Started on	Thursday, 30 January 2025, 11:55 AM
State	Finished
Completed on	Thursday, 30 January 2025, 12:00 PM
Time taken	4 mins 59 secs
Marks	10.00/10.00
Grade	**100.00** out of 100.00

Question 1

Complete

Mark 1.00 out of 1.00

What is the size or shape of Q, K, and V in the ScaledDotProduct Attention forward pass ? (Assuming batch size is B, the sequence length is L, and the hidden size is H)

- a. (H, B, L)
- b. (B, L, H)
- c. (B, H, L)
- d. (L, B, H)

Question 2

Complete

Mark 1.00 out of 1.00

What does the term "scaled dot-product attention" refer to?

- a. Scaling the dot product of query and key vectors by the softmax operation
- b. Applying batch normalization to the query and key vectors
- c. Computing attention weights using convolution operations
- d. Scaling the dot product of query and key vectors by the square root of the key vector dimension

Question 3

Complete

Mark 1.00 out of 1.00

What role does the "key" vector play in the attention mechanism?

- a. It is the final output of the attention module
- b. It is an initialization for positional embeddings
- c. It is used to compute the similarity with the query
- d. It determines the scaling factor for normalization

What are the three main components of the attention mechanism?

- ⦿ a. Queries, Keys, and Values
- ○ b. Keys, Weights, and Queries
- ○ c. Inputs, Outputs, and Weights
- ○ d. Scores, Values, and Gradients

Which of the following is a key step in the attention mechanism?

- ○ a. Ignoring the context vector
- ○ b. Selecting a random set of features
- ○ c. Applying dropout to all layers
- ⦿ d. Calculating the softmax values over all inputs

In the context of the attention mechanism, what does the "query" represent?

- ○ a. The output of the decoder
- ⦿ b. The representation of the task or input focus
- ○ c. The weight matrix of the model
- ○ d. The input sequence to the model

What is the primary purpose of the attention mechanism in neural networks?

- ⦿ a. To focus on relevant parts of the input while processing data
- ○ b. To improve activation functions like ReLU
- ○ c. To increase the number of hidden layers
- ○ d. To reduce the size of the dataset

Question 8

Complete

Mark 1.00 out of 1.00

In the tokenization process, what does word_to_index[token] = len(word_to_index) + 1 accomplish?

- ⦿ a. It maps unique tokens to their corresponding indices.
- ○ b. It removes duplicates from the token list.
- ○ c. It ensures tokens are indexed starting from 0.
- ○ d. It initializes the word embedding layer.

Question 9

Complete

Mark 1.00 out of 1.00

Which of the following types of attention is commonly used in Transformer models?

- ○ a. Local attention
- ○ b. Temporal attention
- ○ c. Global attention
- ⦿ d. Multi-head self-attention

Question 10

Complete

Mark 1.00 out of 1.00

What does the DataLoader class achieve in the code?

- ○ a. It computes the loss and gradients during backpropagation.
- ⦿ b. It batches the data and allows shuffling for training.
- ○ c. It preprocesses the text and sentiment labels.
- ○ d. It splits the data into training and testing sets.