LLM Assignment 3

Avneesh Sharma

2021030

1. Accuracy comparison between the pretrained and fine-tuned models on the test set.

Pre-trained model accuracy: 34%

Fine-tuned model accuracy: 25%

2. Time taken to fine-tune the model using QLoRA.

Approx 200 minutes for 4 epochs

0.36 iterations/second

3. Total parameters in the model and the number of parameters fine-tuned.

Total parameters: 2666915840

Total trainable parameters:

ΔW=A·B

- A has a shape of $d \times r$
- B has a shape of $r \times d$

Trainable Parameters from LoRA=(d×r)+(r×d)=2·d·r
2·d·r=2·4096·16=131072

4. Resources used (e.g., hardware, memory) during fine-tuning.

Memory used: Less than 12.67 GB RAM

Hardware: T4 GPU on Colab

5. Failure cases of the pretrained model that were corrected by the fine-tuned model, as well as those that were not corrected. Provide possible explanations for both.