Thursday Mar 27, 2025

Name:

Discuss and complete the following questions with the person nearest you. You **may** be asked to share your thoughts with the class.

1. How long will winter last here?

2. Recall, the attention mechanism is represented by this equation (we will ignore the generation of K, Q, and V from weights here):

$$\text{Softmax}(Q \cdot K^T)V$$

What is the complexity of this operation assuming a sequence length of n and a dimensionality of d (e.g., each element in the sequence n has a value vector of size d)? Help: Consult this wikipedia article, Computational complexity of mathematical operations.

3. Give an attention matrix with the attention weights (including normalization) that you would need to have every word attend to its left and right neighbor (if applicable) for a sequence length of 5.