NLP ODD ROLL NO. ASSIGNMENT CODE WALKTHROUGH OF TRANSFORMER

RollNo.: 423181

Name: Sumit Sharma

1)Positioning Encoding Module

Initialize the module, where d_model is the embedding size and max_len is the maximum supported sequence length.

Create a matrix to store positional encodings for each position and embedding dimension.

Applies sine to even dimensions and cosine to odd dimensions to generate unique position vectors.

Then add a batch dimension so it can be added to input embeddings.

Then add positional encoding to input embeddings based on sequence length.

2)Transformer Language Model Class

First initializes the model with vocabulary size, embedding dimension, number of attention heads, and number of decoder layers.

Add positional information to embeddings.

Create one Transformer decoder layer with masked multi-head self-attention and feed-forward network. Stacks multiple decoder layers to increase model capacity. Then store embedding dimension for scaling.

3) Forward Pass

It defines how input tokens pass through the model. Adds positional information to embeddings.

Reorders dimensions to match Transformer input format (seq_len, batch, d_model). Creates a causal mask to prevent attention to future tokens.

Applies masked self-attention and feed-forward transformations.

Restores tensor shape to (batch_size, seq_len, d_model) and then generates vocabulary logits for next-word prediction.

4) Training Process

Shifts the sentence to create input–target pairs for next-word prediction.

Computes loss between predicted logits and true word indices.

Computes gradients and updates model parameters.

5) Prediction Step

Selects the most probable next word from the final time step.

Github link : https://github.com/Sumit0717/NLP-Assignment

I have used a set of sentences for training corpus so model performance is not so good. It is just for demonstration how we can use transformer architecture for language modelling task.