

Aist ass3

Tsang Cheuk Hang (SID:1155167650)

April 14, 2025

1 Quesiton 1

1.1 Why is positional encoding necessary in Transformer models?

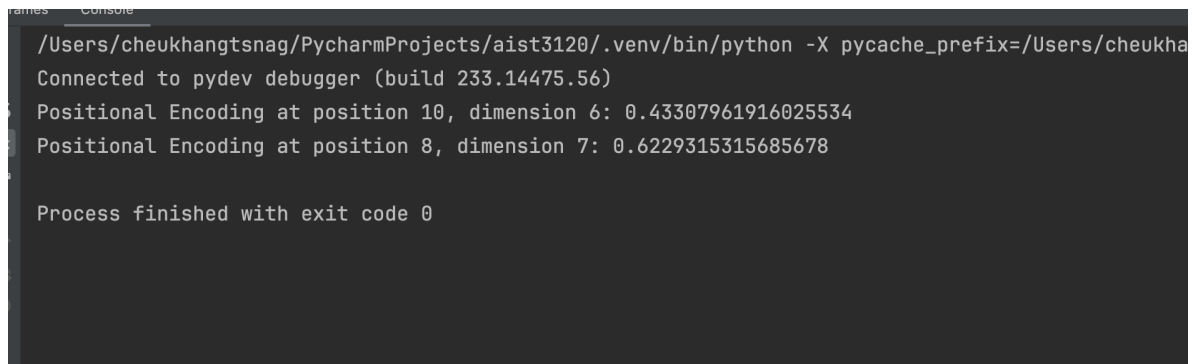
For the model to make use of the order of the sequence, the Transformer model inject some information about the relative or absolute position of the tokens in the sequence. Therefore, the Transformer model add "positional encodings" to the input embeddings at the bottoms of the encoder and decoder stacks. The positional encodings have the same dimension d_{model} as the embeddings, so that the two can be summed.

1.2 Why are sine and cosine functions chosen for positional encoding?

The sine and cosine functions chosen for positional encoding in Transformer models are because the designers hypothesized that it would allow the model to easialy learn to attend by relative positions, since for any fixed offset k , PE_{pos+k} can be represented as a linear function of PE_{pos} . Additionally, the authors of the Transformer models also explored using different learned positional embeddings, but the results were nearly the same.

2 Quesiton 2

2.1 (b) Evaluation



```

/Users/cheukhangtsnag/PycharmProjects/aist3120/.venv/bin/python -X pycache_prefix=/Users/cheukha
Connected to pydev debugger (build 233.14475.56)
Positional Encoding at position 10, dimension 6: 0.43307961916025534
Positional Encoding at position 8, dimension 7: 0.6229315315685678

Process finished with exit code 0

```

Figure 1: Question 2