Q1-1: Are these statements true or false?
(A) Order matters in sequential data.
(B) A batch of sequential data always contains sequences of a same length.

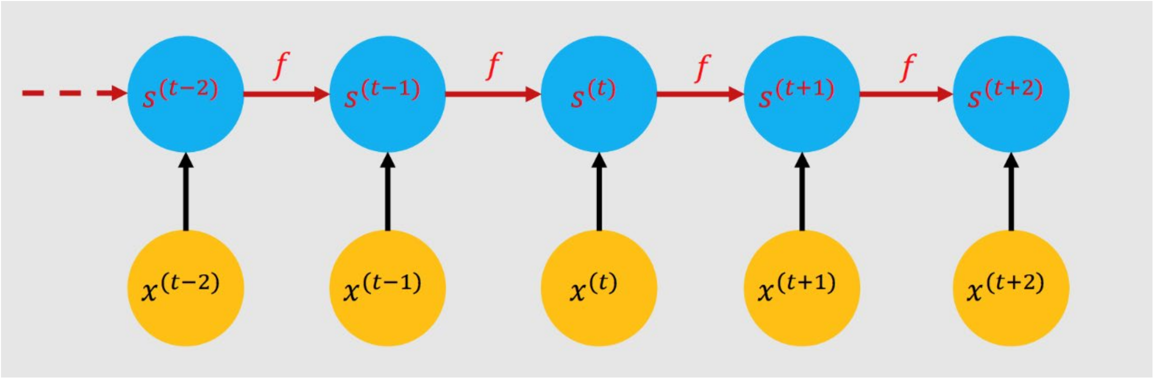
1. True, True
2. True, False
3. False, True
4. False, False

Q1-1: Are these statements true or false?
(A) Order matters in sequential data.
(B) A batch of sequential data always contains sequences of a same length.

1. True, True
2. True, False
3. False, True
4. False, False

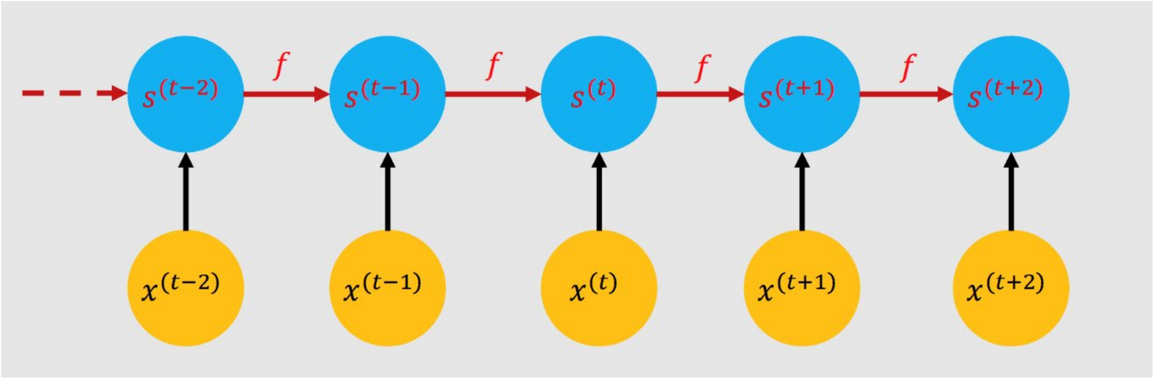
(A) As is shown by its name "sequential", order matters in sequential data.
(B) A batch of sequential data can have different length, such as different sentences.

Q1-2: Please choose the representation of $s^{(t+2)}$ in terms of $s^{(t)}, x^{(t)},\ x^{(t+1)}, x^{(t+2)}$ in the following dynamic system $s^{(t+1)} = f_\theta(s^{(t)}, x^{(t+1)})$.



1. $f_\theta(s^{(t)}, x^{(t+1)})$
2. $f_\theta(s^{(t)}, x^{(t+2)})$
3. $f_\theta(f_\theta(s^{(t)}, x^{(t)}), x^{(t+1)})$
4. $f_\theta(f_\theta(s^{(t)}, x^{(t+1)}), x^{(t+2)})$

Q1-2: Please choose the representation of $s^{(t+2)}$ in terms of $s^{(t)}, x^{(t)}, \ x^{(t+1)}, x^{(t+2)}$ in the following dynamic system $s^{(t+1)} = f_\theta(s^{(t)}, x^{(t+1)})$.



1. $f_\theta(s^{(t)}, x^{(t+1)})$
2. $f_\theta(s^{(t)}, x^{(t+2)})$
3. $f_\theta(f_\theta(s^{(t)}, x^{(t)}), x^{(t+1)})$
4. $f_\theta(f_\theta(s^{(t)}, x^{(t+1)}), x^{(t+2)})$ ⬅

As is shown in this dynamic system, we have $s^{(t+2)} = f_\theta\big(s^{(t+1)}, x^{(t+2)}\big) = f_\theta\big(f_\theta\big(s^{(t)}, x^{(t+1)}\big), x^{(t+2)}\big)$, as $s^{(t+1)} = f_\theta\big(s^{(t)}, x^{(t+1)}\big)$.

Q2-1: Are these statements true or false?
(A) Shared functions and parameters in RNN can help improve the generalization ability of the model.
(B) Hidden states $s^{(t)}$ in RNN contain only previous information before $x^{(t)}$ but not about $x^{(t)}$ in sequential data.

1. True, True
2. True, False
3. False, True
4. False, False

Q2-1: Are these statements true or false?
(A) Shared functions and parameters in RNN can help improve the generalization ability of the model.
(B) Hidden states $s^{(t)}$ in RNN contain only previous information before $x^{(t)}$ but not about $x^{(t)}$ in sequential data.

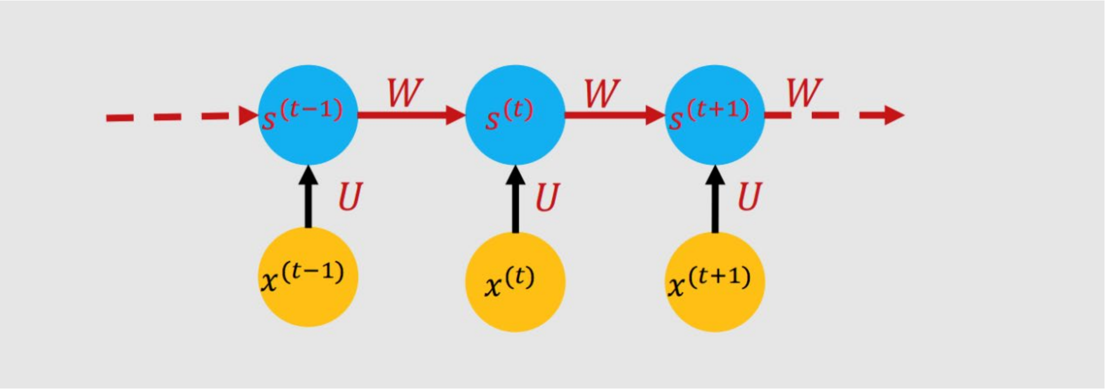
1. True, True
2. True, False ⟵
3. False, True
4. False, False

(A) As is shown in the lecture.
(B) By the formula $s^{(t)} = f_\theta\left(s^{(t-1)}, x^{(t)}\right)$, we can see that $s^{(t)}$ contains the information about $x^{(t)}$ in the sequential data.

Q2-2: Are these statements true or false?

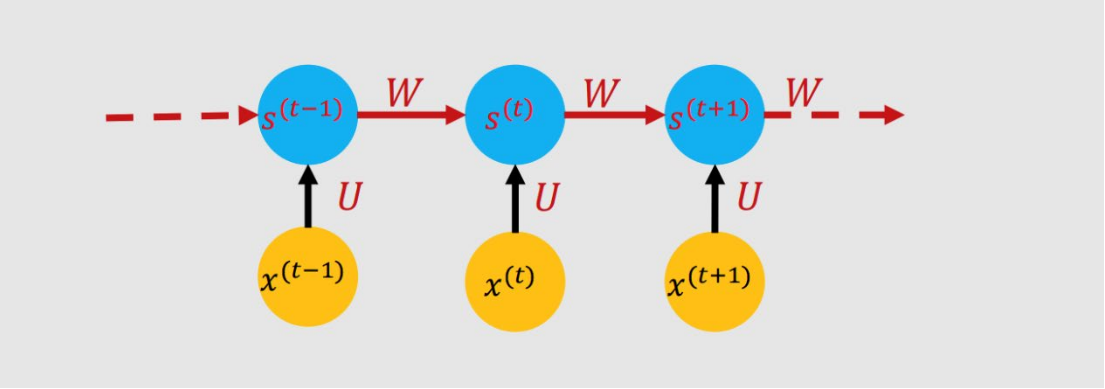
(A) The hidden state $s^{(t)}$ is the linear combination of the previous hidden state $s^{(t-1)}$ and the external data $x^{(t)}$.

(B) Sharing functions and parameters in RNN leads to inherent limitation on the learning ability of the model.



1. True, True
2. True, False
3. False, True
4. False, False

Q2-2: Are these statements true or false?
(A) The hidden state $s^{(t)}$ is the linear combination of the previous hidden state $s^{(t-1)}$ and the external data $x^{(t)}$.
(B) Sharing functions and parameters in RNN leads to inherent limitation on the learning ability of the model.

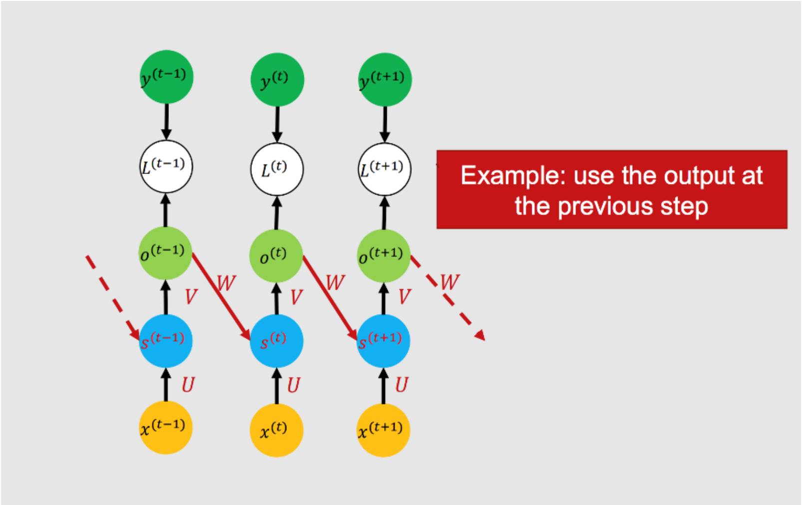


1. True, True
2. True, False
3. False, True
4. False, False

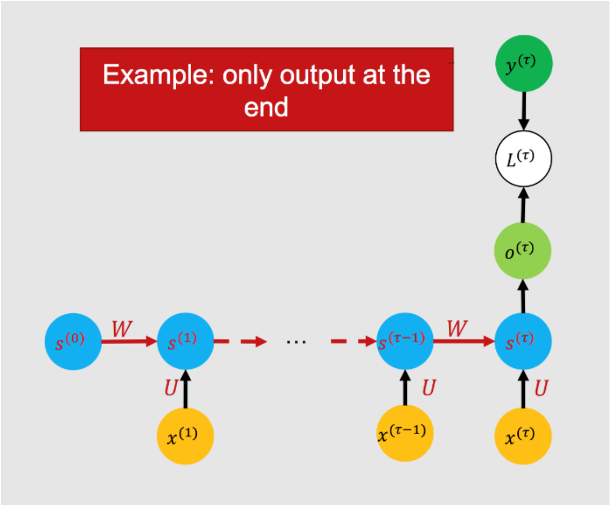
(A) We need to use an activation function to compute the hidden states, so it's not linear.
(B) As is shown in the lecture, such RNN of a finite size can be universal.

Q3-1: Are these statements for the RNN variants true or false?
(A) These RNNs are usually less powerful than those with hidden-to-hidden recurrent connection.
(B) These RNNs can be used to produce a fixed-length representation of the sequence.

1. True, True
2. True, False
3. False, True
4. False, False



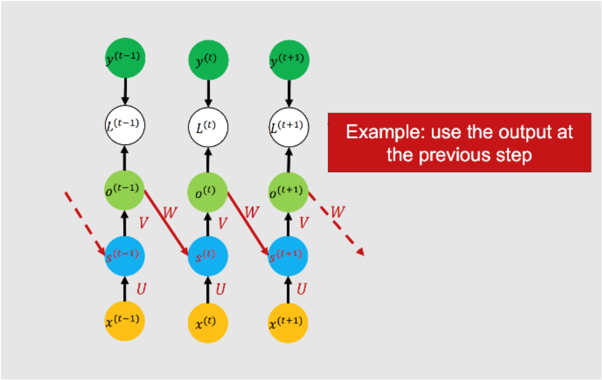
(A)



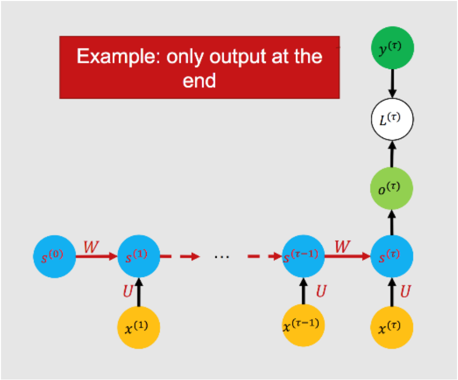
(B)

Q3-1: Are these statements for the RNN variants true or false?
(A) These RNNs are usually less powerful than those with hidden-to-hidden recurrent connection.
(B) These RNNs can be used to produce a fixed-length representation of the sequence.

1. True, True
2. True, False
3. False, True
4. False, False



(A)

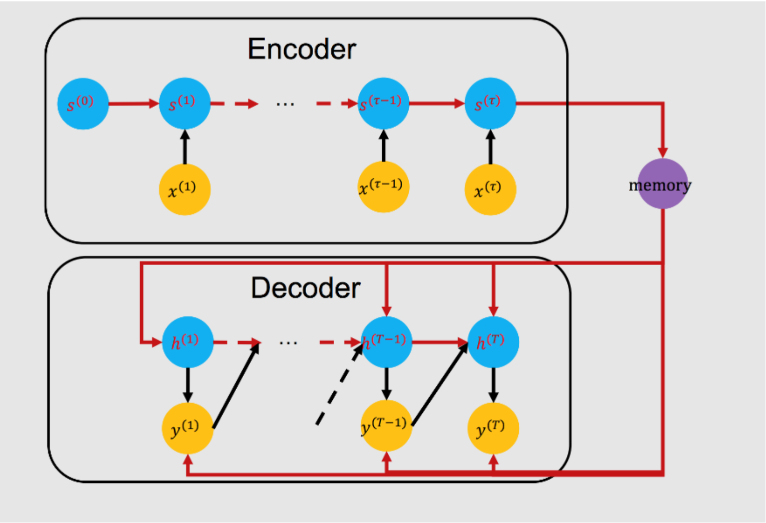


(B)

(A) If we have hidden-to-hidden recurrent connection, RNN can choose to put any information it wants about the past into current hidden state and transmit it to the future. However, RNN here is trained to put a specific output value into o, and o is the only information allowed to send to the future. So intuitively speaking, unless o is very high-dimensional and rich, RNN here will lack some important information from the past.
(B) It's true because the last state can carry all information in the sequence.

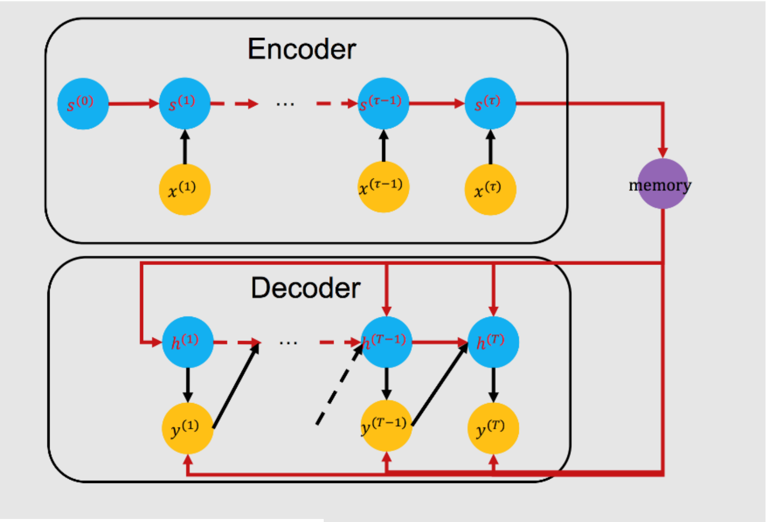
Q3-2: Are these statements for Encoder-Decoder RNNs true or false?
(A) The last hidden state $s^{(\tau)}$ of the encoder RNN is typically used as a representation of the input sequence and is provided as input to the decoder RNN.
(B) If the last hidden state $s^{(\tau)}$ of the encoder RNN has too small a dimension, it may not be able to represent a long input sequence.

1. True, True
2. True, False
3. False, True
4. False, False



Q3-2: Are these statements for Encoder-Decoder RNNs true or false?
(A) The last hidden state $s^{(\tau)}$ of the encoder RNN is typically used as a representation of the input sequence and is provided as input to the decoder RNN.
(B) If the last hidden state $s^{(\tau)}$ of the encoder RNN has too small a dimension, it may not be able to represent a long input sequence.

1. True, True
2. True, False
3. False, True
4. False, False



(A) As is shown in the architecture in the picture.
(B) The last hidden state has a fixed dimension. Hence, if we encounter very long input sequence, such a dimension may be too short to contain all the information.