Assignment 3- Report

1. Approach and Methodology

Dataset Preparation

The SNLI dataset on Hugging Face was used, and samples were selected for training, validation, and testing:

- **Training set**: 1000 samples selected by taking every 550th sample from the full 550k dataset.
- **Validation set**: 100 samples by choosing every 100th sample from the 10k validation set.
- **Testing set**: 100 samples by choosing every 100th sample from the 10k testing set.

This subset selection helped reduce computational costs while providing a manageable dataset for evaluation.

Prompt

Choose - 0 - entailment, 1 - neutral, 2 - contradiction

Premise:{premise}

Hypothesis:{hypothesis}

Output: {label}

Model and Fine-Tuning

Using QLoRA, the Phi2 model was optimized over five epochs, with the model being saved at each one. To better capture NLI-specific linguistic traits, the model's pre-trained parameters were modified in response to the training data.

Hyperparameters

The fine-tuning was done for 5 epochs, monitoring loss over epochs to ensure convergence.

2. Results

Accuracy Comparison

The table below compares the accuracy between the pretrained and fine-tuned models on the test set:

Model	Test Accuracy
Pretrained Model	0.31
Fine-Tuned Model	0.89

The fine-tuned model successfully adapted to the NLI challenge, as evidenced by its improvement over the pre-trained model.

Model Parameters

- Total Parameters: 1,563,335,680
- Trainable Parameters: 41,943,040
- Percentage of Trainable Parameters: 2.68%

Time taken

Training time taken - 22 min 50 s

Resources Used
Hardware - T4 GPU
Memory used - 1.25 GB of CPU RAM for storing model of each epoch
3.1 GB of GPU RAM for storing the original model

3. Observations on Failure Cases

Fine-tuning fixed a few of the pretrained model's failure scenarios, particularly when it came to complex linguistic conclusions. Even after fine-tuning, certain intricate sentence forms continued to produce inaccurate assumptions. These shortcomings may be due to:

- **Ambiguous Language**: Sentences with ambiguous or complex phrases that the fine-tuning process might need more data to fully resolve.
- **Data Limitations**: The small training set may have limited the model's exposure to diverse sentence structures.

Example-

Wrong outputs of the pretrained model that were correctly predicted by the fine-tuned model-

Prompt -

Premise:many children play in the water.

Hypothesis:Some kids splash in the water and interact with each other.

Choose the correct option - 0 - entailment, 1 - neutral, 2 - contradiction -

Correct Label -

1 (Neutral) since children may or may not be interacting with each other

Original Model Output -

0 (Entailment) it may have confused interacting and splashing with playing.

QLORA fine-tuned output-

1 (Neutral) the model now more carefully checks all the details and does not correlate interacting with playing.

Prompt -

Premise: A woman within an orchestra is playing a violin.

Hypothesis: A woman is playing a concert.

 Choose the correct option - 0 - entailment, 1 - neutral, 2 - contradiction -

Correct Label -

1 (Neutral) since children may or may not be interacting with each other

Original Model Output -

0 (Entailment) Model has confused orchestra with a concert

QLORA fine-tuned output-

1 (Neutral) the model now understands that orchestra and concerts are different things.

Wrong outputs of the pretrained model that were still wrongly predicted by the fine-tuned model-

Prompt-

Premise: A bicycle rider wearing racing gear pedals a yellow bike past the wire fence at the edge of a field, with a stand of trees in the background.

Hypothesis: Someone was trying to take a shortcut home.

Choose the correct option - 0 - entailment, 1 - neutral, 2 - contradiction-

Correct Label -

2 (Contradiction) since the rider seems to be suffering from an accident, not taking a shortcut.

Original Model Output -

0 (Entailment) since the model feels like the rider is taking a shortcut.

QLORA fine-tuned output-

1 (Neutral) since the model believes now thinks that the model may or may not be taking a shortcut.

Prompt-

Premise: Three soccer players, two in orange one in yellow, running for the ball on a soccer field.

Hypothesis: Some teams were playing ball in the field.

Choose the correct option - 0 - entailment, 1 - neutral, 2 - contradiction-

Correct Label -

1 (Neutral) since players may or may not be in the same team.

Original Model Output -

0 (Entailment) since the model has confused that two different colours of jerseys represent two different teams.

QLORA fine-tuned output-

0 (Entailment) since the model still makes the same assumption due to the fact that it is very common that two different colours represent the different teams.