

GAI Project 3 PEFT on GLUE benchmarks

- Topic: PEFT on GLUE
- If you have any questions, please e-mail to nckudm@gmail.com (<mailto:nckudm@gmail.com>).

Scoring Criteria

GLUE website: <https://gluebenchmark.com/tasks> (<https://gluebenchmark.com/tasks>)

1. Data (Task1, 20 pts)

- You can choose any perfering dataset in GLUE benchmark to fine-tuning.
 1. Load the train and validation split of data (10 pts). You can use pandas, Dataset and Dataloader.
 2. **Select two datasets on GLUE benchmarks** (10 pts). Tokenize the text. You can design your own tokenizer or use any API (if available).

2. PEFT Validation (Task3, 10 pts):

1. Evaluate your model when you are training. (10 pts)

3. PEFT Models (Task2, 15 pts)

1. Model design (5 pts). You can use any Pretrained-Models for your base model
2. Train(finetune) the model (separately by PEFT methods) (10 pts). You should mark some of the parameter as trainable while others are untrainable.

Analysis (45 pts)

1. Model analysis (5 pts)

Include your model design and the process of loss reduction and the results of validation and testing (if available).

2. PEFT Discussion (25 pts)

What is the main feature of Bitfit? Why does it work? (10 pts) What are the best hyper-parameters(e.g. learning rate, batch size, warmup ...) you find? Comparing with full-finetuning, why do we need larger/smaller learning rate? (10 pts)

Baselines :

dataset	metrics	baseline
CoLA	Matthew's Corr	0.6
SST2	Accuracy	0.88
MRPC	Accuracy	0.8
STSB	Pearson-Spearman Corr	0.8
QQP	F1 / Accuracy	0.8/0.8
MNLI_Matched	Accuracy	0.8
MNLI_Mismatched	Accuracy	0.8
QNLI	Accuracy	0.85
RTE	Accuracy	0.7
WNLI	Accuracy	0.8

3. PEFT Comparison (25 pts)

Compare the performance of Bitfit with Lora on the two datasets you select. (10 pts)

How does the rank(r) affect LoRA's parameter count and performance? (10 pts)

Submission

- Structure
 - You should submit a .zip file with the name {student_id}_GAI_Project3 (eg. F1234567_GAI_Project3). It should be unzipped into a directory with the same name, and the directory structure should be:

```
{student_id}_GAI_Project3
├─ main.py (.ipynb)  // the code you use to run the language models & generate figur
...                // include other scripts if you have more
├─ requirements.txt // pip freeze > requirements.txt
└─ report.pdf      // your report file, be sure it is .pdf
```

- TA will not run your code for this project, but please make sure that you hand in the code that train the model(s) and executes the generation.
 - Make it readable **with comments**, lest we would need to refer to it under any circumstances.

- If your code does not look like it can reproduce the results described in your report, we would consider a grade discount/ask you for a demo.
- Submission Deadline: 5/16 (Thursday) 9:00 a.m.
 - Note that the deadline is 9:00 **in the morning**.
 - Late submission within 1 week will get a 10% discount, and 3 weeks will get a 30% discount.
 - Submissions later than that will not be graded and you get a 0 as a result.

Appendix

We recommend you to run your code on Colab or Kaggle. Of course, if you have your own hardware resources, you can use them as well.