

Marked exercises, lecture 5

This is the last set of marked exercises, which are a part of Submission 1. In this assignment, you will get familiar with attention and positional encoding parts of a transformer architecture. Detailed instructions and explanations are provided in the accompanying Jupyter notebook. In total, you can get 20 points out of 60 points for Submission 1 for completing all exercises related to lecture 5.

1.5.1 [2 points] Implementation of Scaled dot-product.

1.5.2 [8 points] Implementation of Multi-head attention.

1.5.3 [2 points] Matching manual and PyTorch implementation of Multi-head attention.

1.5.4 [4 points] Implementation of Absolute Positional Encoding.

1.5.5 [4 points] Get familiar with and answer questions about Rotary Positional Embedding.