Set A

CSE440: Natural Language Processing II - QUIZ 5

Date: 22 December 2025

This is an MCQ quiz. Each question is worth 1. Write your answers serially on the side.

Name: **Student ID:** **Section:**

1. The hidden state update in a simple RNN is primarily responsible for:
   A. Feature extraction only
   B. Combining current input with past information
   C. Encoding current input only
   D. Preventing overfitting

2. Which property allows an RNN to handle variable-length sequences?
   A. Fixed-size hidden layers
   B. One-hot encoded inputs
   C. Weight sharing across time steps
   D. Softmax output layer

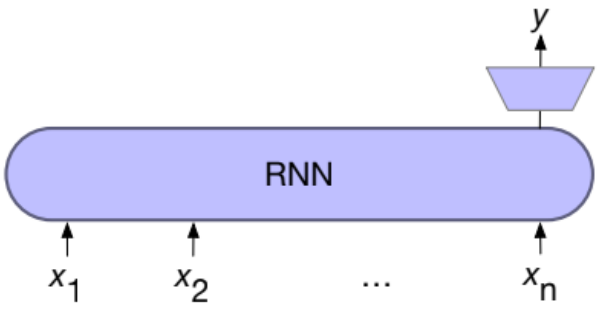
3. What happens if the forget gate outputs values close to zero in LSTM?
   A. Previous memory is preserved
   B. Previous cell state is mostly forgotten
   C. Gradients explode
   D. Output becomes zero

4. In a Bidirectional RNN, how are the outputs of the forward and backward RNN combined at each time step?
   A. By averaging the two outputs
   B. By element-wise multiplication of the two outputs
   C. By concatenating or summing the two hidden states before computing the output
   D. By selecting the output with the higher activation value

Answers:

1	**B**
2	**C**
3	**B**
4	**C**
5	**B**
6	**C**
7	**B**
8	**C**
9	**B**
10	**C**

5. For the following image, which name is most appropriate?



   A. Sequence labeling
   B. Sequence classification
   C. Language modeling
   D. Encoder-Decoder

6. Which issue mainly prevents simple RNNs from learning long-term dependencies?
   A. Overfitting
   B. High computational cost
   C. Vanishing gradients
   D. Lack of non-linearity

7. The LSTM cell state is designed to:
   A. Store short-term memory only
   B. Preserve long-term information
   C. Replace hidden states
   D. Normalize gradients

Answer question 8 to 10 by considering the following RNN:

$$o^{(t)} = softmax(Vh^{(t)})$$
$$h^{(t)} = Uh^{(t-1)} + Wx^{(t)}$$

$$U = \begin{vmatrix} 1 & 1 \\ 0 & 2 \end{vmatrix} \quad W = \begin{vmatrix} 2 & 1 & 0 \\ 1 & 0 & 1 \end{vmatrix} \quad V = \begin{vmatrix} 1 & 0 \\ 0 & 1 \\ -1 & 2 \end{vmatrix}$$

$$h^{(0)} = \begin{vmatrix} 1 \\ 0 \end{vmatrix} \quad x^{(1)} = \begin{vmatrix} 0 \\ 1 \\ 0 \end{vmatrix}$$

8. What is the hidden state $\boldsymbol{h^{(1)}}$?
A. $[0\ 1]^T$
B. $[1\ 2]^T$
C. $[2\ 0]^T$
D. $[1\ 0]^T$

9. Using the $\boldsymbol{h^{(1)}}$ obtained in question 8, what is the pre-softmax score vector $\boldsymbol{Vh^{(1)}}$?
A. $[1\ 2\ 1]^T$
B. $[2\ 0\ -2]^T$
C. $[-2\ 0\ 2]^T$
D. $[2\ 2\ -1]^T$

10. If index 0 is NN, 1 is VB, 2 is DET, what will be the prediction for $\boldsymbol{x^{(1)}}$?
A. DET
B. VB
C. NN
D. Tie among NN and VB

Set B

CSE440: Natural Language Processing II - QUIZ 5

Date: 22 December 2025

This is an MCQ quiz. Each question is worth 1. Write your answers serially on the side.

Name: **Student ID:** **Section:**

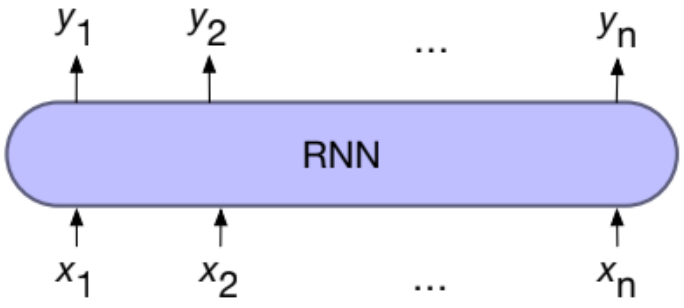
1. In a simple RNN, which parameters remain the same at every time step?
   A. Weights U, W, V
   B. Inputs only
   C. Hidden states only
   D. Outputs only

2. The hidden state update in a simple RNN is primarily responsible for:
   A. Feature extraction only
   B. Combining current input with past information
   C. Encoding current input only
   D. Preventing overfitting

3. Which issue mainly prevents simple RNNs from learning long-term dependencies?
   A. Vanishing gradients
   B. Overfitting
   C. High computational cost
   D. Lack of non-linearity

4. The LSTM cell state is designed to:
   A. Store short-term memory only
   B. Preserve long-term information
   C. Replace hidden states
   D. Normalize gradients

5. For the following image, which name is most appropriate?



   A. Sequence labeling
   B. Sequence classification
   C. Language modeling
   D. Encoder-Decoder

6. What happens if the forget gate outputs values close to one in LSTM?
   A. Previous memory is not preserved
   B. Previous cell state is mostly remembered
   C. Gradients explode
   D. Output becomes zero

Answers:

1	**A**
2	**B**
3	**A**
4	**B**
5	**A**
6	**B**
7	**A**
8	**B**
9	**A**
10	**B**

7. In a Bidirectional RNN, how are the outputs of the forward and backward RNN combined at each time step?

A. By concatenating or summing the two hidden states before computing the output
B. By averaging the two outputs
C. By element-wise multiplication of the two outputs
D. By selecting the output with the higher activation value

Answer question 8 to 10 by considering the following RNN:

$$o^{(t)} = softmax(Vh^{(t)})$$
$$h^{(t)} = Uh^{(t-1)} + Wx^{(t)}$$

$$U = \begin{vmatrix} 1 & 3 \\ 0 & 2 \end{vmatrix} \quad W = \begin{vmatrix} 3 & 1 & 2 \\ 1 & 0 & 1 \end{vmatrix} \quad V = \begin{vmatrix} 1 & 2 \\ 3 & 1 \\ -1 & 2 \end{vmatrix}$$

$$h^{(0)} = \begin{vmatrix} 1 \\ 0 \end{vmatrix} \quad x^{(1)} = \begin{vmatrix} 0 \\ 1 \\ 0 \end{vmatrix}$$

8. What is the hidden state $\boldsymbol{h^{(1)}}$?

A. $[0\ 1]^T$
B. $[5\ 2]^T$
C. $[2\ 5]^T$
D. $[1\ 0]^T$

9. Using the $\boldsymbol{h^{(1)}}$ obtained in question 8, what is the pre-softmax score vector $\boldsymbol{Vh^{(1)}}$?

A. $[9\ 17\ -1]^T$
B. $[1\ 4\ 5]^T$
C. $[-17\ 4\ 12]^T$
D. $[13\ 2\ -1]^T$

10. If index 0 is NN, 1 is VB, 2 is DET, what will be the prediction for $\boldsymbol{x^{(1)}}$?

A. DET
B. VB
C. NN
D. Tie among NN and VB