

Homework 10

Student Number:

Name:

Problem 1. (20 points) Please provide the calculation formula for self-attention in Transformers, and explain why \sqrt{dk} is introduced as the denominator in softmax.

Problem 2. (30 points) Please answer the following questions: What is the difference between the structure of the Decoder and Encoder in Transformers? Briefly explain the differences between the Bert, Bart, T5, and GPT models, and what tasks they are suitable for.

Problem 3. (20 points) Here's a question on calculating ROUGE-L, Given two sentences:
Reference sentence: "The quick brown fox jumps over the lazy dog."
Hypothesis sentence: "A brown fox jumps over a lazy dog quickly."
Calculate the ROUGE-L score between the reference and hypothesis sentences.

Problem 4. (30 points) Can you explain the advantages of Prompt Tuning mentioned in Slice compared to traditional text-based prompts? Also, why don't we use Fine-Tuning to handle LLM?