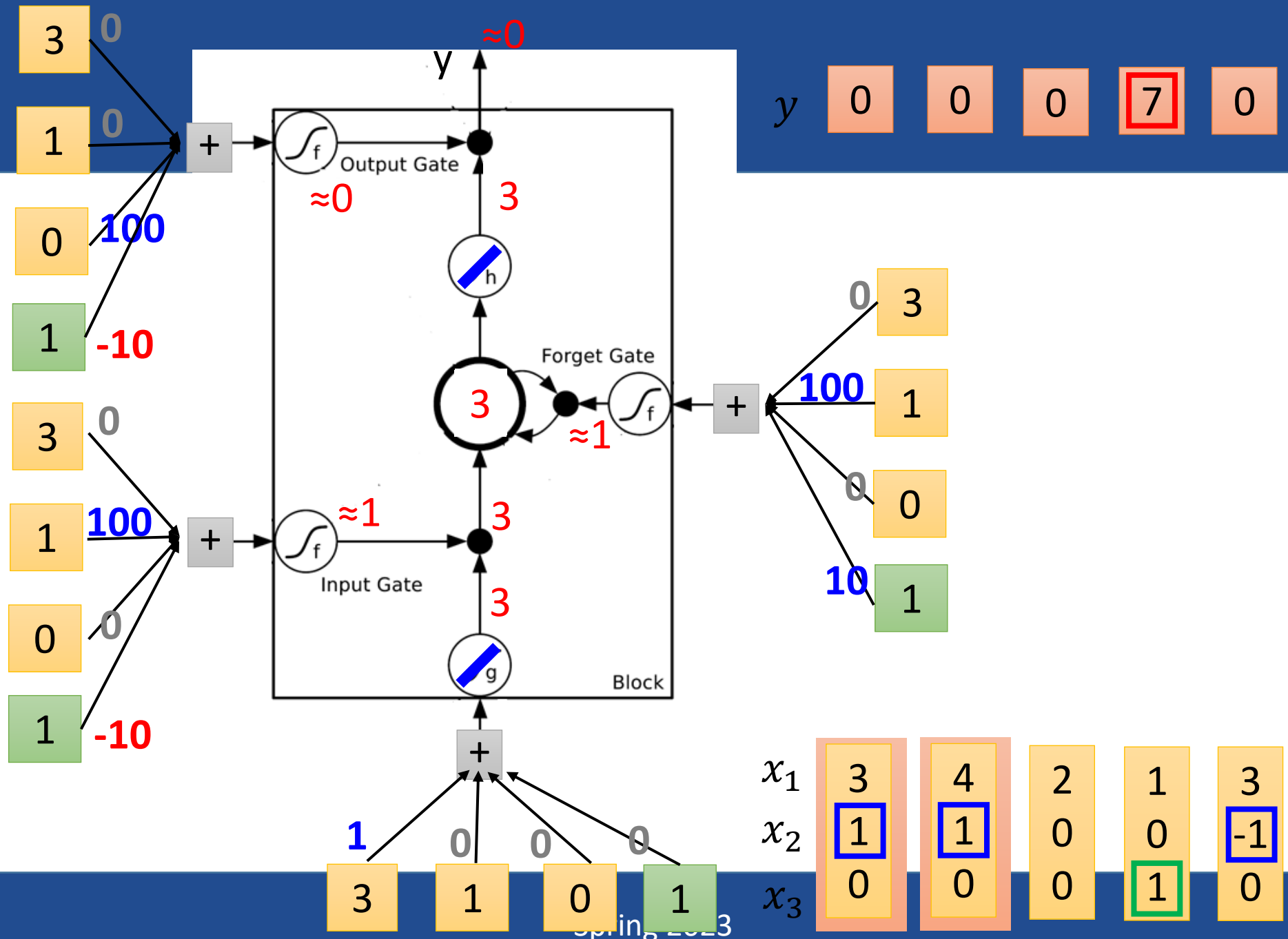
When x_2 = -1, reset the memory重置存储

When x_3 = 1, output the number from the memory.从存储输出



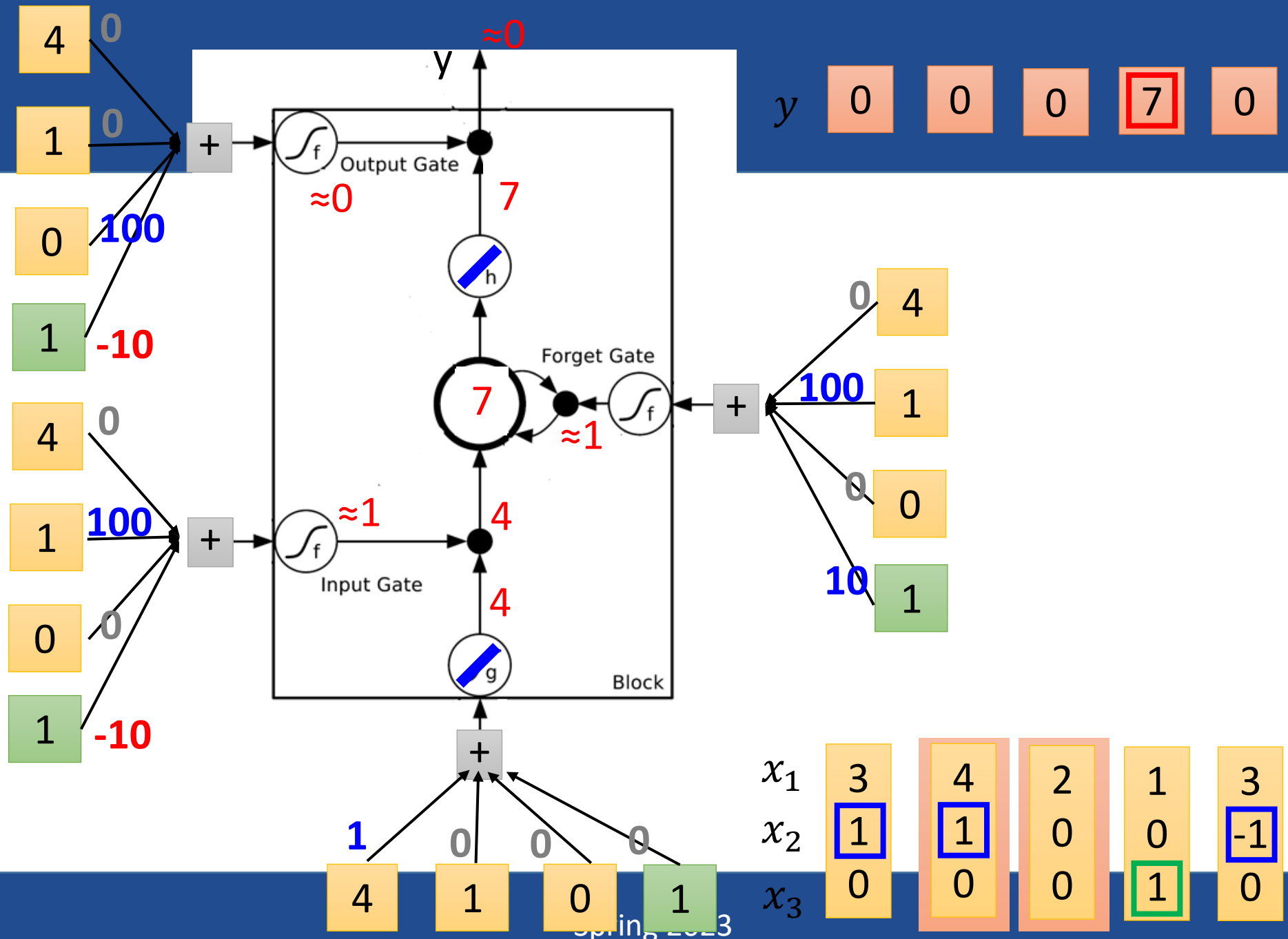
y					
	0	0	0	7	0

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



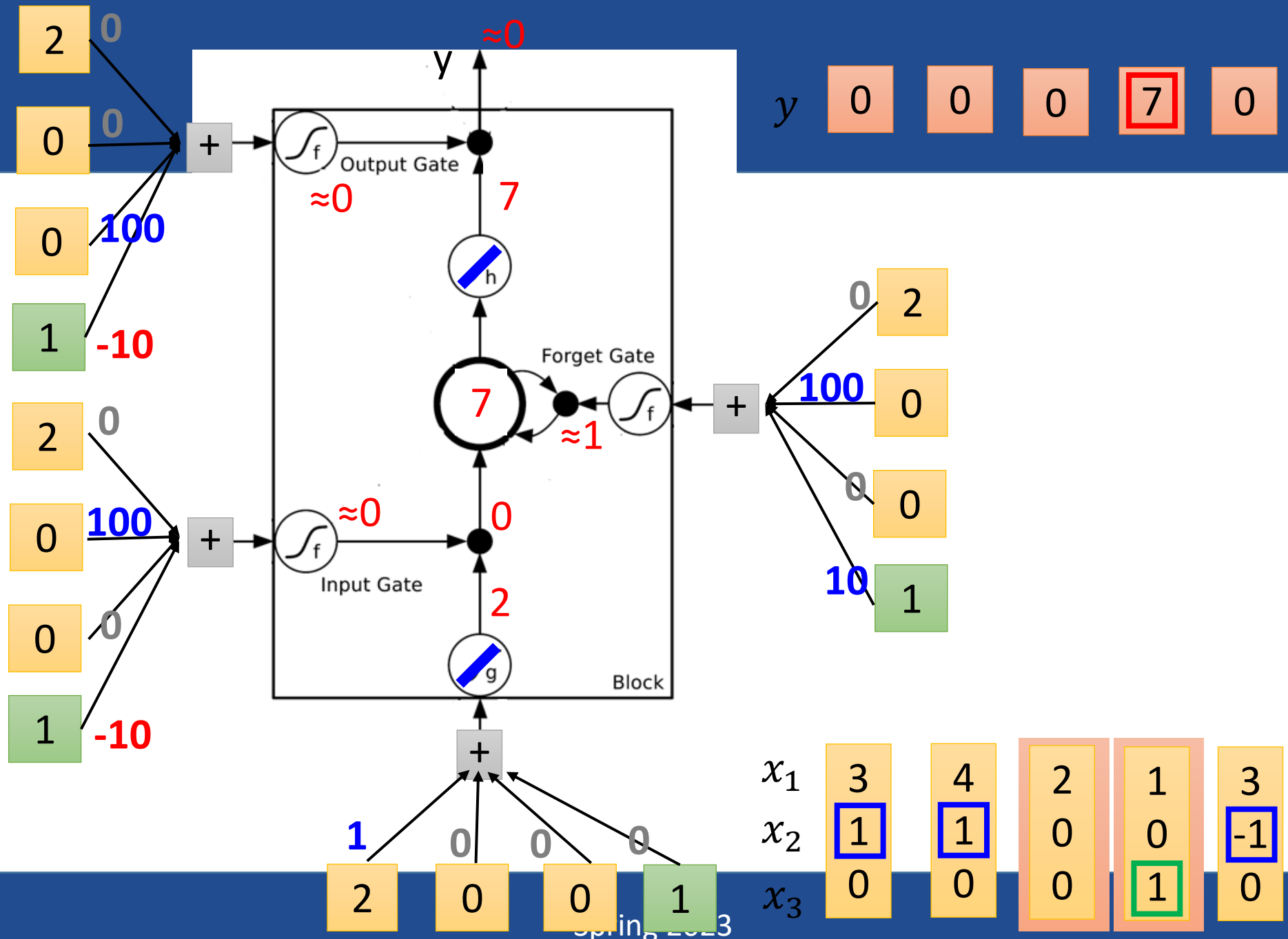
y	0	0	0	7	0

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



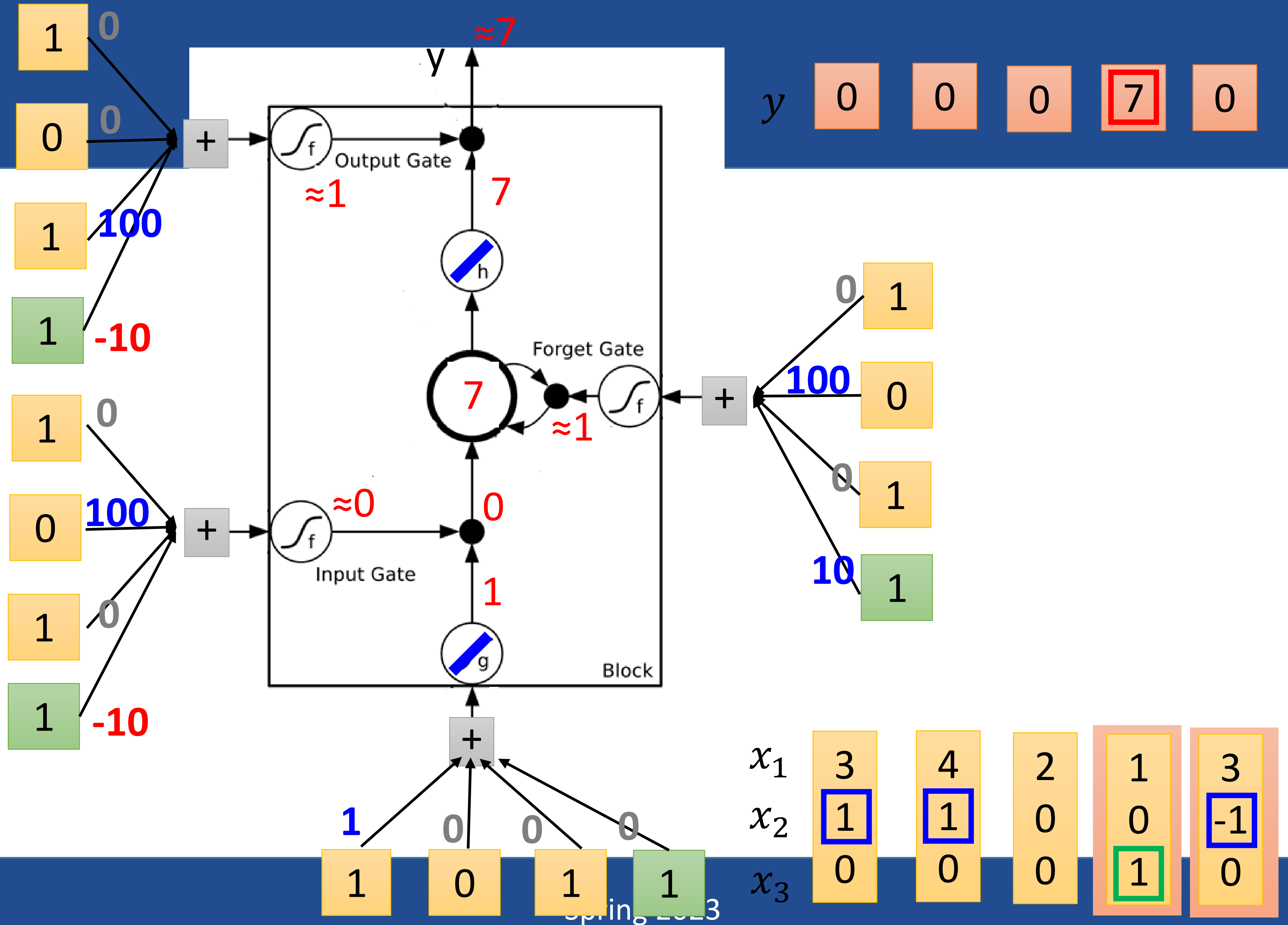
y	0	0	0	7	0

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

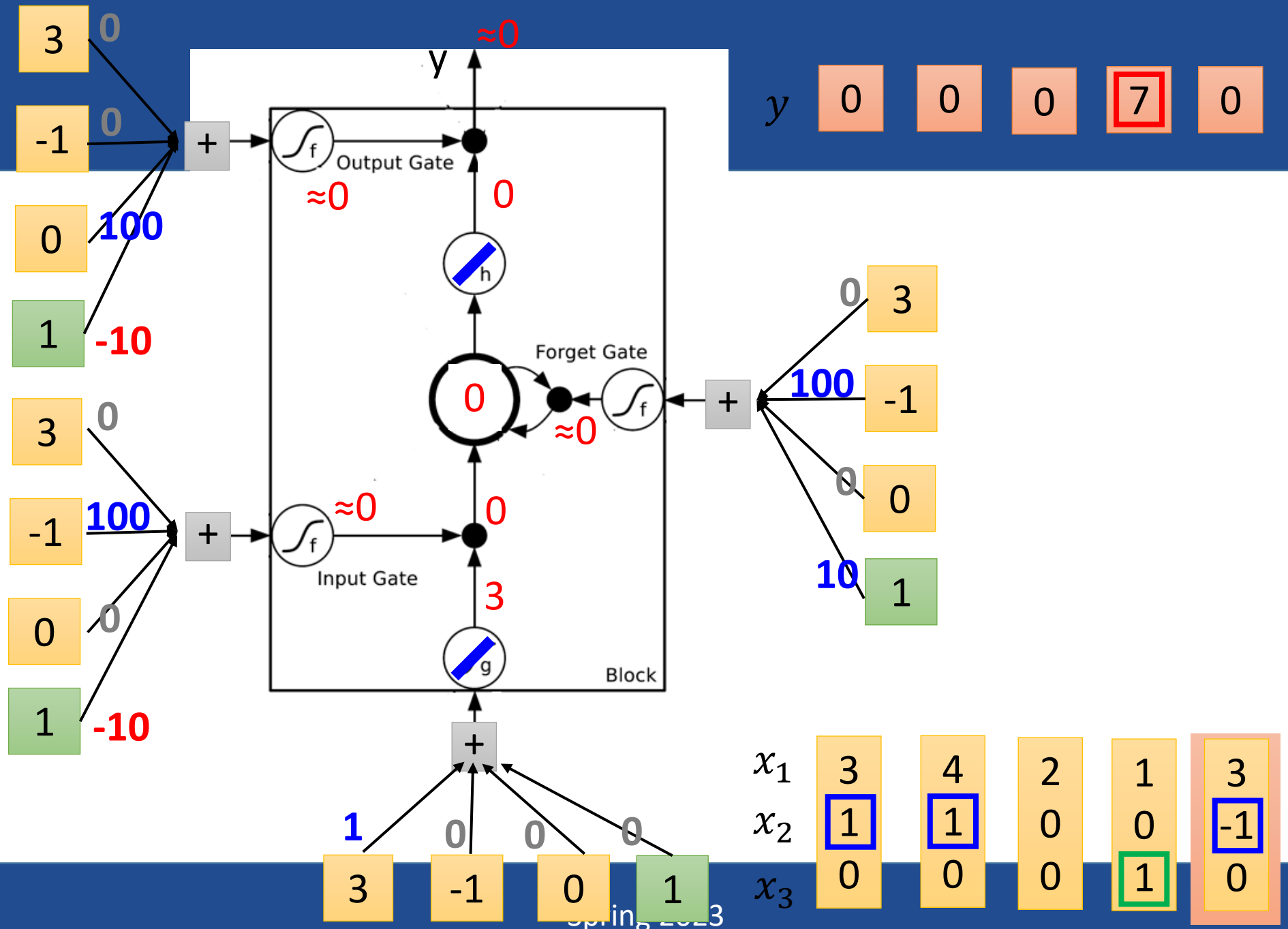


y	0	0	0	7	0

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



≈7
y
Output Gate
≈1
7
h
Forget Gate
7
≈1
≈0
0
Input Gate
1
g
Block
1 0 0 100 1 -10
1 0 0 100 1 0 1 -10
1 0 0 100 1 0 1 10
1 1 0 0 1 0 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

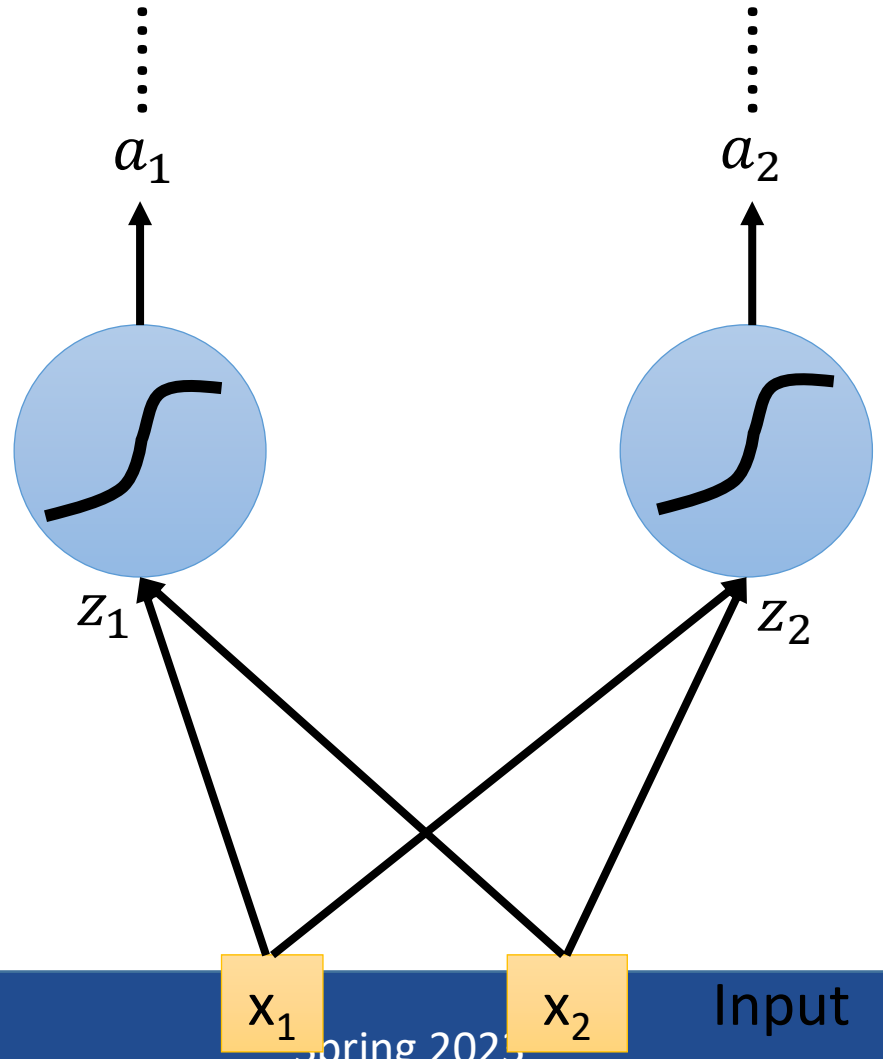


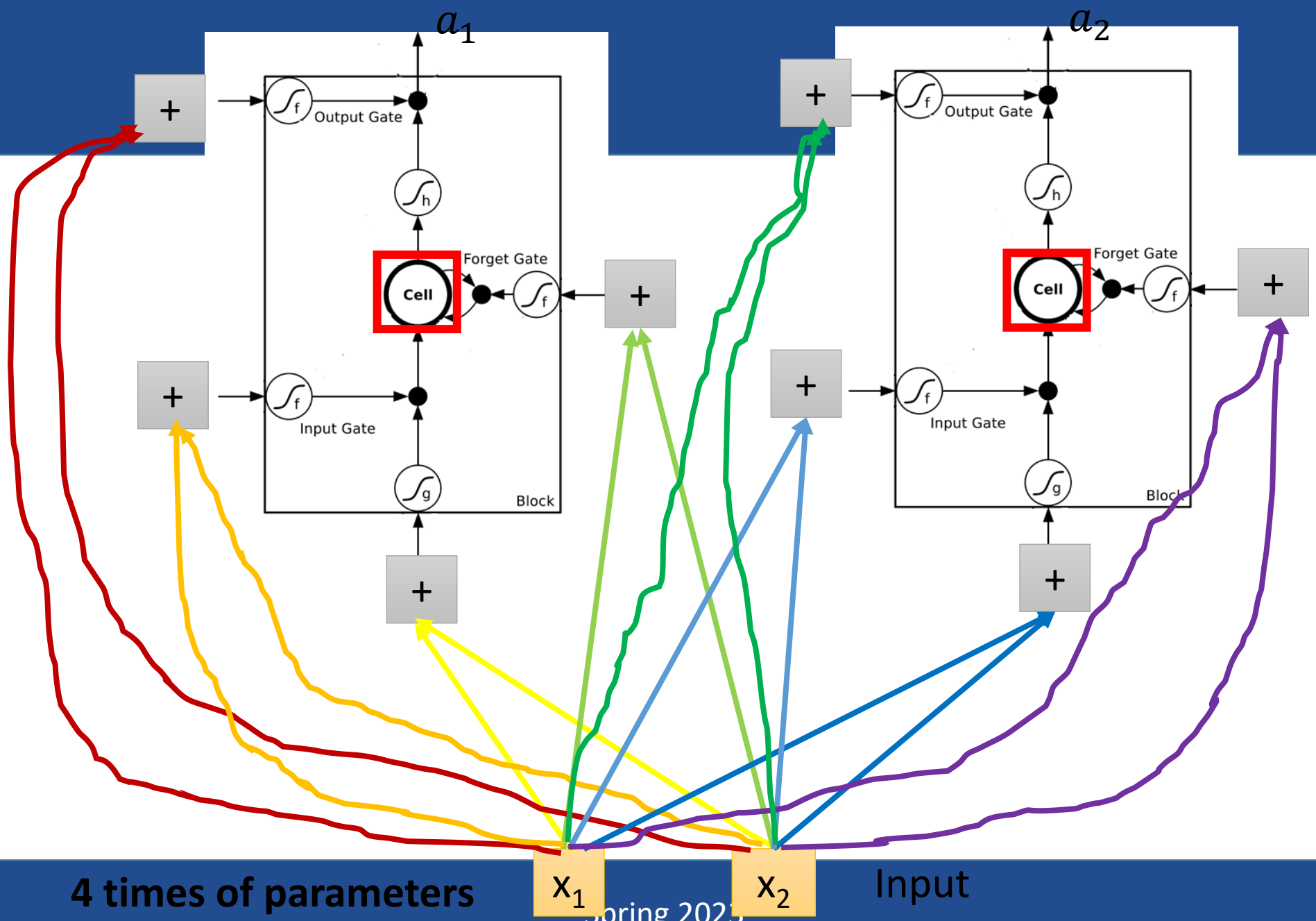
y	0	0	0	7	0

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

Original Network:原始的网络

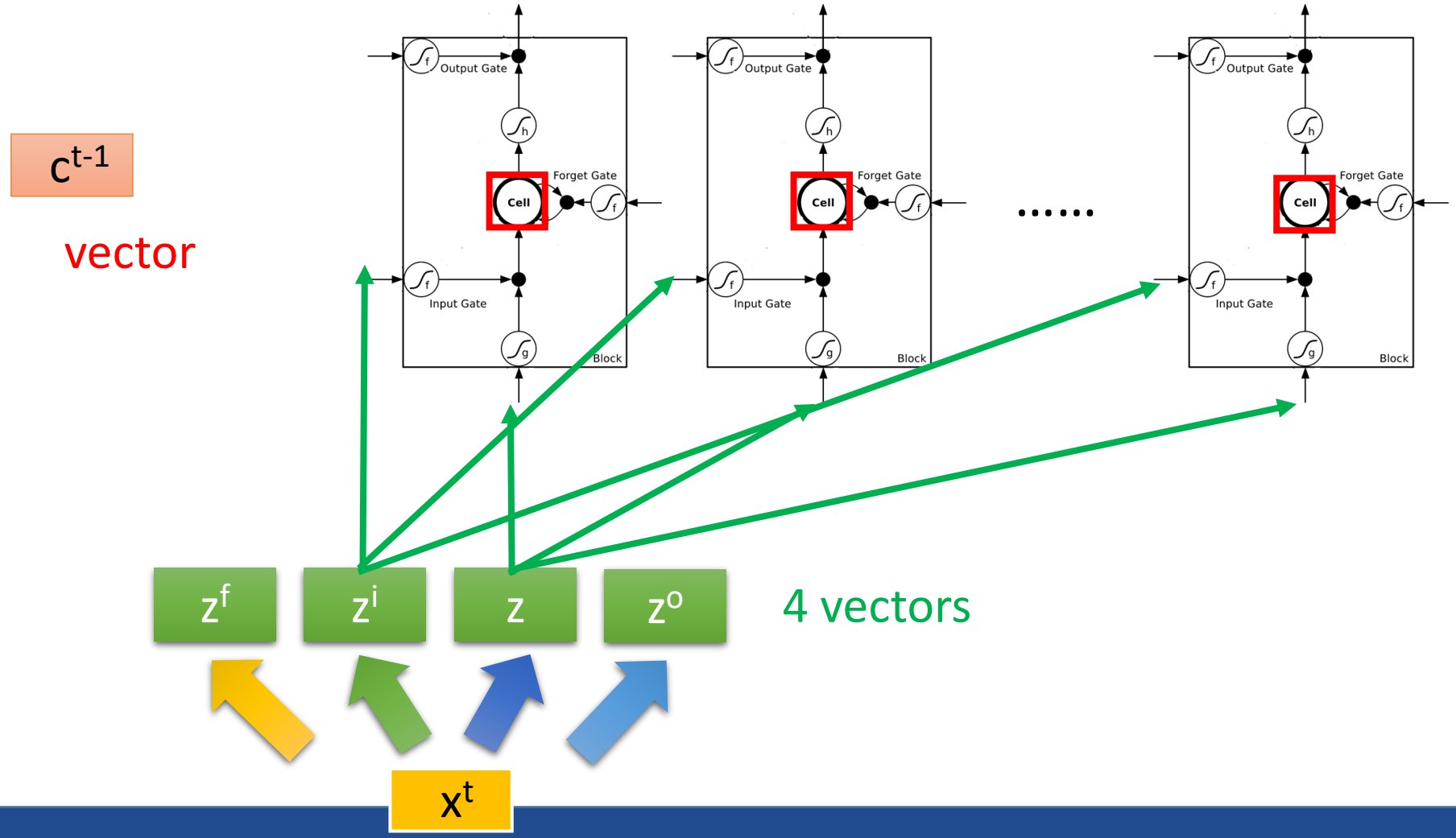
- Simply replace the neurons with LSTM



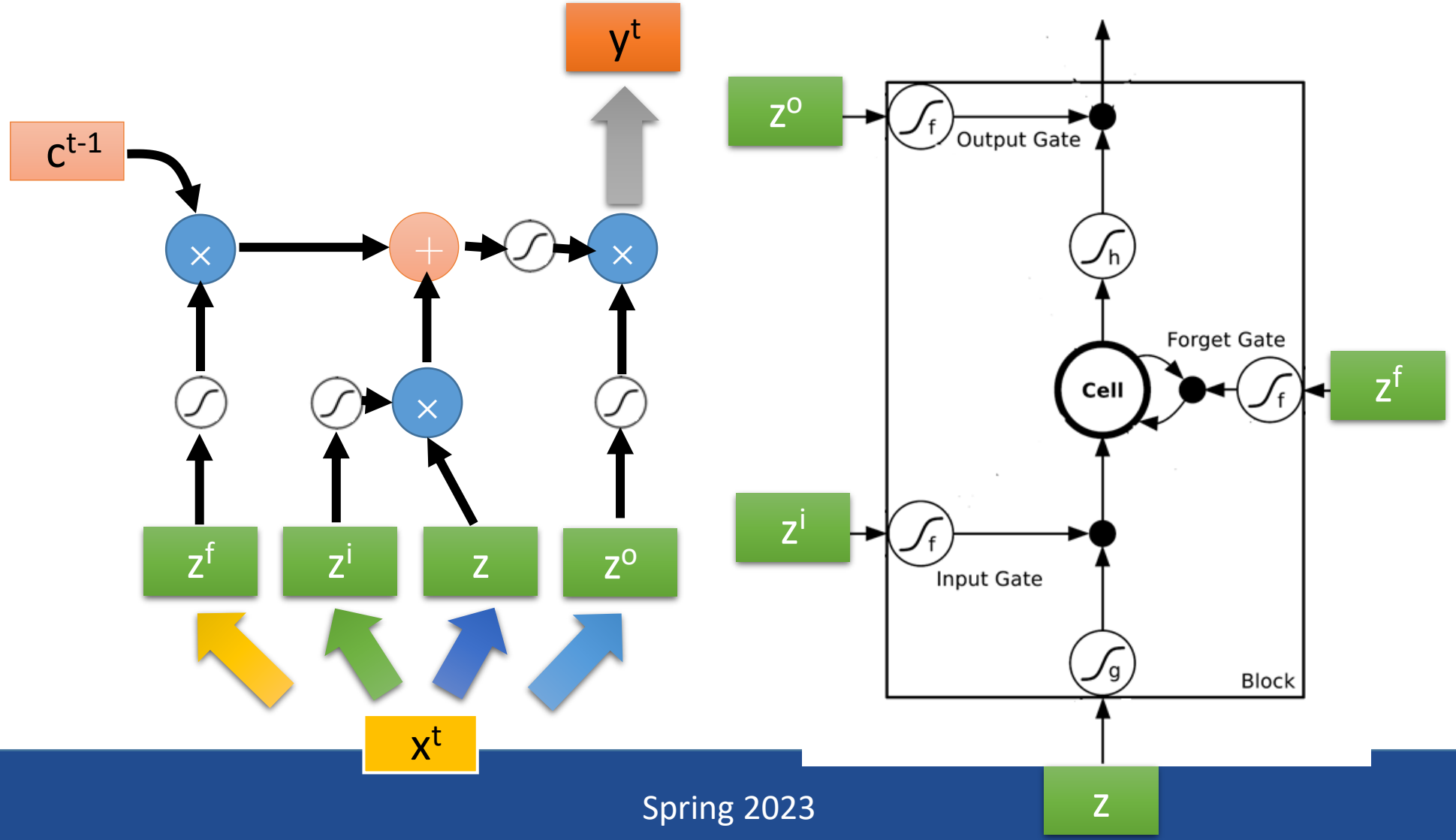


4 times of parameters

Long Short-term Memory (LSTM)

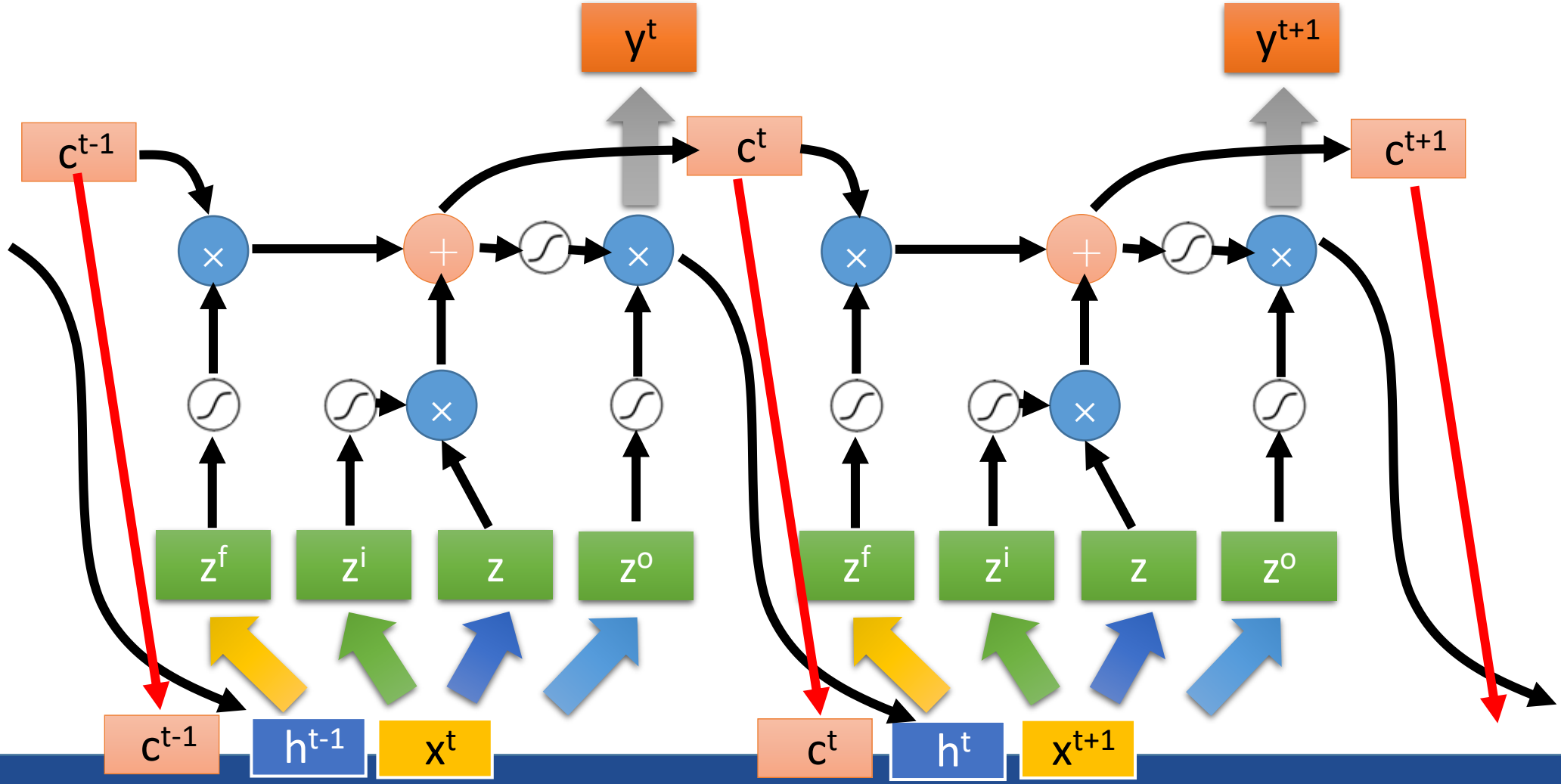


Long Short-term Memory (LSTM)

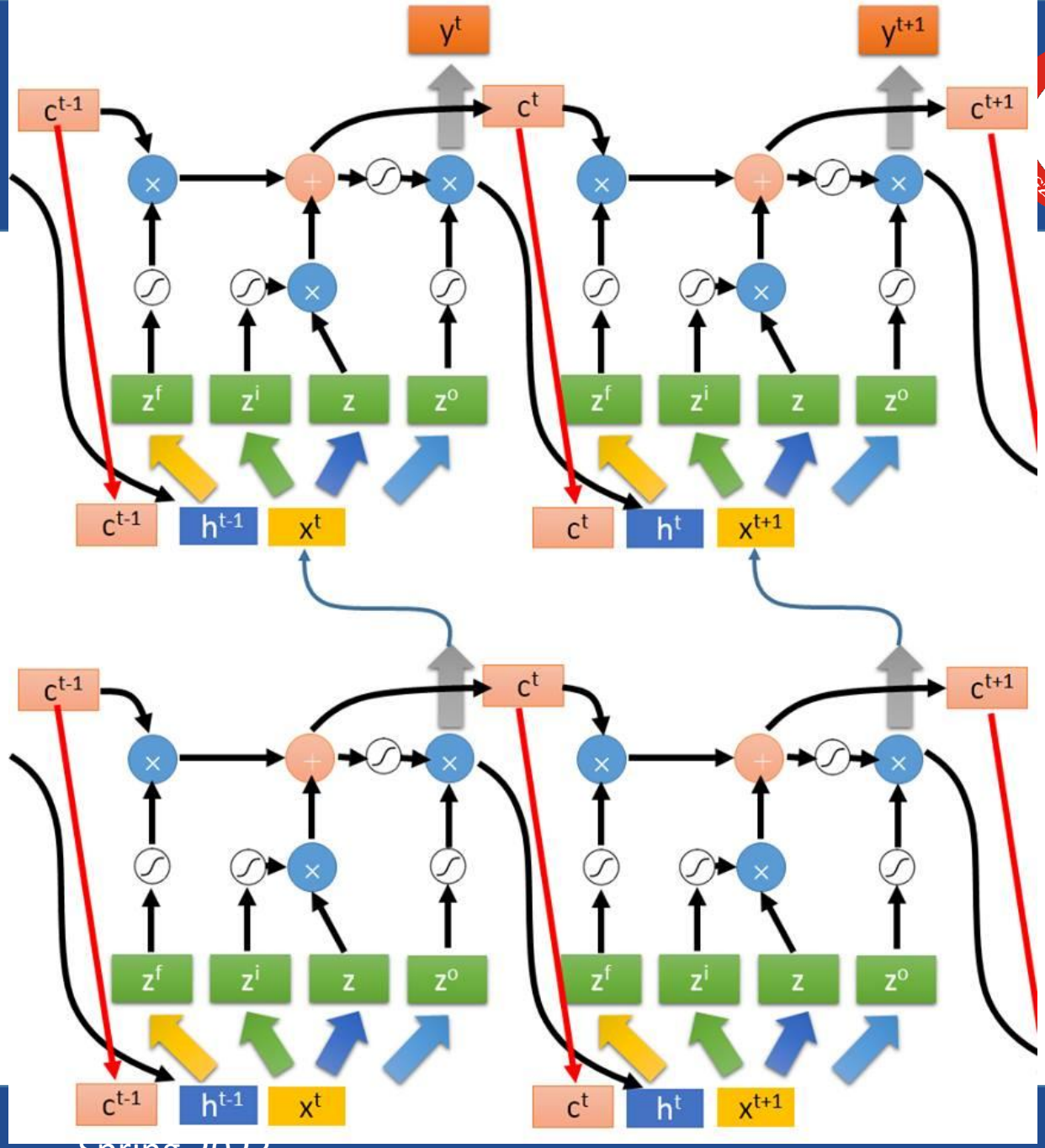


LSTM

Extension: “peephole”



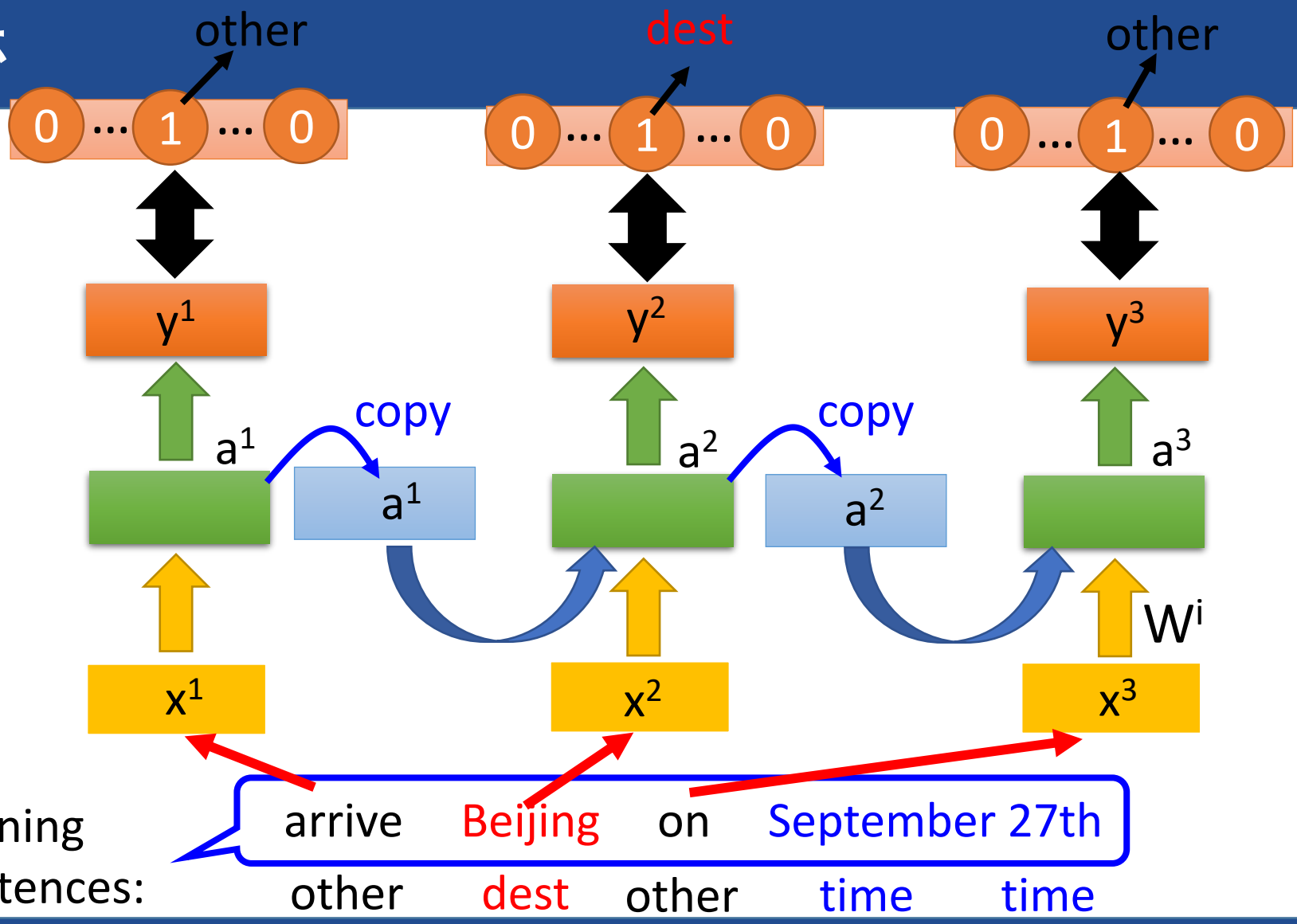
Multiple-layer LSTM



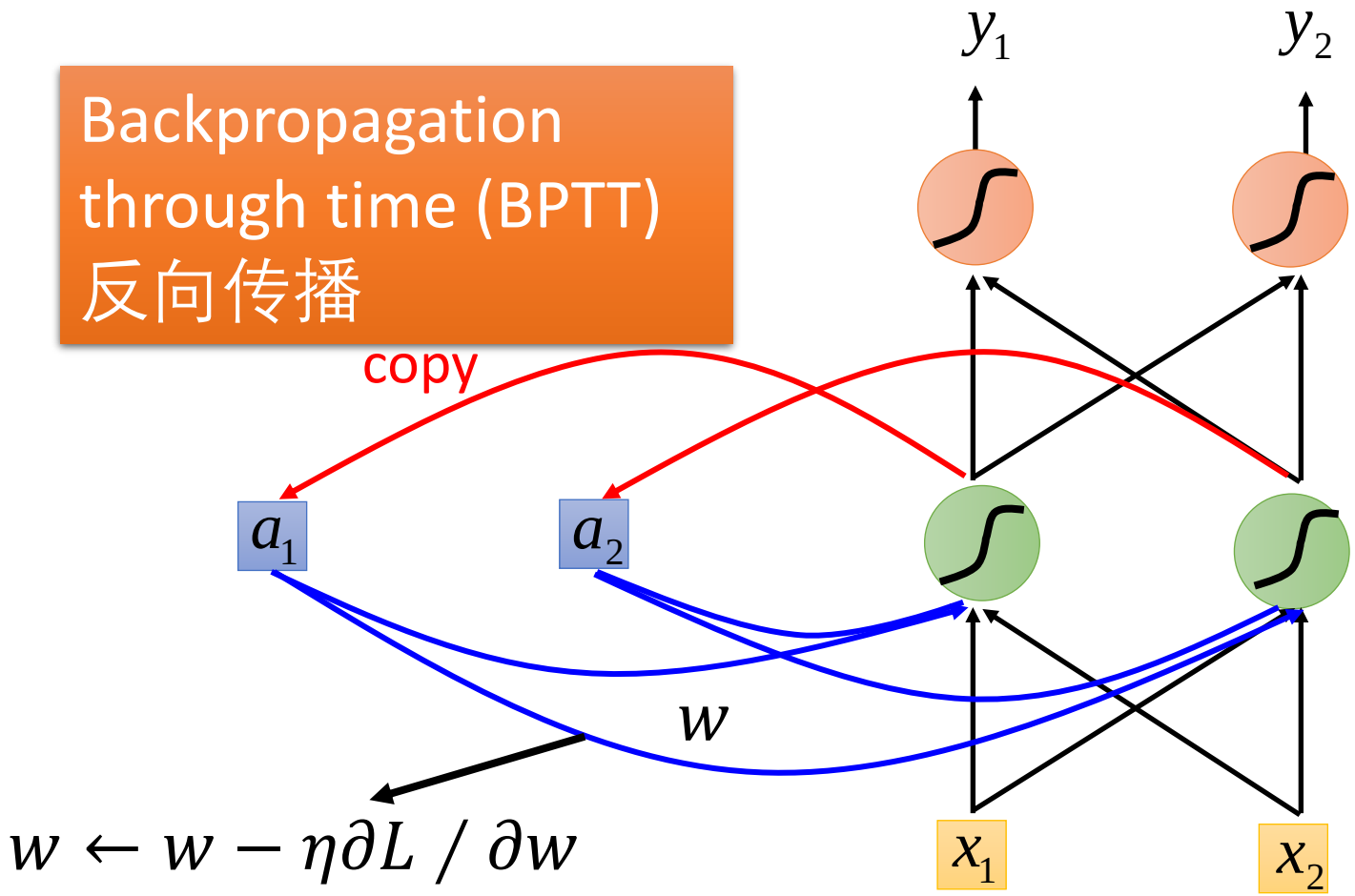
This is quite standard now.

Learning Target

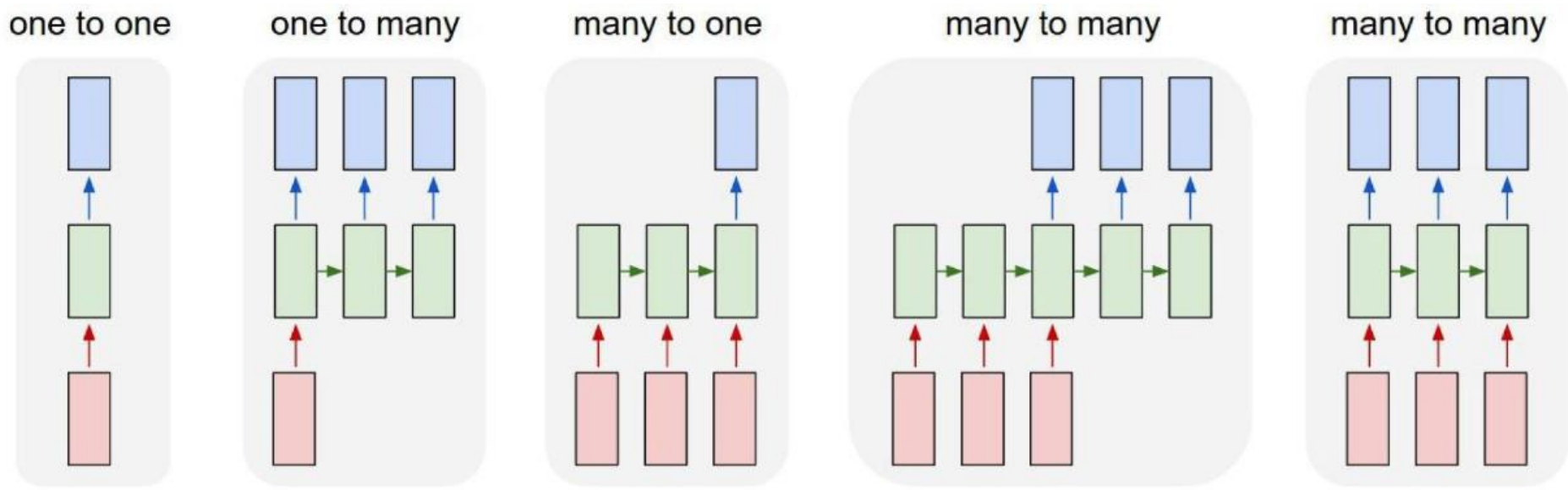
学习目标



Learning


Backpropagation through time (BPTT)
反向传播
copy
y_1
y_2
a_1
a_2
w
$w \leftarrow w - \eta \partial L / \partial w$
x_1
x_2

More Architecture



Vanilla Neural Network

Image Captioning

Sentiment Classification

Machine translation

video classification on frame level

You Like More?

https://www.youtube.com/watch?v=fLvJ8VdHLA0&t=77s

Q&A