Recurrent Neural Networks

CS 6956: Deep Learning for NLP

Overview

1. Modeling sequences
2. Recurrent neural networks: An abstraction
3. Usage patterns for RNNs
4. BiDirectional RNNs
5. A concrete example: The Elman RNN
6. The vanishing gradient problem
7. Long short-term memory units

Overview

1. Modeling sequences
2. Recurrent neural networks: An abstraction
3. Usage patterns for RNNs
4. BiDirectional RNNs
5. A concrete example: The Elman RNN
6. The vanishing gradient problem
7. Long short-term memory units

A simple RNN

- What we saw so far is just a template for a recurrent neural network
  - Did not specify what the functions inside it are

- Let's look at a simple instantiation, first introduced by Elman 1990

A simple RNN

At each step, an RNN:

- Computes the next cell state: $\mathbf{s}_t = \mathrm{R}(\mathbf{s}_{t-1}, \mathbf{x}_t)$
- Computes the output: $\boldsymbol{y}_t = \mathrm{O}(\mathbf{s}_t)$

Need to specify two functions:

1. How to generate the current state using the previous state and the current input?

2. How to generate the current output using the current state?

A simple RNN

At each step, an RNN:

- Computes the next cell state: $\mathbf{s}_t = \mathrm{R}(\mathbf{s}_{t-1}, \mathbf{x}_t)$
- Computes the output: $\boldsymbol{y}_t = \mathrm{O}(\mathbf{s}_t)$

Need to specify two functions:

1. How to generate the current state using the previous state and the current input?

2. How to generate the current output using the current state?

   The output is the state. That is, $\boldsymbol{y}_t = \mathbf{s}_t$

Computing the value of a state

1. How to generate the current state using the previous state and the current input?

$\mathbf{s}_{t-1}$

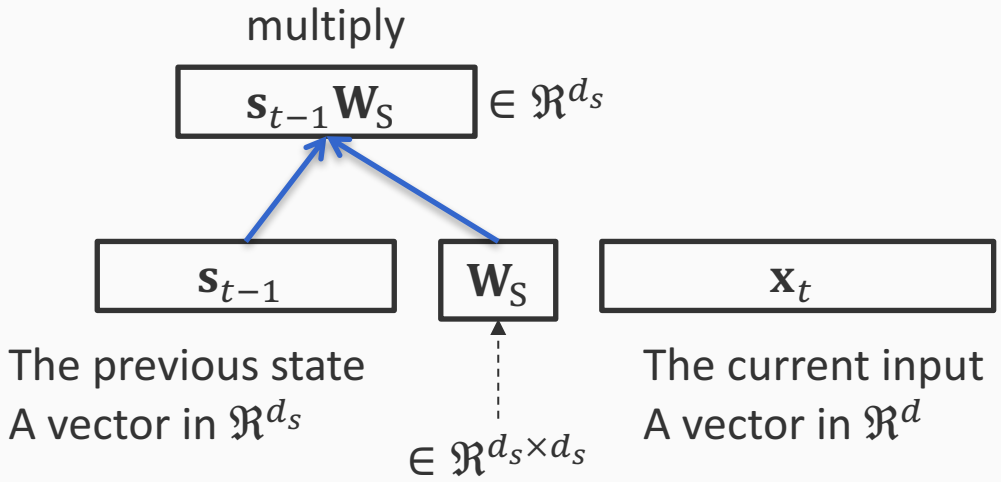
The previous state
A vector in $\Re^{d_s}$

$\mathbf{x}_t$

The current input
A vector in $\Re^{d}$

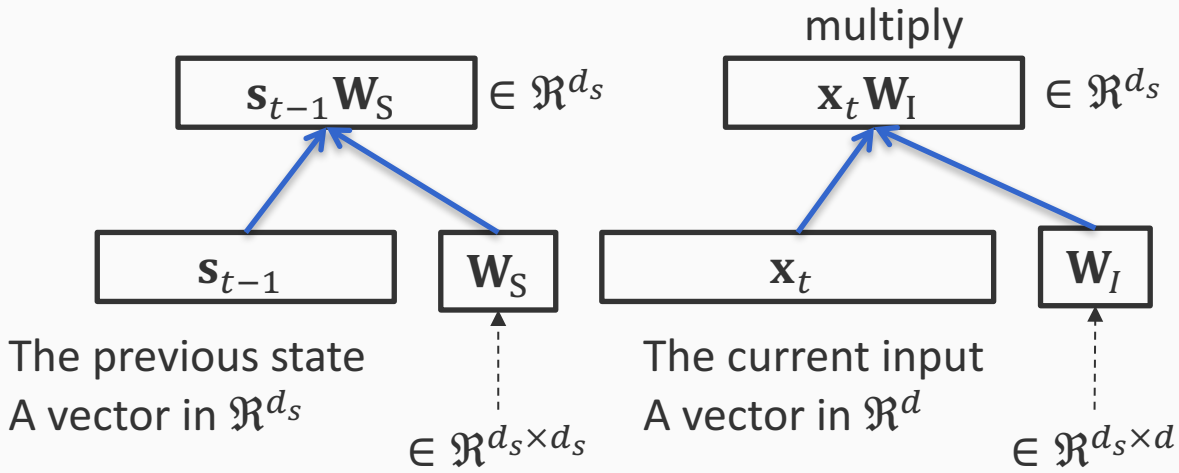
Computing the value of a state

1. How to generate the current state using the previous state and the current input?



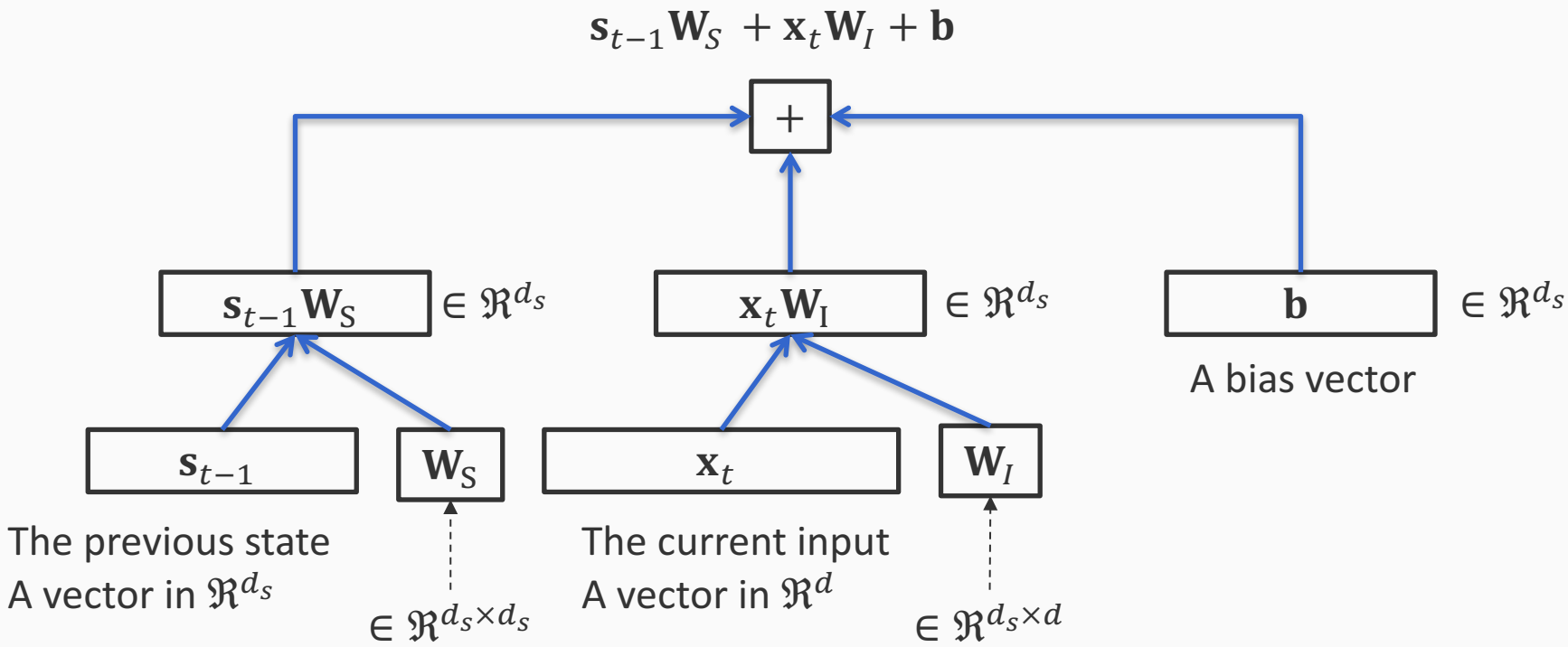
Computing the value of a state

1. How to generate the current state using the previous state and the current input?



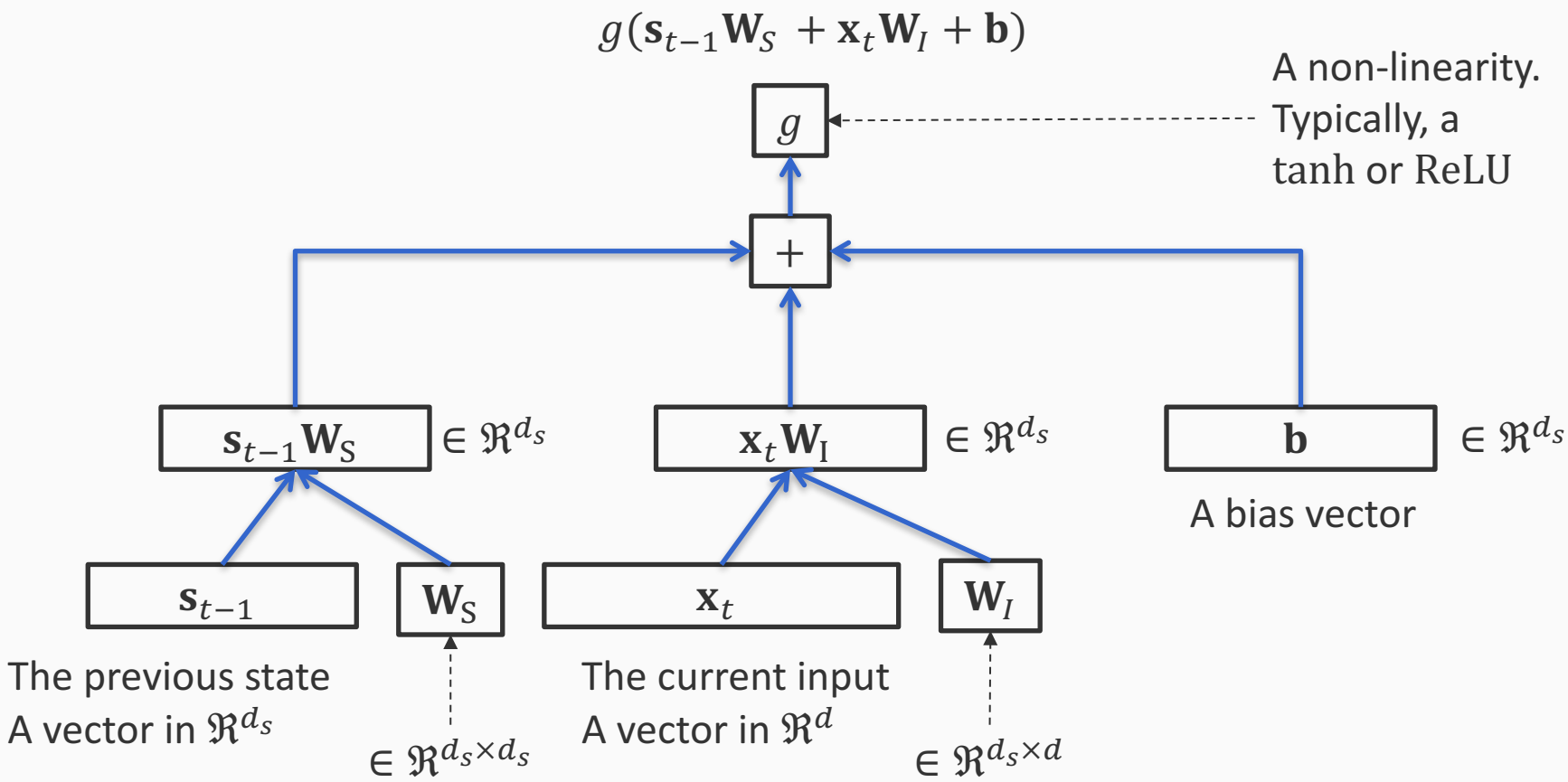
Computing the value of a state

1. How to generate the current state using the previous state and the current input?



Computing the value of a state

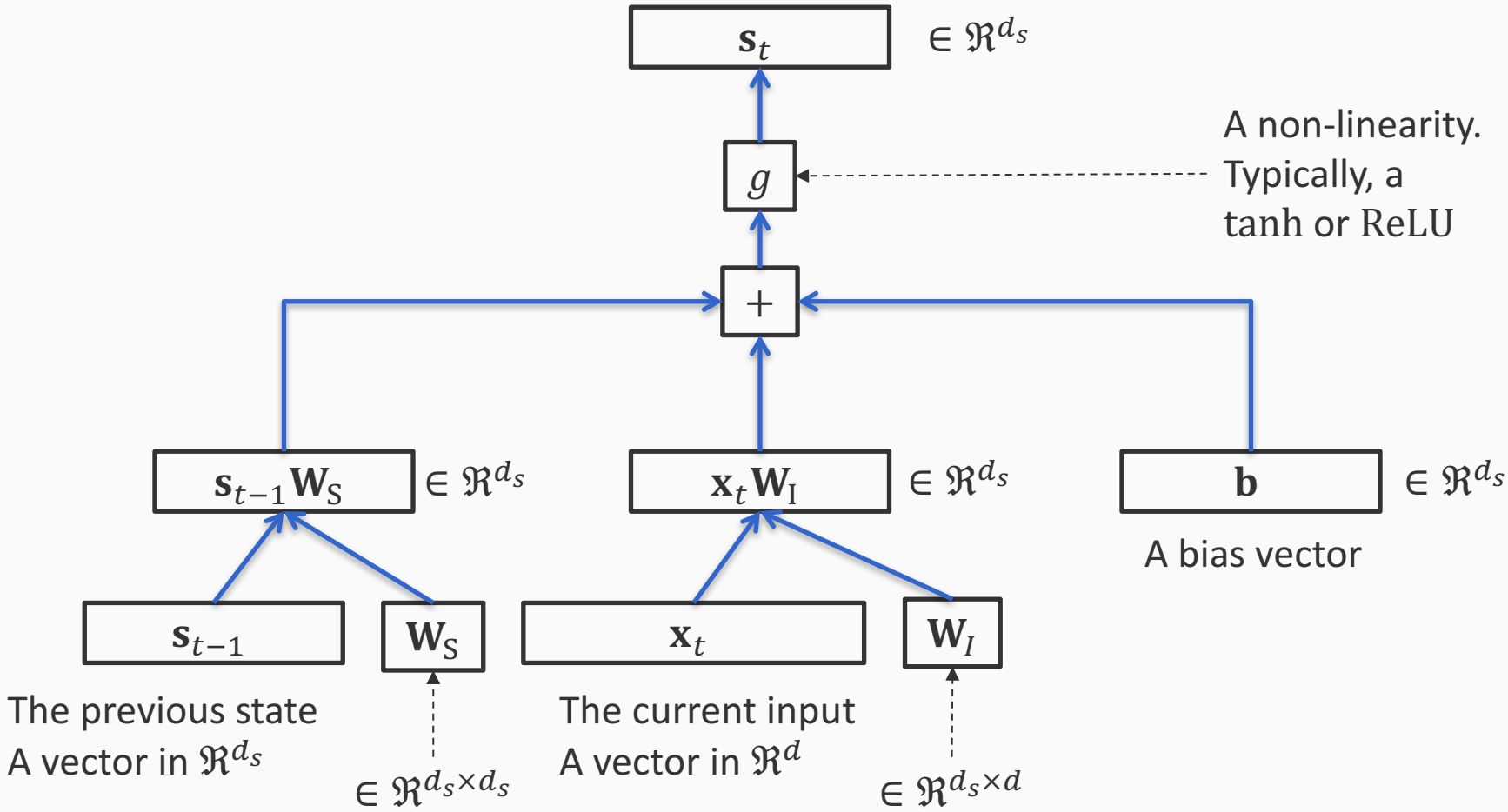
1. How to generate the current state using the previous state and the current input?



Computing the value of a state

1. How to generate the current state using the previous state and the current input?

Next state $\mathbf{s}_t = g(\mathbf{s}_{t-1}\mathbf{W}_S + \mathbf{x}_t\mathbf{W}_I + \mathbf{b})$



A simple RNN

At each step, an RNN:

- Computes the next cell state: $\mathbf{s}_t = \mathrm{R}(\mathbf{s}_{t-1}, \mathbf{x}_t)$
- Computes the output: $\boldsymbol{y}_t = \mathrm{O}(\mathbf{s}_t)$

Need to specify two functions:

1. How to generate the current state using the previous state and the current input?

2. How to generate the current output using the current state?

   The output is the state. That is, $\boldsymbol{y}_t = \mathbf{s}_t$

A simple RNN

At each step, an RNN:

- Computes the next cell state: $\mathbf{s}_t = \mathrm{R}(\mathbf{s}_{t-1}, \mathbf{x}_t)$
- Computes the output: $\boldsymbol{y}_t = \mathrm{O}(\mathbf{s}_t)$

Need to specify two functions:

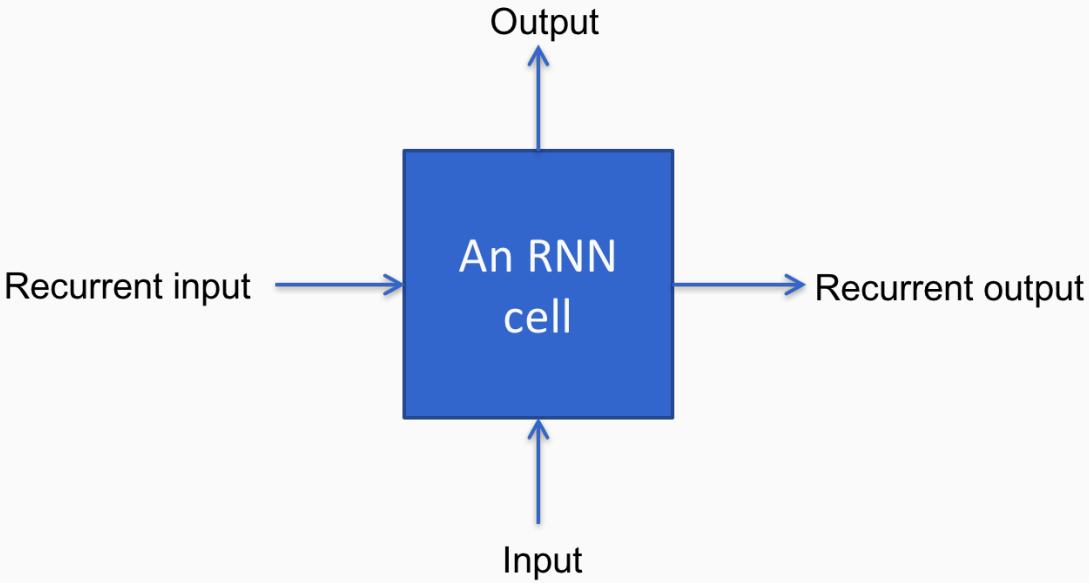
1. How to generate the current state using the previous state and the current input?

   Next state $\mathbf{s}_t = g(\mathbf{s}_{t-1}\mathbf{W}_S + \mathbf{x}_t\mathbf{W}_I + \mathbf{b})$

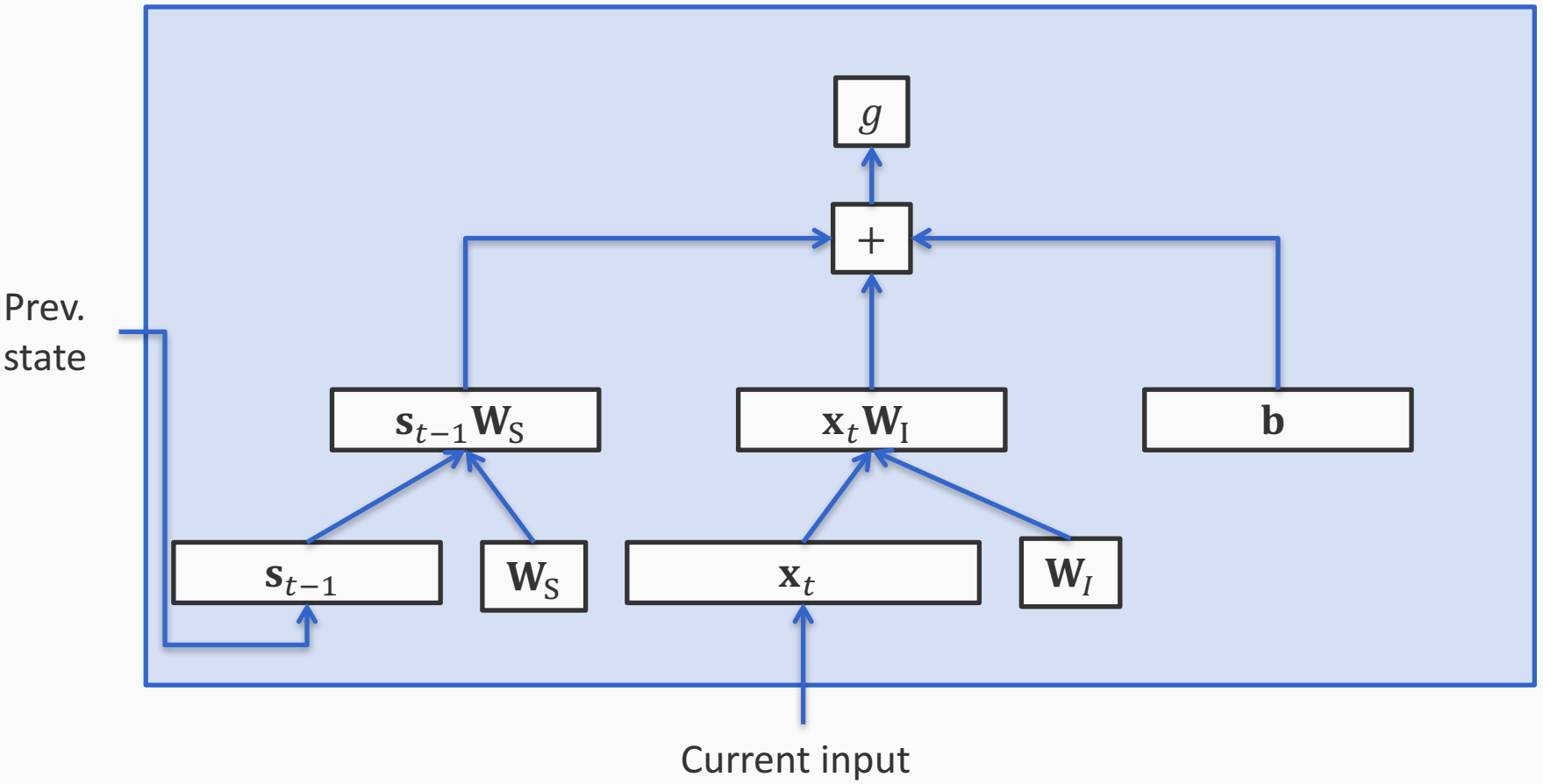
2. How to generate the current output using the current state?

   The output is the state. That is, $\boldsymbol{y}_t = \mathbf{s}_t$

The Elman RNN



The Elman RNN



The Elman RNN