Recurrent neural networks

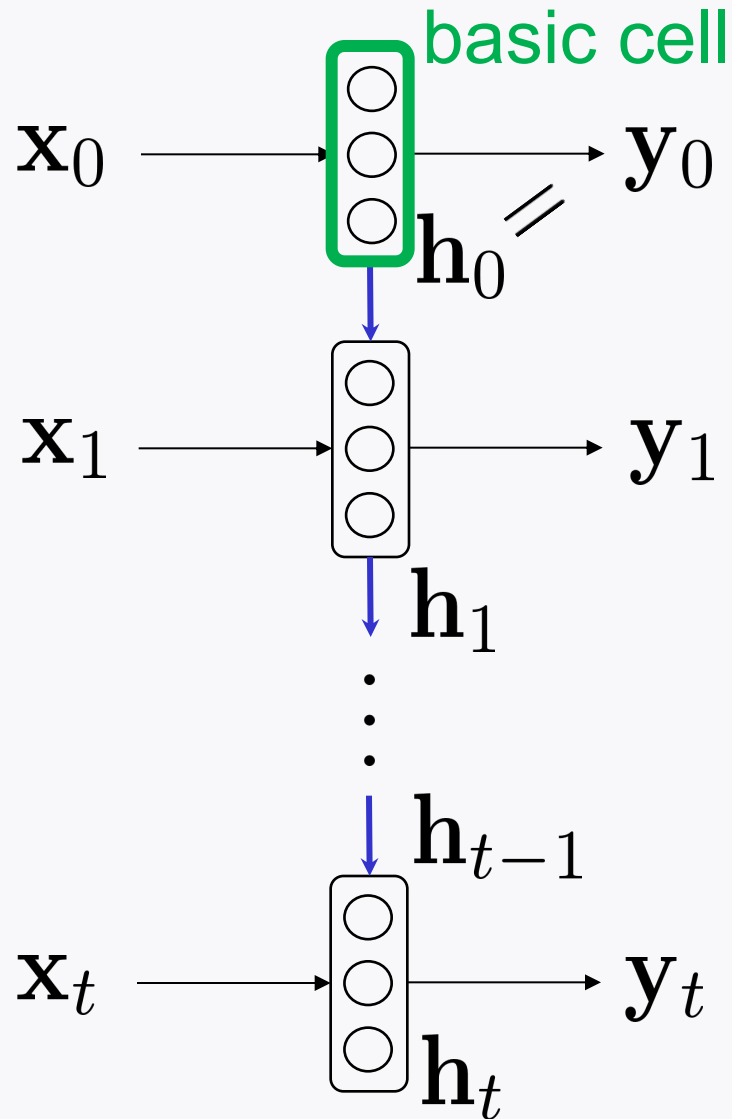# Lecture 11

Changho Suh

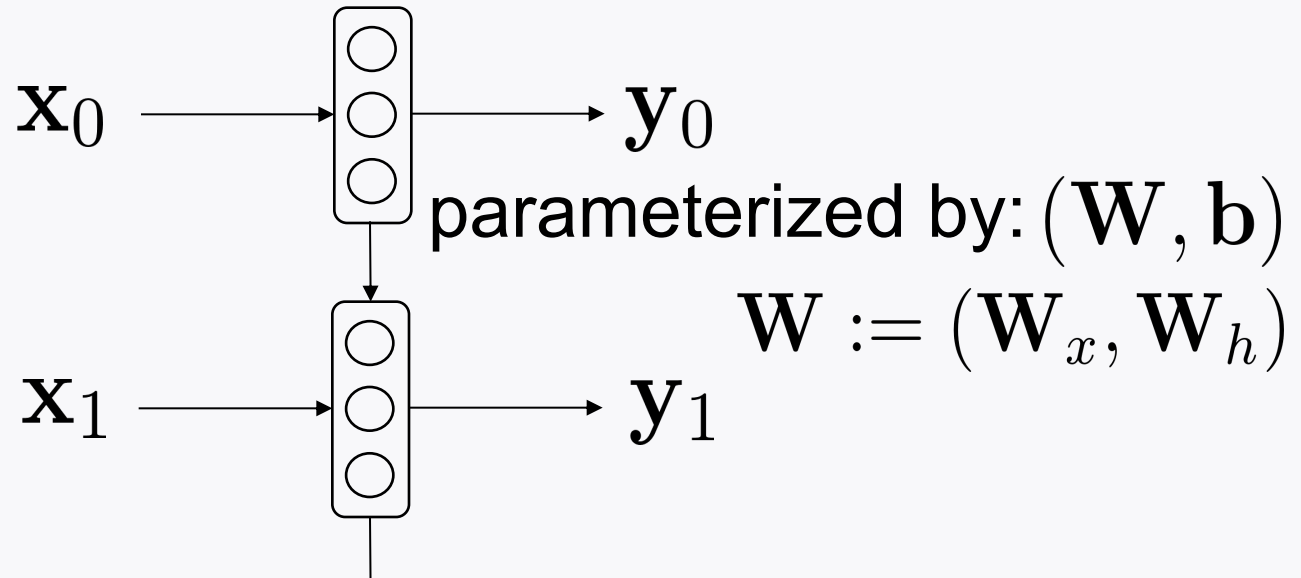September 30, 2021

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.
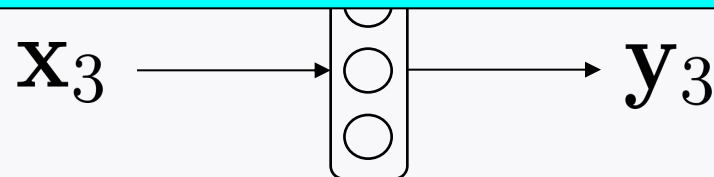
Recall: Basic RNNs

basic cell

$$\mathbf{x}_0 \rightarrow \mathbf{y}_0$$
$$\mathbf{h}_0 =$$
$$\mathbf{x}_1 \rightarrow \mathbf{y}_1$$
$$\mathbf{h}_1$$
$$\mathbf{h}_{t-1}$$
$$\mathbf{x}_t \rightarrow \mathbf{y}_t$$
$$\mathbf{h}_t$$

Unrolled version of basic RNNs

$$\mathbf{x}_0 \longrightarrow \boxed{\circ\circ\circ} \longrightarrow \mathbf{y}_0$$

parameterized by: $(\mathbf{W}, \mathbf{b})$

$$\mathbf{W} := (\mathbf{W}_x, \mathbf{W}_h)$$

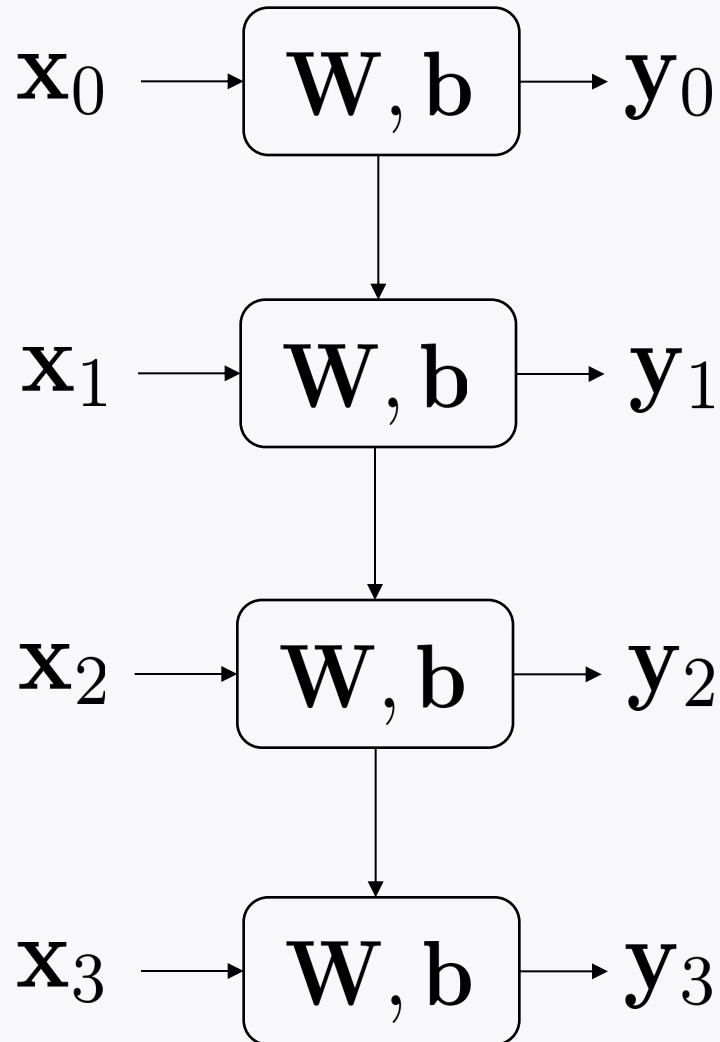$$\mathbf{x}_1 \longrightarrow \boxed{\circ\circ\circ} \longrightarrow \mathbf{y}_1$$

Recall:

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

$$\mathbf{x}_3 \longrightarrow \boxed{\circ\circ\circ} \longrightarrow \mathbf{y}_3$$

Unrolled version of basic RNNs

$x_0 \longrightarrow \boxed{W, b} \longrightarrow y_0$

$x_1 \longrightarrow \boxed{W, b} \longrightarrow y_1$

$x_2 \longrightarrow \boxed{W, b} \longrightarrow y_2$

$x_3 \longrightarrow \boxed{W, b} \longrightarrow y_3$

Why sharing the same parameters?

How to train?
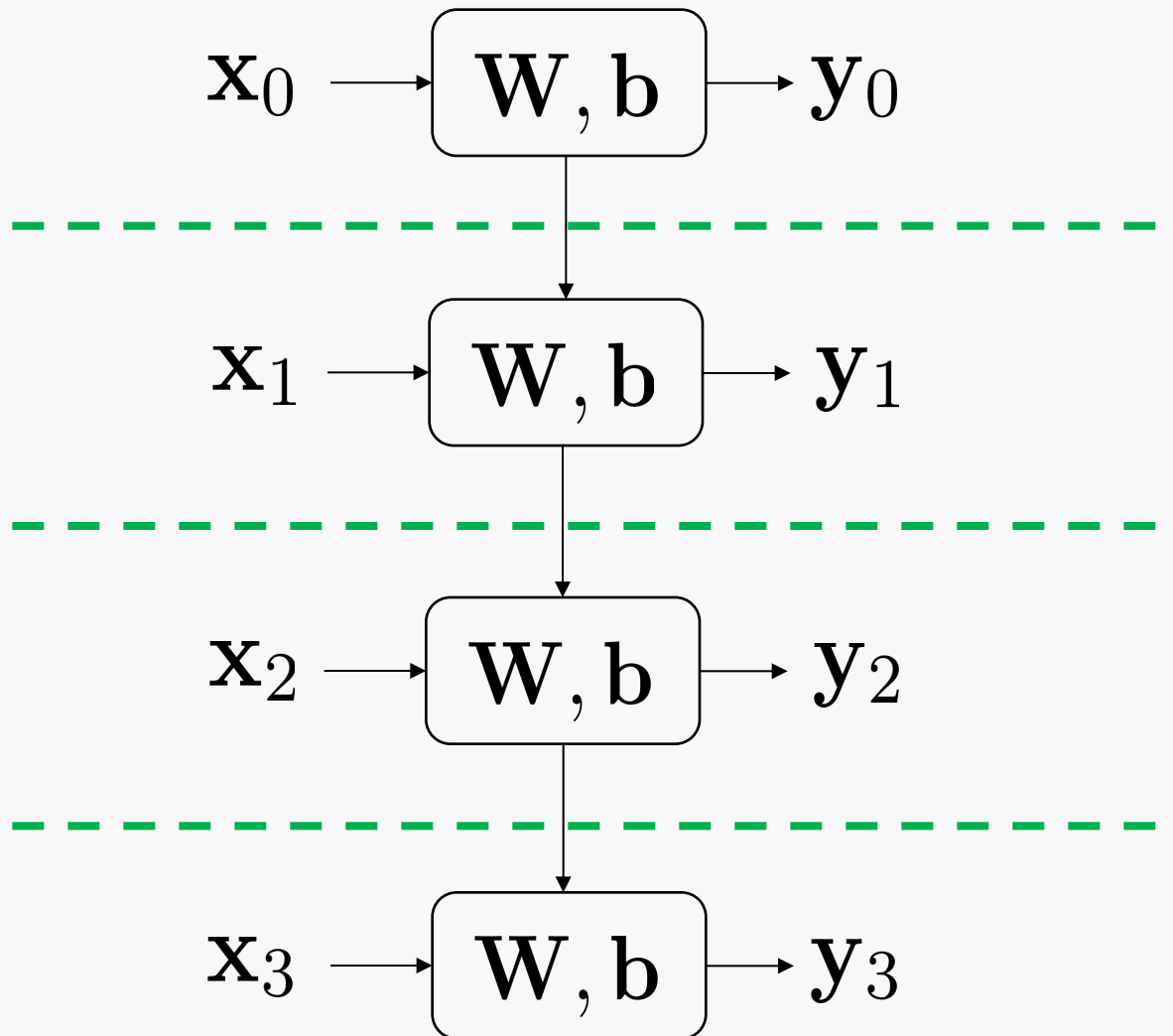
Idea: **BackProp Through Time**

$$\mathbf{x}_0 \longrightarrow \boxed{\mathbf{W}, \mathbf{b}} \longrightarrow \mathbf{y}_0$$

$$\mathbf{x}_1 \longrightarrow \boxed{\mathbf{W}, \mathbf{b}} \longrightarrow \mathbf{y}_1$$

$$\mathbf{x}_2 \longrightarrow \boxed{\mathbf{W}, \mathbf{b}} \longrightarrow \mathbf{y}_2$$

$$\mathbf{x}_3 \longrightarrow \boxed{\mathbf{W}, \mathbf{b}} \longrightarrow \mathbf{y}_3$$

3. Backprop

2. Compute a cost function:

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

1. Select a few last outputs.

7

A challenge

$\mathbf{x}_0 \longrightarrow \boxed{\mathbf{W}, \mathbf{b}} \longrightarrow \mathbf{y}_0$

$\mathbf{x}_1 \longrightarrow \boxed{\mathbf{W}, \mathbf{b}} \longrightarrow \mathbf{y}_1$

$\mathbf{x}_{T-2} \longrightarrow \boxed{\mathbf{W}, \mathbf{b}} \longrightarrow \mathbf{y}_{T-2}$

$\mathbf{x}_{T-1} \longrightarrow \boxed{\mathbf{W}, \mathbf{b}} \longrightarrow \mathbf{y}_{T-1}$
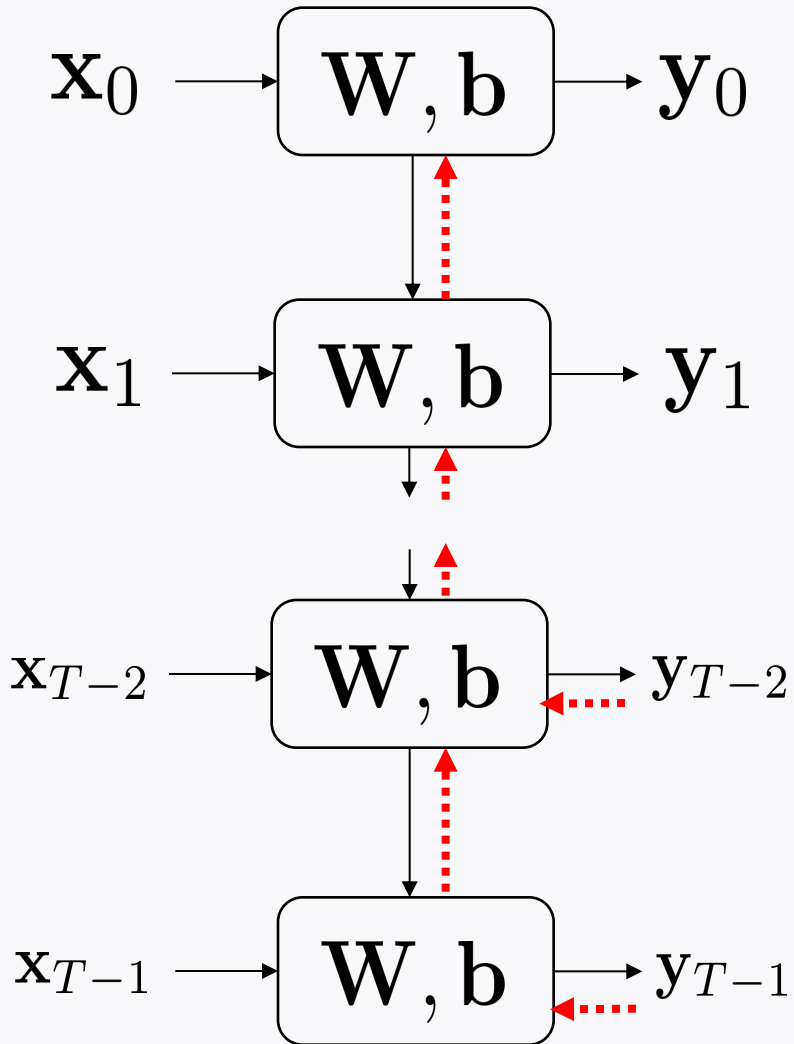
For large *T*, we often face:

Vanishing gradient problem!

8

The simplest and most common solution

Reduce *T*!

This is called: **Truncated BPTT**

Problem of the truncated BPTT

The model cannot learn long-term patterns.

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

As 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 details.