Introduction to Large Language Models

Week-6 Assignment

Number of questions: 8   **Total mark: 6 X 1 + 2 X 2 = 10**

Question 1: [1 mark]

True or False:
RoPE uses additive embeddings like sinusoidal encoding.

Answer: False
Solution: Please refer to slides.

Question 2: [1 mark]

Which of the following is true about *multi-head attention*?

a. It increases model interpretability by using a single set of attention weights
 b. Each head operates on different parts of the input in parallel
 c. It reduces the number of parameters in the model
 d. Heads are averaged before applying the softmax function

Answer: b
Solution: Each attention head processes different learned projections of the input, enabling the model to capture different features.

Question 3: [1 mark]

What is the role of the residual connection in the Transformer architecture?

a. Improve gradient flow during backpropagation

b. Normalize input embeddings

c. Reduce computational complexity

d. Prevent overfitting

Answer: a

Solution: Please refer to lecture slides.

Question 4: [1 mark]

True or False:
The feedforward network in a Transformer block introduces non-linearity between attention layers.

Answer: True

Solution: Please refer to lecture slides.

Question 5: [1 mark]

Fill in the blank:
The sinusoidal positional encoding uses sine for even dimensions and ___ for odd dimensions.

 a. sine
 b. cosine
 c. tangent
 d. None of these

Answer: b

Solution: Please refer to lecture slides.

Question 6: [1 mark]

Why is positional encoding added to input embeddings in Transformers?
 a. To provide unique values for each word
 b. To indicate the position of tokens since Transformers are non-sequential
 c. To scale embeddings
 d. To avoid vanishing gradients

Answer: b

Solution: Please refer to lecture slides.

Question 7: [2 marks]

You are given a self-attention layer with input dimension 512, using 8 heads. What is the output dimension per head?

a. 64
b. 128

c. 32
d. 256

Answer: a
Solution: Each head processes 512/8 = 64 dimensions

QUESTION 8: [2 marks]
For a transformer with d_{model} = 512, calculate the positional encoding for position p=14 and dimensions 6 and 7 using the sinusoidal formula:

$$PE(p, 2i) = sin\left(\frac{p}{10000^{2i/d_{model}}}\right) \quad PE(p, 2i + 1) = cos\left(\frac{p}{10000^{2i/d_{model}}}\right)$$

a. $sin\left(\frac{14}{10000^{3/256}}\right), cos\left(\frac{14}{10000^{3/256}}\right)$

b. $cos\left(\frac{14}{10000^{6/256}}\right), sin\left(\frac{14}{10000^{7/256}}\right)$

c. $cos\left(\frac{14}{10000^{3/256}}\right), sin\left(\frac{14}{10000^{3/256}}\right)$

d. $sin\left(\frac{14}{10000^{3/512}}\right), cos\left(\frac{14}{10000^{3/256}}\right)$

Correct Answer: a

Solution:

For dimension 6, $PE(14,6) = sin\left(\frac{14}{10000^{6/512}}\right) = sin\left(\frac{14}{10000^{3/256}}\right)$

For dimension 7, $PE(14,7) = cos\left(\frac{14}{10000^{6/512}}\right) =$

$cos\left(\frac{14}{10000^{3/256}}\right)$ _____

_____