CSE-5368 Neural Networks
Quiz 05

Consider the encoder part of a multi-head Transformer
Assuming:

- Input embedding size: $d_{model} = 200$
- Sentence length: $L = 1000$ (number of tokens in the sentence)
- Number of attention heads: $n_{heads} = 10$
- Dimension of the expansion layer: 800

Ignore Biases

Assume $net = XW$

What are the shapes of the following matrices?

W_Q^i (The shape of the W_Q for the head number i) Note: Show shape for ONE HEAD

W_K^i (The shape of the W_K for the head number i) Note: Show shape for ONE HEAD

W_V^i (The shape of the W_V for the head number i) Note Show shape for ONE HEAD

W_O Output Projection Matrix

W_1 First layer of Feed-Forward Network (Expansion Layer)

W_2 Second layer of Feed-Forward Network (Compression Layer)