Started on	Thursday, 12 June 2025, 4:53 PM
State	Finished
Completed on	Thursday, 12 June 2025, 5:01 PM
Time taken	7 mins 50 secs
Marks	13.00/15.00
Grade	**86.67** 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. $h_t = \tanh(W x_t + b)$
- b. $h_t=\tanh(W x_t + U h_{t-1} + b)$
- c. $h_t = \text{ReLU}(x_t)$
- d. $h_t=\sigma(W x_t + b)$

Question 2

Complete

Mark 1.00 out of 1.00

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

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

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. Input gate
- c. Memory gate
- d. Output 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. Feeding the ground truth output at time t-1 to predict time t
- b. Using the model's own output as input
- 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. Reducing batch size to avoid overfitting
- b. Limiting updates to only the final layer
- c. Restricting the magnitude of gradients to prevent exploding gradients
- d. Applying dropout to avoid vanishing gradients

Question 7

Complete

Mark 1.00 out of 1.00

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

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

Question 8

Complete

Mark 1.00 out of 1.00

What is the primary advantage of using bidirectional RNNs?

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

Question 9

Complete

Mark 1.00 out of 1.00

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

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

Question 10

Complete

Mark 1.00 out of 1.00

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

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

Question 11

Complete

Mark 1.00 out of 1.00

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

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

Question 12

Complete

Mark 0.00 out of 1.00

Which of the following statements about GRU is incorrect?

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

Question 13

Complete

Mark 1.00 out of 1.00

Which one is not a typical application of RNNs?

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

Question 14

Complete

Mark 1.00 out of 1.00

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

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

Question 15

Complete

Mark 1.00 out of 1.00

Why are RNNs not inherently parallelizable across time steps?

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