ASSIGNMENT 3: POSITIONAL ENCODING IN TRANSFORMERS

Deadline: 11:59 pm, April 14, 2025

Submit via Blackboard with VeriGuide receipt.

Please follow the course policy and the school's academic honesty policy.

Submission Instructions:

Students must submit a zip file containing:

- A PDF file with their written solutions.
- The implemented Python file(s) for the assignment.

For the implementation part, students must also include the output of their code execution (either as pasted text or a screenshot) in the PDF file.

1. [20 points] Answer the following in a few sentences:
   (a) Why is positional encoding necessary in Transformer models?
   (b) Why are sine and cosine functions chosen for positional encoding?

2. [80 points] In this problem, you are required to implement the positional encoding mechanism used in the Transformer architecture. Positional encoding injects information about the relative or absolute position of tokens in the sequence, enabling the model to capture token order.

   For each position pos and dimension pair indexed by $i = 0, 1, \ldots, \frac{d_{\text{model}}}{2} - 1$, the positional encoding is defined as:

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

   where d_{model} is the model dimension (assume d_{model} is even).

   (a) Implementation [50 points]: Write a Python function `get_positional_encoding(seq_len, d_model)` that returns a NumPy array of shape $(\text{seq\_len}, d_{\text{model}})$ containing the positional encodings for each position in the sequence. You are allowed to use only the `numpy` library for numerical computations.

   (b) Evaluation [30 points]: Given a model dimension $d_{\text{model}} = 512$:
      i. Report the value of the positional encoding at dimension index 6 for position 10.
      ii. Report the value of the positional encoding at dimension index 7 for position 8.

   You may use or adapt the implementation from part (a) to obtain these values.

*** END ***