Started on	Thursday, 12 June 2025, 4:53 PM
State	Finished
Completed on	Thursday, 12 June 2025, 4:58 PM
Time taken	5 mins 9 secs
Marks	7.00/15.00
Grade	**46.67** out of 100.00

Question 1

Complete

Mark 1.00 out of 1.00

In a standard RNN, the hidden state h_t is updated as:

- a. $h_t=\tanh(W x_t + U h_{t-1} + b)$
- b. $h_t=\sigma(W x_t + b)$
- c. $h_t = \tanh(W x_t + b)$
- d. $h_t = \text{ReLU}(x_t)$

Question 2

Complete

Mark 0.00 out of 1.00

In an LSTM cell, what is the function of the cell state C_t?

- a. Acts as the output layer
- b. Calculates gradients
- c. Stores long-term memory
- d. Stores hidden layers

Question 3

Complete

Mark 0.00 out of 1.00

In an LSTM cell, which gate controls how much of the previous hidden state should be carried forward?

- a. Forget gate
- b. Memory gate
- c. Output gate
- d. Input gate

Question 4

Complete

Mark 1.00 out of 1.00

In sequence-to-sequence models, what is the role of the encoder?

- a. Translate output sequence
- b. Predict next token
- c. Encode input sequence into a fixed representation
- d. Update output vocabulary

Question 5

Complete

Mark 1.00 out of 1.00

What does teacher forcing refer to during RNN training?

- a. Using the model's own output as input
- b. Feeding the ground truth output at time t-1 to predict time t
- c. Pre-training the encoder before decoder
- d. Resetting hidden states between batches

Question 6

Complete

Mark 1.00 out of 1.00

What is gradient clipping in the context of training RNNs?

- a. Limiting updates to only the final layer
- b. Reducing batch size to avoid overfitting
- c. Restricting the magnitude of gradients to prevent exploding gradients
- d. Applying dropout to avoid vanishing gradients

Question 7

Complete

Mark 0.00 out of 1.00

What is the main reason RNNs struggle with learning long-term dependencies?

- a. Gradient explosion
- b. Vanishing gradients
- c. Lack of activation functions
- d. Insufficient parameters

Question 8

Complete

Mark 1.00 out of 1.00

What is the primary advantage of using bidirectional RNNs?

- a. Reduced computation time
- b. Works with images
- c. Access to both past and future context
- d. Replaces the need for attention mechanisms

Question 9

Complete

Mark 0.00 out of 1.00

What technique is commonly used during inference in seq2seq models to improve generation quality?

- a. Adam optimizer
- b. Batch normalization
- c. Dropout
- d. Beam search

Question 10

Complete

Mark 0.00 out of 1.00

Which loss function is most commonly used in training sequence-to-sequence models with RNNs for classification?

- a. Hinge Loss
- b. Categorical Crossentropy
- c. Mean Squared Error
- d. Binary Crossentropy

Question 11

Complete

Mark 0.00 out of 1.00

Which mechanism allows RNN-based models to focus on specific parts of the input during decoding?

- a. Dropout
- b. Beam search
- c. Batch normalization
- d. Attention

Question 12

Complete

Mark 0.00 out of 1.00

Which of the following statements about GRU is incorrect?

- a. GRU has fewer parameters than LSTM
- b. GRU combines the forget and input gates into a single update gate
- c. GRU has a separate memory cell c_t like LSTM
- d. GRU is generally faster to train than LSTM

Question 13

Complete

Mark 0.00 out of 1.00

Which one is not a typical application of RNNs?

- a. Machine translation
- b. Speech recognition
- c. Object detection
- d. Sentiment analysis

Question 14

Complete

Mark 1.00 out of 1.00

Which RNN variant is specifically designed to solve the vanishing gradient problem?

- a. LSTM
- b. GRU
- c. Vanilla RNN
- d. Bidirectional RNN

Question 15

Complete

Mark 1.00 out of 1.00

Why are RNNs not inherently parallelizable across time steps?

- a. Each output depends on previous output
- b. Due to weight sharing
- c. They use convolutional filters
- d. They have attention layers