Recurrent neural networks

Lecture 11

Changho Suh

January 25, 2024

Basic RNNs

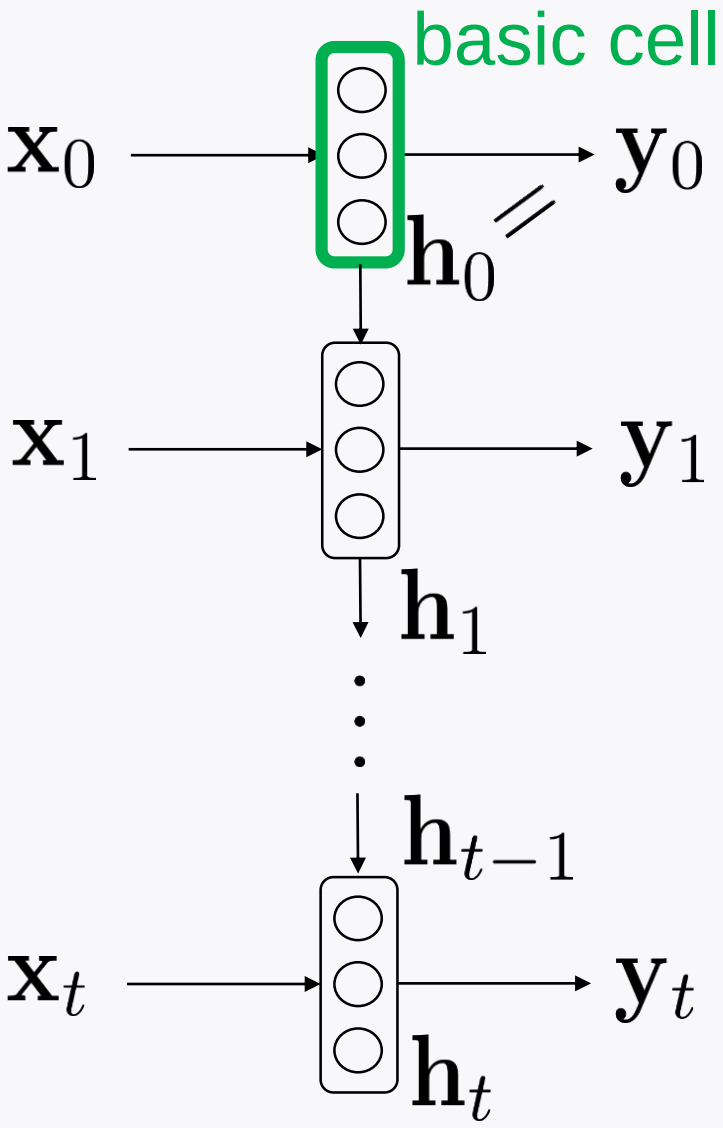
Outline

Will explore details on basic RNNs.

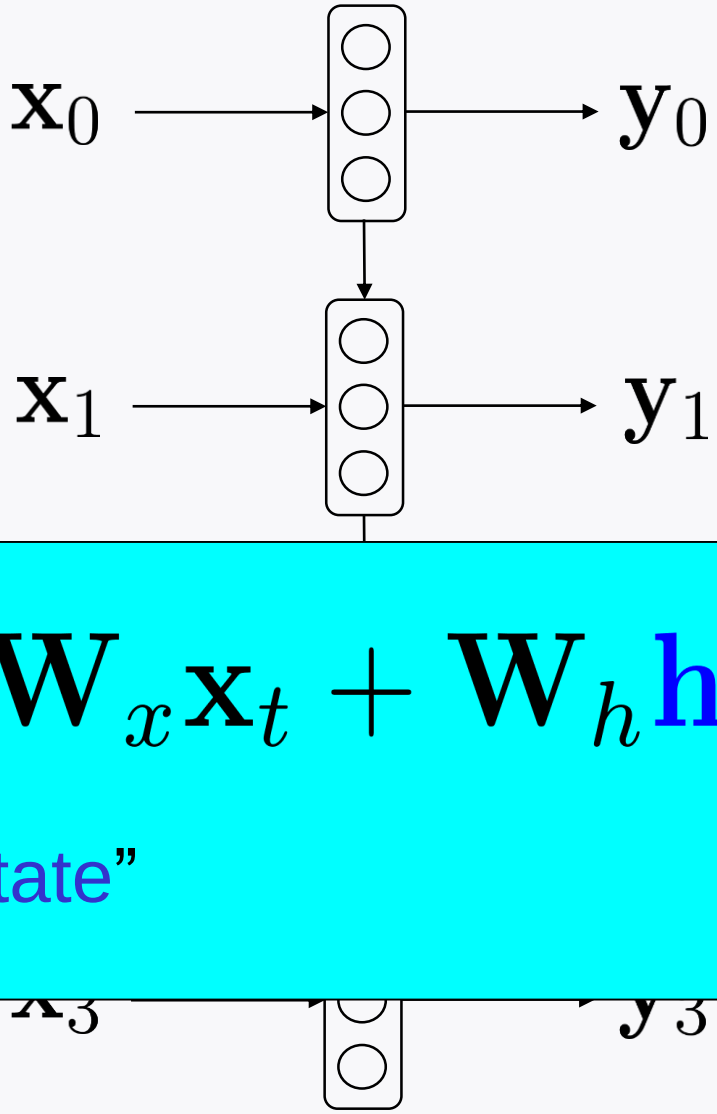
1. Study how to train basic RNNs;

2. Emphasize one challenge that basic RNNs face during training.

Recall: Basic RNNs



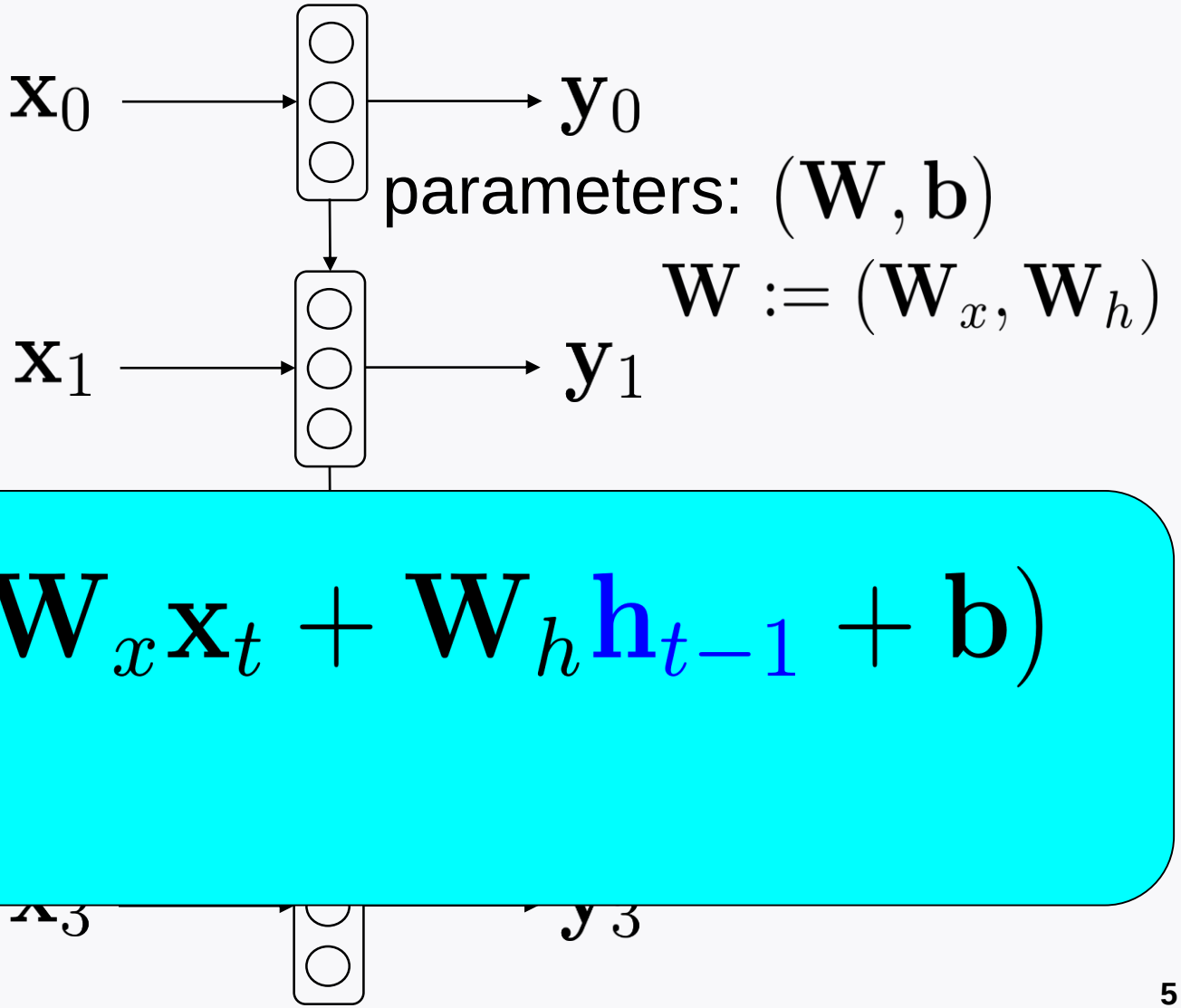
What to feed back?



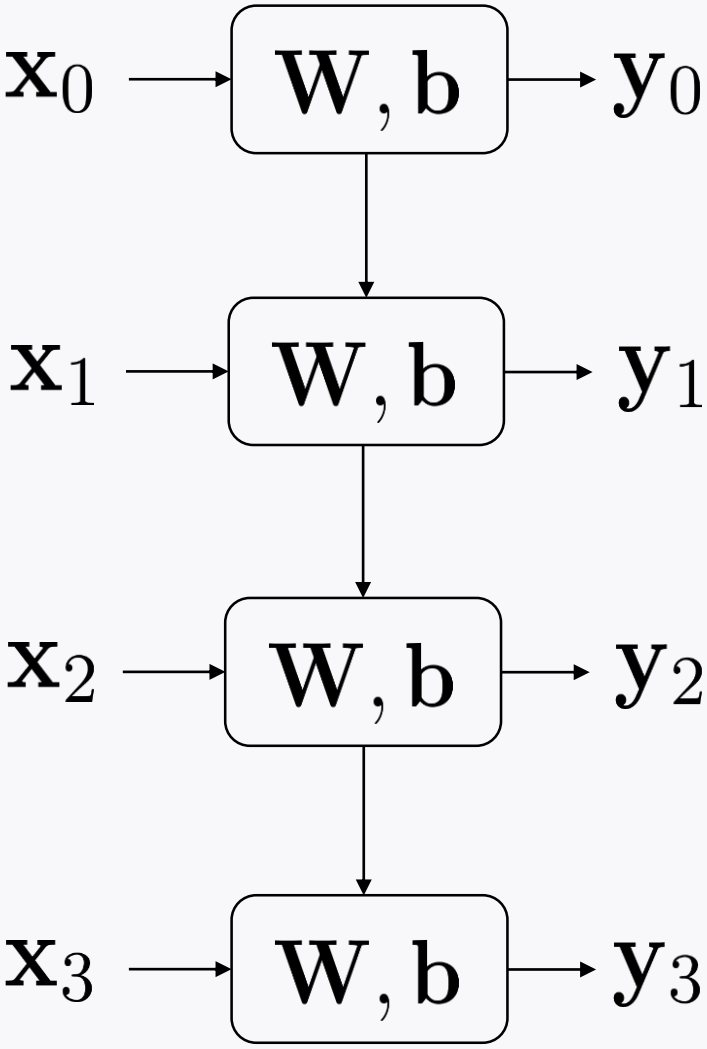
$$\mathbf{y}_t = \phi\left(\mathbf{W}_x\mathbf{x}_t + \mathbf{W}_h\mathbf{h}_{t-1} + \mathbf{b}\right)$$

Feed back "state"

Parameters to train?

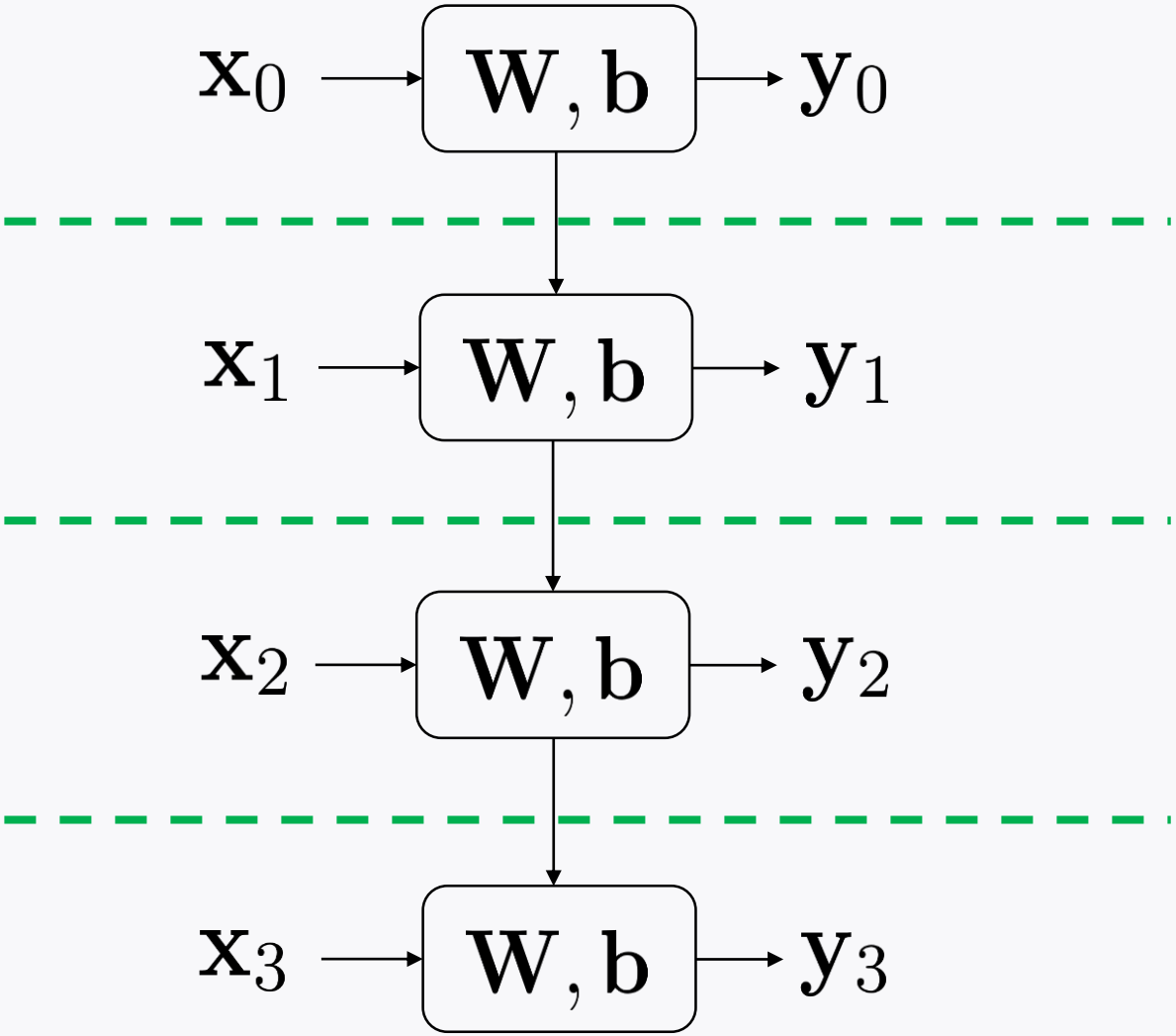


Parameters to train?

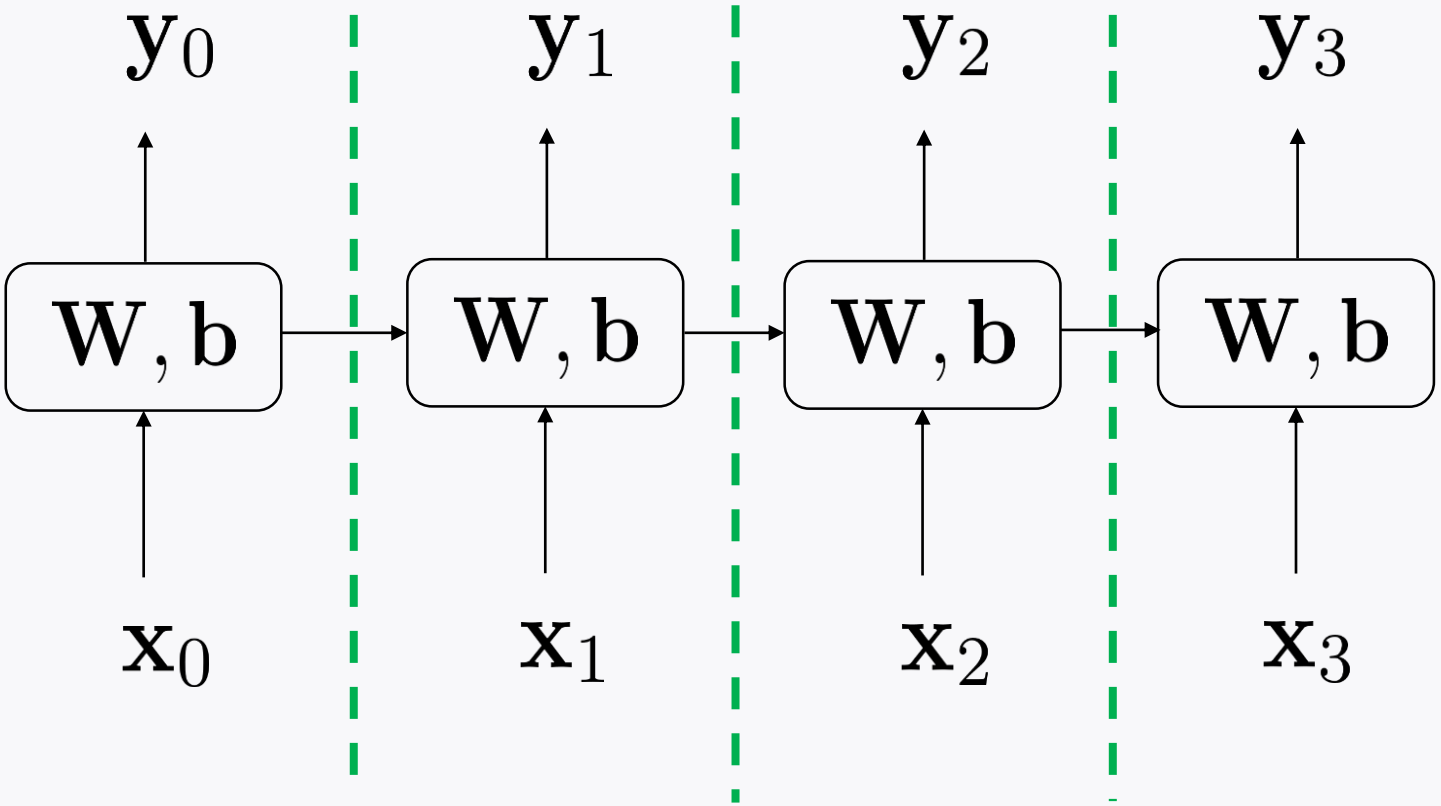


Why sharing the same parameters?

How to train?

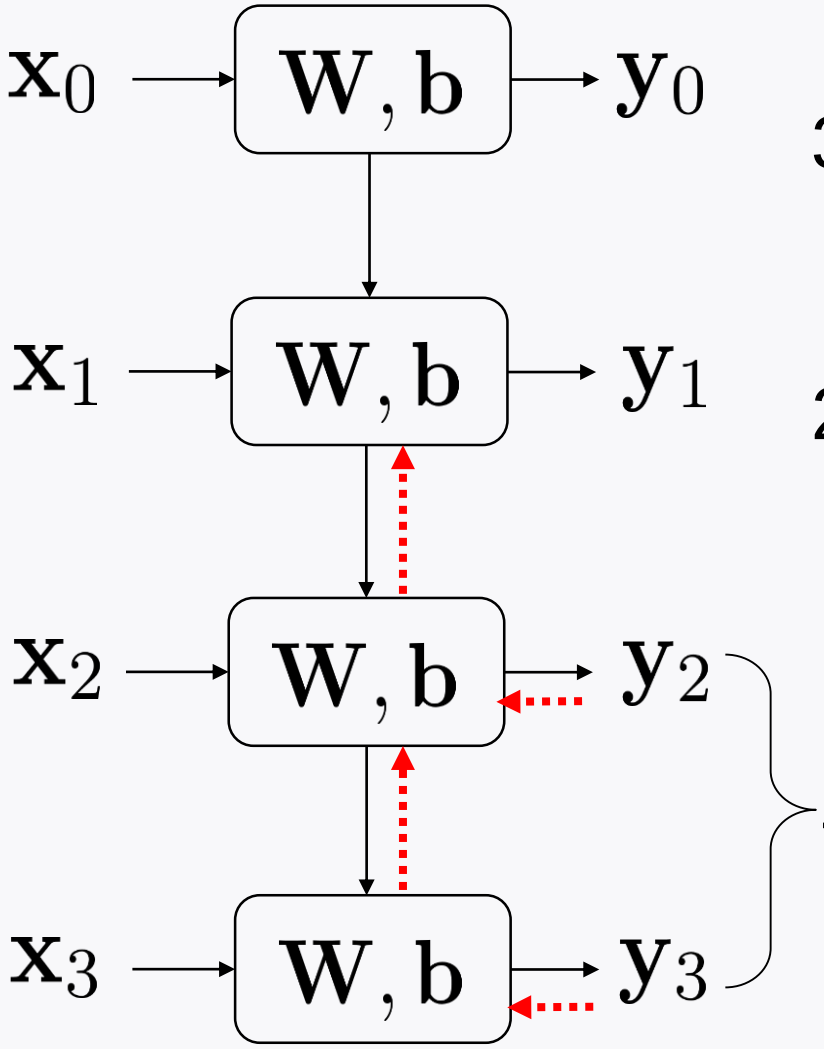


How to train?



Looks like a layered network!

Idea: BackProp Through Time



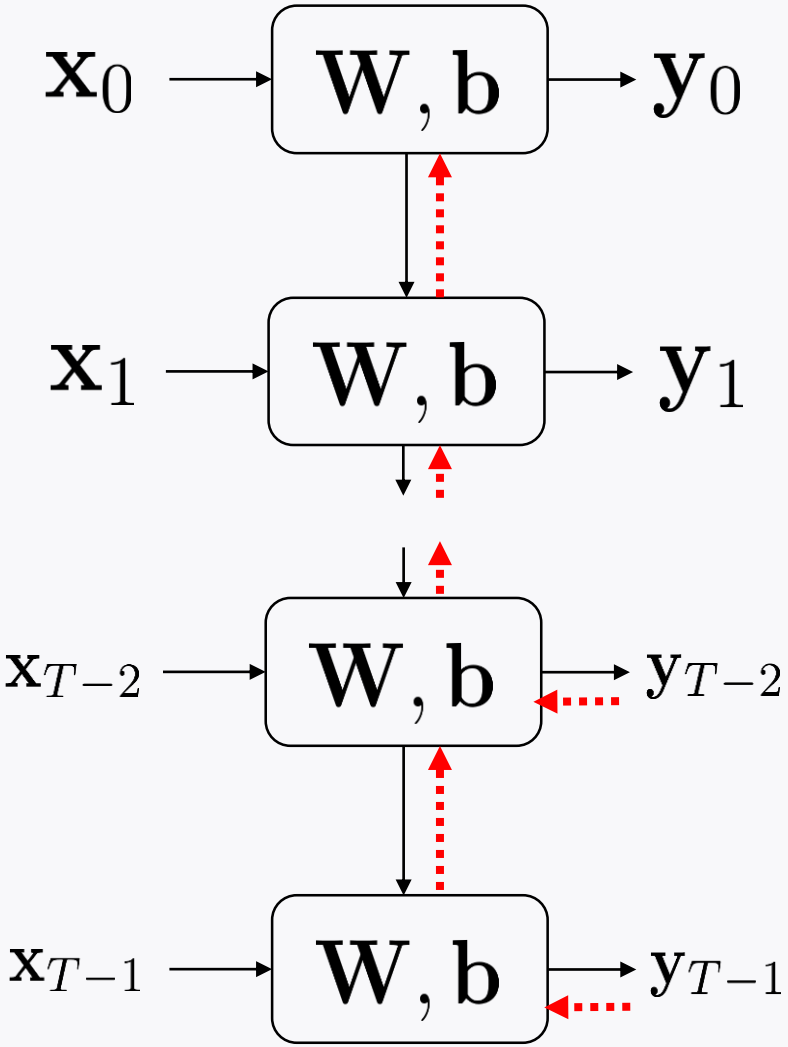
3. Backprop

2. Compute a cost function:

$$J(\mathbf{y}_2, \mathbf{y}_3)$$

1. Select a few last outputs.

A challenge



For large *T*, we often face:

Vanishing gradient problem!

The simplest and most common solution

Reduce *T*!

This technique is called: **Truncated BPTT**

Problem of the truncated BPTT

The model cannot learn long-term patterns.

If not well keeping memory, the states fade away quickly.

In an effort to address such problems, various types of cells have been introduced.

The most popular one is:

Long Short-Term Memory (LSTM) cell

Look ahead

Next lecture: Will study LSTM cells in detail.