

Indian Institute of Technology Gandhinagar



Assignment 3:

Pretraining and fine-tuning an LLM

CS 613: NLP

Group T10: Exquisiters

Aditi Dey (20110007)

Anavart Pandya (20110016)

Ayush Gupta (20110031)

Mann Kumar Jain (20110108)

Pinki Kumari(22210028)

Ronak Kalra (20110171)

Sakshi Jain (20110181)

Vedang Chavan (20110222)

Task 1: Bert-base-uncased model.

Task 2: Parameters of the selected model from the code

Modules	Parameters
embeddings.word_embeddings.weight	23440896
embeddings.position_embeddings.weight	393216
embeddings.token_type_embeddings.weight	1536
embeddings.LayerNorm.weight	768
embeddings.LayerNorm.bias	768
encoder.layer.0.attention.self.query.weight	589824
encoder.layer.0.attention.self.query.bias	768
encoder.layer.0.attention.self.key.weight	589824
encoder.layer.0.attention.self.key.bias	768
encoder.layer.0.attention.self.value.weight	589824
encoder.layer.0.attention.self.value.bias	768
encoder.layer.0.attention.output.dense.weight	589824
encoder.layer.0.attention.output.dense.bias	768
encoder.layer.0.attention.output.LayerNorm.weight	768
encoder.layer.0.attention.output.LayerNorm.bias	768
encoder.layer.0.intermediate.dense.weight	2359296
encoder.layer.0.intermediate.dense.bias	3072
encoder.layer.0.output.dense.weight	2359296
encoder.layer.0.output.dense.bias	768
encoder.layer.0.output.LayerNorm.weight	768
encoder.layer.0.output.LayerNorm.bias	768
pooler.dense.weight	589824
pooler.dense.bias	768

There were 12 such encoder layers.

Hence, Total Trainable Params: 109482240

The model's parameter count was computed to be **109,482,240** through the code. This count closely aligns with the parameter figure of **108,432,384** as reported in the corresponding paper. The slight variance between the calculated and reported counts can potentially stem from differences in implementation details, such as parameter initialization methods and biases. The exact number of parameters from the different layers were also calculated, and have been attached in the end.

Task 3: Pretrain the selected model on the train split of 'wikitext-2-raw-v1'

The Models have been pushed to 🤗. We used the following hyperparameters:

Learning Rate = 5e-5

Epochs = 5

Batch Size: 36718
Sequence Length: 512

Models were trained using the specified hyperparameters for 5 epochs, each epoch involving batches of 36718 samples, with sequences padded or truncated to a maximum length of 512 tokens. The above hyperparameters are also inspired by the research paper.

Task 4: The Perplexity scores

The perplexity score decreases after each epoch during training. However, the improvement saturates. There are multiple reasons for this:

- Initially, as the model learns more from the training data, its ability to predict sequences improves, resulting in decreased perplexity on the test data.
- The choice of hyperparameters and the model architecture influence the convergence pattern.
- After some epochs, the model starts memorizing the training data, causing overfitting. This might lead to an increase in perplexity on the test data.

Task 5: The pre-trained model has been pushed to 🤗.

Here is the link to the pushed model: [link](#)

Task 6 : The final pre-trained model has been fine-tuned on the train split of the dataset for the below tasks:

- a. Classification: SST-2
- b. Question-Answering: SQuAD

Task 7: For each of the following metrics corresponding to their particular task, the scores have been calculated on the test splits of the dataset:

- a. Classification: Accuracy, Precision, Recall, F1 Scores:

METRIC	SCORE
Recall	1.0
Precision	0.551
Accuracy	0.551
F1	0.7105

- b. Question-Answering: squad_v2, F1, METEOR, BLEU, ROUGE, exact-match Scores:

METRIC	SCORE
exact	7.479154384
f1	12.1049688
total	11873
HasAns_exact	14.97975709
HasAns_f1	24.24465159
HasAns_total	5928
NoAns_exact	0
NoAns_f1	0
NoAns_total	5945
best_exact	50.08843595
best_exact_thresh	0
best_f1	50.09257061
best_f1_thresh	0
meteor	0.0009801603959
bleu	0
precisions	[1.5039629423531004e-05, 0.0, 0.0, 0.0]
brevity_penalty	1
length_ratio	5.600185294
translation_length	66491
reference_length	11873
rouge1	2.11E-05
rouge2	0
rougeL	2.11E-05
rougeLsum	2.11E-05

In assessing the ROPES (Reading for Open-Ended Problem Solving) dataset, F1 score is the preferred metric, measuring a system's adeptness at applying passage knowledge to novel situations. For the evaluation of generative Question-Answering (QA) models, METEOR stands out as the optimal metric, providing a thorough evaluation of the model's performance, especially in generating responses that emulate human-like language and expression.

Task 8 : Number of parameters in the model after fine-tuning are:

1.model_finetuned_QA:

+-----+-----+	
Modules	Parameters
+-----+-----+	
bert.embeddings.word_embeddings.weight	23440896
bert.embeddings.position_embeddings.weight	393216
bert.embeddings.token_type_embeddings.weight	1536
bert.embeddings.LayerNorm.weight	768
bert.embeddings.LayerNorm.bias	768
bert.encoder.layer.0.attention.self.query.weight	589824
bert.encoder.layer.0.attention.self.query.bias	768
bert.encoder.layer.0.attention.self.key.weight	589824
bert.encoder.layer.0.attention.self.key.bias	768
bert.encoder.layer.0.attention.self.value.weight	589824
bert.encoder.layer.0.attention.self.value.bias	768
bert.encoder.layer.0.attention.output.dense.weight	589824
bert.encoder.layer.0.attention.output.dense.bias	768
bert.encoder.layer.0.attention.output.LayerNorm.weight	768
bert.encoder.layer.0.attention.output.LayerNorm.bias	768
bert.encoder.layer.0.intermediate.dense.weight	2359296
bert.encoder.layer.0.intermediate.dense.bias	3072
bert.encoder.layer.0.output.dense.weight	2359296
bert.encoder.layer.0.output.dense.bias	768
bert.encoder.layer.0.output.LayerNorm.weight	768
bert.encoder.layer.0.output.LayerNorm.bias	768
bert.encoder.layer.1.attention.self.query.weight	589824
bert.encoder.layer.1.attention.self.query.bias	768
bert.encoder.layer.1.attention.self.key.weight	589824
bert.encoder.layer.1.attention.self.key.bias	768
bert.encoder.layer.1.attention.self.value.weight	589824
bert.encoder.layer.1.attention.self.value.bias	768
bert.encoder.layer.1.attention.output.dense.weight	589824
bert.encoder.layer.1.attention.output.dense.bias	768
bert.encoder.layer.1.attention.output.LayerNorm.weight	768
bert.encoder.layer.1.attention.output.LayerNorm.bias	768
bert.encoder.layer.1.intermediate.dense.weight	2359296
bert.encoder.layer.1.intermediate.dense.bias	3072
bert.encoder.layer.1.output.dense.weight	2359296
bert.encoder.layer.1.output.dense.bias	768
bert.encoder.layer.1.output.LayerNorm.weight	768
bert.encoder.layer.1.output.LayerNorm.bias	768
bert.encoder.layer.2.attention.self.query.weight	589824
bert.encoder.layer.2.attention.self.query.bias	768

bert.encoder.layer.2.attention.self.key.weight	589824	
bert.encoder.layer.2.attention.self.key.bias	768	
bert.encoder.layer.2.attention.self.value.weight	589824	
bert.encoder.layer.2.attention.self.value.bias	768	
bert.encoder.layer.2.attention.output.dense.weight	589824	
bert.encoder.layer.2.attention.output.dense.bias	768	
bert.encoder.layer.2.attention.output.LayerNorm.weight	768	
bert.encoder.layer.2.attention.output.LayerNorm.bias	768	
bert.encoder.layer.2.intermediate.dense.weight	2359296	
bert.encoder.layer.2.intermediate.dense.bias	3072	
bert.encoder.layer.2.output.dense.weight	2359296	
bert.encoder.layer.2.output.dense.bias	768	
bert.encoder.layer.2.output.LayerNorm.weight	768	
bert.encoder.layer.2.output.LayerNorm.bias	768	
bert.encoder.layer.3.attention.self.query.weight	589824	
bert.encoder.layer.3.attention.self.query.bias	768	
bert.encoder.layer.3.attention.self.key.weight	589824	
bert.encoder.layer.3.attention.self.key.bias	768	
bert.encoder.layer.3.attention.self.value.weight	589824	
bert.encoder.layer.3.attention.self.value.bias	768	
bert.encoder.layer.3.attention.output.dense.weight	589824	
bert.encoder.layer.3.attention.output.dense.bias	768	
bert.encoder.layer.3.attention.output.LayerNorm.weight	768	
bert.encoder.layer.3.attention.output.LayerNorm.bias	768	
bert.encoder.layer.3.intermediate.dense.weight	2359296	
bert.encoder.layer.3.intermediate.dense.bias	3072	
bert.encoder.layer.3.output.dense.weight	2359296	
bert.encoder.layer.3.output.dense.bias	768	
bert.encoder.layer.3.output.LayerNorm.weight	768	
bert.encoder.layer.3.output.LayerNorm.bias	768	
bert.encoder.layer.4.attention.self.query.weight	589824	
bert.encoder.layer.4.attention.self.query.bias	768	
bert.encoder.layer.4.attention.self.key.weight	589824	
bert.encoder.layer.4.attention.self.key.bias	768	
bert.encoder.layer.4.attention.self.value.weight	589824	
bert.encoder.layer.4.attention.self.value.bias	768	
bert.encoder.layer.4.attention.output.dense.weight	589824	
bert.encoder.layer.4.attention.output.dense.bias	768	
bert.encoder.layer.4.attention.output.LayerNorm.weight	768	
bert.encoder.layer.4.attention.output.LayerNorm.bias	768	
bert.encoder.layer.4.intermediate.dense.weight	2359296	
bert.encoder.layer.4.intermediate.dense.bias	3072	
bert.encoder.layer.4.output.dense.weight	2359296	
bert.encoder.layer.4.output.dense.bias	768	

bert.encoder.layer.4.output.LayerNorm.weight	768
bert.encoder.layer.4.output.LayerNorm.bias	768
bert.encoder.layer.5.attention.self.query.weight	589824
bert.encoder.layer.5.attention.self.query.bias	768
bert.encoder.layer.5.attention.self.key.weight	589824
bert.encoder.layer.5.attention.self.key.bias	768
bert.encoder.layer.5.attention.self.value.weight	589824
bert.encoder.layer.5.attention.self.value.bias	768
bert.encoder.layer.5.attention.output.dense.weight	589824
bert.encoder.layer.5.attention.output.dense.bias	768
bert.encoder.layer.5.attention.output.LayerNorm.weight	768
bert.encoder.layer.5.attention.output.LayerNorm.bias	768
bert.encoder.layer.5.intermediate.dense.weight	2359296
bert.encoder.layer.5.intermediate.dense.bias	3072
bert.encoder.layer.5.output.dense.weight	2359296
bert.encoder.layer.5.output.dense.bias	768
bert.encoder.layer.5.output.LayerNorm.weight	768
bert.encoder.layer.5.output.LayerNorm.bias	768
bert.encoder.layer.6.attention.self.query.weight	589824
bert.encoder.layer.6.attention.self.query.bias	768
bert.encoder.layer.6.attention.self.key.weight	589824
bert.encoder.layer.6.attention.self.key.bias	768
bert.encoder.layer.6.attention.self.value.weight	589824
bert.encoder.layer.6.attention.self.value.bias	768
bert.encoder.layer.6.attention.output.dense.weight	589824
bert.encoder.layer.6.attention.output.dense.bias	768
bert.encoder.layer.6.attention.output.LayerNorm.weight	768
bert.encoder.layer.6.attention.output.LayerNorm.bias	768
bert.encoder.layer.6.intermediate.dense.weight	2359296
bert.encoder.layer.6.intermediate.dense.bias	3072
bert.encoder.layer.6.output.dense.weight	2359296
bert.encoder.layer.6.output.dense.bias	768
bert.encoder.layer.6.output.LayerNorm.weight	768
bert.encoder.layer.6.output.LayerNorm.bias	768
bert.encoder.layer.7.attention.self.query.weight	589824
bert.encoder.layer.7.attention.self.query.bias	768
bert.encoder.layer.7.attention.self.key.weight	589824
bert.encoder.layer.7.attention.self.key.bias	768
bert.encoder.layer.7.attention.self.value.weight	589824
bert.encoder.layer.7.attention.self.value.bias	768
bert.encoder.layer.7.attention.output.dense.weight	589824
bert.encoder.layer.7.attention.output.dense.bias	768
bert.encoder.layer.7.attention.output.LayerNorm.weight	768
bert.encoder.layer.7.attention.output.LayerNorm.bias	768

bert.encoder.layer.7.intermediate.dense.weight	2359296	
bert.encoder.layer.7.intermediate.dense.bias	3072	
bert.encoder.layer.7.output.dense.weight	2359296	
bert.encoder.layer.7.output.dense.bias	768	
bert.encoder.layer.7.output.LayerNorm.weight	768	
bert.encoder.layer.7.output.LayerNorm.bias	768	
bert.encoder.layer.8.attention.self.query.weight	589824	
bert.encoder.layer.8.attention.self.query.bias	768	
bert.encoder.layer.8.attention.self.key.weight	589824	
bert.encoder.layer.8.attention.self.key.bias	768	
bert.encoder.layer.8.attention.self.value.weight	589824	
bert.encoder.layer.8.attention.self.value.bias	768	
bert.encoder.layer.8.attention.output.dense.weight	589824	
bert.encoder.layer.8.attention.output.dense.bias	768	
bert.encoder.layer.8.attention.output.LayerNorm.weight	768	
bert.encoder.layer.8.attention.output.LayerNorm.bias	768	
bert.encoder.layer.8.intermediate.dense.weight	2359296	
bert.encoder.layer.8.intermediate.dense.bias	3072	
bert.encoder.layer.8.output.dense.weight	2359296	
bert.encoder.layer.8.output.dense.bias	768	
bert.encoder.layer.8.output.LayerNorm.weight	768	
bert.encoder.layer.8.output.LayerNorm.bias	768	
bert.encoder.layer.9.attention.self.query.weight	589824	
bert.encoder.layer.9.attention.self.query.bias	768	
bert.encoder.layer.9.attention.self.key.weight	589824	
bert.encoder.layer.9.attention.self.key.bias	768	
bert.encoder.layer.9.attention.self.value.weight	589824	
bert.encoder.layer.9.attention.self.value.bias	768	
bert.encoder.layer.9.attention.output.dense.weight	589824	
bert.encoder.layer.9.attention.output.dense.bias	768	
bert.encoder.layer.9.attention.output.LayerNorm.weight	768	
bert.encoder.layer.9.attention.output.LayerNorm.bias	768	
bert.encoder.layer.9.intermediate.dense.weight	2359296	
bert.encoder.layer.9.intermediate.dense.bias	3072	
bert.encoder.layer.9.output.dense.weight	2359296	
bert.encoder.layer.9.output.dense.bias	768	
bert.encoder.layer.9.output.LayerNorm.weight	768	
bert.encoder.layer.9.output.LayerNorm.bias	768	
bert.encoder.layer.10.attention.self.query.weight	589824	
bert.encoder.layer.10.attention.self.query.bias	768	
bert.encoder.layer.10.attention.self.key.weight	589824	
bert.encoder.layer.10.attention.self.key.bias	768	
bert.encoder.layer.10.attention.self.value.weight	589824	
bert.encoder.layer.10.attention.self.value.bias	768	

bert.encoder.layer.10.attention.output.dense.weight	589824
bert.encoder.layer.10.attention.output.dense.bias	768
bert.encoder.layer.10.attention.output.LayerNorm.weight	768
bert.encoder.layer.10.attention.output.LayerNorm.bias	768
bert.encoder.layer.10.intermediate.dense.weight	2359296
bert.encoder.layer.10.intermediate.dense.bias	3072
bert.encoder.layer.10.output.dense.weight	2359296
bert.encoder.layer.10.output.dense.bias	768
bert.encoder.layer.10.output.LayerNorm.weight	768
bert.encoder.layer.10.output.LayerNorm.bias	768
bert.encoder.layer.11.attention.self.query.weight	589824
bert.encoder.layer.11.attention.self.query.bias	768
bert.encoder.layer.11.attention.self.key.weight	589824
bert.encoder.layer.11.attention.self.key.bias	768
bert.encoder.layer.11.attention.self.value.weight	589824
bert.encoder.layer.11.attention.self.value.bias	768
bert.encoder.layer.11.attention.output.dense.weight	589824
bert.encoder.layer.11.attention.output.dense.bias	768
bert.encoder.layer.11.attention.output.LayerNorm.weight	768
bert.encoder.layer.11.attention.output.LayerNorm.bias	768
bert.encoder.layer.11.intermediate.dense.weight	2359296
bert.encoder.layer.11.intermediate.dense.bias	3072
bert.encoder.layer.11.output.dense.weight	2359296
bert.encoder.layer.11.output.dense.bias	768
bert.encoder.layer.11.output.LayerNorm.weight	768
bert.encoder.layer.11.output.LayerNorm.bias	768
qa_outputs.weight	1536
qa_outputs.bias	2

+-----+-----+

Total Trainable Params: 108893186

2.model_finetuned_classification:

+-----+-----+

Modules	Parameters
+-----+-----+	
bert.embeddings.word_embeddings.weight	23440896
bert.embeddings.position_embeddings.weight	393216
bert.embeddings.token_type_embeddings.weight	1536
bert.embeddings.LayerNorm.weight	768
bert.embeddings.LayerNorm.bias	768
bert.encoder.layer.0.attention.self.query.weight	589824
bert.encoder.layer.0.attention.self.query.bias	768
bert.encoder.layer.0.attention.self.key.weight	589824
bert.encoder.layer.0.attention.self.key.bias	768

bert.encoder.layer.0.attention.self.value.weight	589824	
bert.encoder.layer.0.attention.self.value.bias	768	
bert.encoder.layer.0.attention.output.dense.weight	589824	
bert.encoder.layer.0.attention.output.dense.bias	768	
bert.encoder.layer.0.attention.output.LayerNorm.weight	768	
bert.encoder.layer.0.attention.output.LayerNorm.bias	768	
bert.encoder.layer.0.intermediate.dense.weight	2359296	
bert.encoder.layer.0.intermediate.dense.bias	3072	
bert.encoder.layer.0.output.dense.weight	2359296	
bert.encoder.layer.0.output.dense.bias	768	
bert.encoder.layer.0.output.LayerNorm.weight	768	
bert.encoder.layer.0.output.LayerNorm.bias	768	
bert.encoder.layer.1.attention.self.query.weight	589824	
bert.encoder.layer.1.attention.self.query.bias	768	
bert.encoder.layer.1.attention.self.key.weight	589824	
bert.encoder.layer.1.attention.self.key.bias	768	
bert.encoder.layer.1.attention.self.value.weight	589824	
bert.encoder.layer.1.attention.self.value.bias	768	
bert.encoder.layer.1.attention.output.dense.weight	589824	
bert.encoder.layer.1.attention.output.dense.bias	768	
bert.encoder.layer.1.attention.output.LayerNorm.weight	768	
bert.encoder.layer.1.attention.output.LayerNorm.bias	768	
bert.encoder.layer.1.intermediate.dense.weight	2359296	
bert.encoder.layer.1.intermediate.dense.bias	3072	
bert.encoder.layer.1.output.dense.weight	2359296	
bert.encoder.layer.1.output.dense.bias	768	
bert.encoder.layer.1.output.LayerNorm.weight	768	
bert.encoder.layer.1.output.LayerNorm.bias	768	
bert.encoder.layer.2.attention.self.query.weight	589824	
bert.encoder.layer.2.attention.self.query.bias	768	
bert.encoder.layer.2.attention.self.key.weight	589824	
bert.encoder.layer.2.attention.self.key.bias	768	
bert.encoder.layer.2.attention.self.value.weight	589824	
bert.encoder.layer.2.attention.self.value.bias	768	
bert.encoder.layer.2.attention.output.dense.weight	589824	
bert.encoder.layer.2.attention.output.dense.bias	768	
bert.encoder.layer.2.attention.output.LayerNorm.weight	768	
bert.encoder.layer.2.attention.output.LayerNorm.bias	768	
bert.encoder.layer.2.intermediate.dense.weight	2359296	
bert.encoder.layer.2.intermediate.dense.bias	3072	
bert.encoder.layer.2.output.dense.weight	2359296	
bert.encoder.layer.2.output.dense.bias	768	
bert.encoder.layer.2.output.LayerNorm.weight	768	
bert.encoder.layer.2.output.LayerNorm.bias	768	

bert.encoder.layer.3.attention.self.query.weight	589824	
bert.encoder.layer.3.attention.self.query.bias	768	
bert.encoder.layer.3.attention.self.key.weight	589824	
bert.encoder.layer.3.attention.self.key.bias	768	
bert.encoder.layer.3.attention.self.value.weight	589824	
bert.encoder.layer.3.attention.self.value.bias	768	
bert.encoder.layer.3.attention.output.dense.weight	589824	
bert.encoder.layer.3.attention.output.dense.bias	768	
bert.encoder.layer.3.attention.output.LayerNorm.weight	768	
bert.encoder.layer.3.attention.output.LayerNorm.bias	768	
bert.encoder.layer.3.intermediate.dense.weight	2359296	
bert.encoder.layer.3.intermediate.dense.bias	3072	
bert.encoder.layer.3.output.dense.weight	2359296	
bert.encoder.layer.3.output.dense.bias	768	
bert.encoder.layer.3.output.LayerNorm.weight	768	
bert.encoder.layer.3.output.LayerNorm.bias	768	
bert.encoder.layer.4.attention.self.query.weight	589824	
bert.encoder.layer.4.attention.self.query.bias	768	
bert.encoder.layer.4.attention.self.key.weight	589824	
bert.encoder.layer.4.attention.self.key.bias	768	
bert.encoder.layer.4.attention.self.value.weight	589824	
bert.encoder.layer.4.attention.self.value.bias	768	
bert.encoder.layer.4.attention.output.dense.weight	589824	
bert.encoder.layer.4.attention.output.dense.bias	768	
bert.encoder.layer.4.attention.output.LayerNorm.weight	768	
bert.encoder.layer.4.attention.output.LayerNorm.bias	768	
bert.encoder.layer.4.intermediate.dense.weight	2359296	
bert.encoder.layer.4.intermediate.dense.bias	3072	
bert.encoder.layer.4.output.dense.weight	2359296	
bert.encoder.layer.4.output.dense.bias	768	
bert.encoder.layer.4.output.LayerNorm.weight	768	
bert.encoder.layer.4.output.LayerNorm.bias	768	
bert.encoder.layer.5.attention.self.query.weight	589824	
bert.encoder.layer.5.attention.self.query.bias	768	
bert.encoder.layer.5.attention.self.key.weight	589824	
bert.encoder.layer.5.attention.self.key.bias	768	
bert.encoder.layer.5.attention.self.value.weight	589824	
bert.encoder.layer.5.attention.self.value.bias	768	
bert.encoder.layer.5.attention.output.dense.weight	589824	
bert.encoder.layer.5.attention.output.dense.bias	768	
bert.encoder.layer.5.attention.output.LayerNorm.weight	768	
bert.encoder.layer.5.attention.output.LayerNorm.bias	768	
bert.encoder.layer.5.intermediate.dense.weight	2359296	
bert.encoder.layer.5.intermediate.dense.bias	3072	

bert.encoder.layer.5.output.dense.weight	2359296
bert.encoder.layer.5.output.dense.bias	768
bert.encoder.layer.5.output.LayerNorm.weight	768
bert.encoder.layer.5.output.LayerNorm.bias	768
bert.encoder.layer.6.attention.self.query.weight	589824
bert.encoder.layer.6.attention.self.query.bias	768
bert.encoder.layer.6.attention.self.key.weight	589824
bert.encoder.layer.6.attention.self.key.bias	768
bert.encoder.layer.6.attention.self.value.weight	589824
bert.encoder.layer.6.attention.self.value.bias	768
bert.encoder.layer.6.attention.output.dense.weight	589824
bert.encoder.layer.6.attention.output.dense.bias	768
bert.encoder.layer.6.attention.output.LayerNorm.weight	768
bert.encoder.layer.6.attention.output.LayerNorm.bias	768
bert.encoder.layer.6.intermediate.dense.weight	2359296
bert.encoder.layer.6.intermediate.dense.bias	3072
bert.encoder.layer.6.output.dense.weight	2359296
bert.encoder.layer.6.output.dense.bias	768
bert.encoder.layer.6.output.LayerNorm.weight	768
bert.encoder.layer.6.output.LayerNorm.bias	768
bert.encoder.layer.7.attention.self.query.weight	589824
bert.encoder.layer.7.attention.self.query.bias	768
bert.encoder.layer.7.attention.self.key.weight	589824
bert.encoder.layer.7.attention.self.key.bias	768
bert.encoder.layer.7.attention.self.value.weight	589824
bert.encoder.layer.7.attention.self.value.bias	768
bert.encoder.layer.7.attention.output.dense.weight	589824
bert.encoder.layer.7.attention.output.dense.bias	768
bert.encoder.layer.7.attention.output.LayerNorm.weight	768
bert.encoder.layer.7.attention.output.LayerNorm.bias	768
bert.encoder.layer.7.intermediate.dense.weight	2359296
bert.encoder.layer.7.intermediate.dense.bias	3072
bert.encoder.layer.7.output.dense.weight	2359296
bert.encoder.layer.7.output.dense.bias	768
bert.encoder.layer.7.output.LayerNorm.weight	768
bert.encoder.layer.7.output.LayerNorm.bias	768
bert.encoder.layer.8.attention.self.query.weight	589824
bert.encoder.layer.8.attention.self.query.bias	768
bert.encoder.layer.8.attention.self.key.weight	589824
bert.encoder.layer.8.attention.self.key.bias	768
bert.encoder.layer.8.attention.self.value.weight	589824
bert.encoder.layer.8.attention.self.value.bias	768
bert.encoder.layer.8.attention.output.dense.weight	589824
bert.encoder.layer.8.attention.output.dense.bias	768

bert.encoder.layer.8.attention.output.LayerNorm.weight	768
bert.encoder.layer.8.attention.output.LayerNorm.bias	768
bert.encoder.layer.8.intermediate.dense.weight	2359296
bert.encoder.layer.8.intermediate.dense.bias	3072
bert.encoder.layer.8.output.dense.weight	2359296
bert.encoder.layer.8.output.dense.bias	768
bert.encoder.layer.8.output.LayerNorm.weight	768
bert.encoder.layer.8.output.LayerNorm.bias	768
bert.encoder.layer.9.attention.self.query.weight	589824
bert.encoder.layer.9.attention.self.query.bias	768
bert.encoder.layer.9.attention.self.key.weight	589824
bert.encoder.layer.9.attention.self.key.bias	768
bert.encoder.layer.9.attention.self.value.weight	589824
bert.encoder.layer.9.attention.self.value.bias	768
bert.encoder.layer.9.attention.output.dense.weight	589824
bert.encoder.layer.9.attention.output.dense.bias	768
bert.encoder.layer.9.attention.output.LayerNorm.weight	768
bert.encoder.layer.9.attention.output.LayerNorm.bias	768
bert.encoder.layer.9.intermediate.dense.weight	2359296
bert.encoder.layer.9.intermediate.dense.bias	3072
bert.encoder.layer.9.output.dense.weight	2359296
bert.encoder.layer.9.output.dense.bias	768
bert.encoder.layer.9.output.LayerNorm.weight	768
bert.encoder.layer.9.output.LayerNorm.bias	768
bert.encoder.layer.10.attention.self.query.weight	589824
bert.encoder.layer.10.attention.self.query.bias	768
bert.encoder.layer.10.attention.self.key.weight	589824
bert.encoder.layer.10.attention.self.key.bias	768
bert.encoder.layer.10.attention.self.value.weight	589824
bert.encoder.layer.10.attention.self.value.bias	768
bert.encoder.layer.10.attention.output.dense.weight	589824
bert.encoder.layer.10.attention.output.dense.bias	768
bert.encoder.layer.10.attention.output.LayerNorm.weight	768
bert.encoder.layer.10.attention.output.LayerNorm.bias	768
bert.encoder.layer.10.intermediate.dense.weight	2359296
bert.encoder.layer.10.intermediate.dense.bias	3072
bert.encoder.layer.10.output.dense.weight	2359296
bert.encoder.layer.10.output.dense.bias	768
bert.encoder.layer.10.output.LayerNorm.weight	768
bert.encoder.layer.10.output.LayerNorm.bias	768
bert.encoder.layer.11.attention.self.query.weight	589824
bert.encoder.layer.11.attention.self.query.bias	768
bert.encoder.layer.11.attention.self.key.weight	589824
bert.encoder.layer.11.attention.self.key.bias	768

bert.encoder.layer.11.attention.self.value.weight	589824	
bert.encoder.layer.11.attention.self.value.bias	768	
bert.encoder.layer.11.attention.output.dense.weight	589824	
bert.encoder.layer.11.attention.output.dense.bias	768	
bert.encoder.layer.11.attention.output.LayerNorm.weight	768	
bert.encoder.layer.11.attention.output.LayerNorm.bias	768	
bert.encoder.layer.11.intermediate.dense.weight	2359296	
bert.encoder.layer.11.intermediate.dense.bias	3072	
bert.encoder.layer.11.output.dense.weight	2359296	
bert.encoder.layer.11.output.dense.bias	768	
bert.encoder.layer.11.output.LayerNorm.weight	768	
bert.encoder.layer.11.output.LayerNorm.bias	768	
bert.pooler.dense.weight	589824	
bert.pooler.dense.bias	768	
classifier.weight	1536	
classifier.bias	2	

+-----+-----+

Total Trainable Params: 109483778

Task 9: The fine-tuned model has been pushed to 🤗 : [Fine-Tuned Classification](#), [Fined-Tune-QA](#)

Comments and Discussion:

Performance

The model's performance is poor and the scores from the performance metrics is not up to the mark as:

- The model's Exact Match (EM) score of 7.48% indicates a relatively low accuracy in providing exact answers to questions in the dataset.
- The F1 score of 12.10% suggests a modest balance between precision and recall, but there is room for improvement.
- The model performs better on questions with answers (HasAns) with an EM score of 14.98% and an F1 score of 24.24%. This indicates some ability to answer questions when information is available.
- The model's performance on questions with no answers (NoAns) is notably lower, with an EM and F1 score of 0.00%. This suggests challenges in identifying unanswerable questions.
- The best exact match and F1 scores are relatively higher at 50.09%, indicating that the model can achieve better performance under certain conditions or thresholds.
- The METEOR score of 0.00098 is quite low, suggesting that the model may struggle with generating responses that closely align with human-like language and expression.
- The BLEU score of 0.00 indicates minimal overlap between the model's generated text and reference text.
- The ROUGE scores are also very low, suggesting limited similarity between the model's outputs and reference text.

Work Distribution

Name	Contribution
Aditi	Calculated number of parameters of selected model, Helped in Pre-training
Anavart	Fine-Tuned for Classification
Ayush	Fine-Tuned for Question & Answering, Calculated Metric Scores for Question-Answering, Performance Inference, Documentation
Mann	Pre-Training the BERT Model using MLM and NSP, Pushed model to HuggingFace, Documentation
Pinki	Pre-Training the Bert Model using MLM and NSP, Documentation
Ronak	Fine-Tuned for Question & Answering, Calculated metrics for the Classification task
Sakshi	Fine-Tuned for Question & Answering, Documentation
Vedang	Pre Training the BERT Model, Fine-tuning for classification, Error analysis and debugging, Handling distribution of tasks among members

