

CS 613: NLP

Assignment 3: Pretraining and fine-tuning an LLM

Total marks: 100 Pts.	Submission deadline: 23:59:59 Hrs, November 19, 2023 (Sunday)
-----------------------	--

Assignment Instructions

- Regarding the late submission, we will be following the penalty as per the table:

Late Submission	Penalty (Out of 100)
Till 1-hour past the deadline	5 points
1 to 12 hours past the deadline	10 points
12 to 24 hours past the deadline	20 points
24 to 36 hours past the deadline	40 points
36+ hours past the deadline	100 points

- We will follow the zero plagiarism policy, and any act of plagiarism will result in a zero score for the assignment.
- Please cite and mention others' work and give credit wherever possible.
- If you seek help and discuss it with the stakeholders or individuals, please ask their permission to mention it in the report/submission.
- GPU requirement: Use Colab and write the answers in the colab itself.

Problem Statement

Pretrain an LM and fine-tune the model for a task.

Tasks (100 Pts.)

- Select the Bert-base-uncased [[code](#), [paper](#)] model.
- Calculate the [number of parameters](#) of the selected model from the [code](#). Does your calculated parameters matches with the parameters reported in the respective paper. **[05 pts]**
- Pretrain the selected model on the **train split** of '[wikitext-2-raw-v1](#)'. **[10 pts]**
For 5 epochs. Use the hyperparameters as per your choice.
- Compute and report the Perplexity scores using the [inbuilt function](#) on the test split of '[wikitext-2-raw-v1](#)' for each epoch. Do scores decrease after every epoch? Why and why not? **[10 pts]**

5. Push the pre-trained model to 😊. [2.5 pts]

This will help you understand how to share your trained model with the community [[A good reference to help!](#)]

6. Fine-tune the **final pretrained model** on the following three tasks:

- a. Classification: [SST-2](#) [10 pts]
- b. Question-Answering: [SQuAD](#) [10 pts]

Train-test split should be 80:20, use [random/stratify](#) sampling and seed as 1. Fine-tuning should be performed on the Train split.

7. Calculate the scores for the following metrics on the test splits. Note that metrics depend on the selected task:

- a. Classification: Accuracy, Precision, Recall, F1 [17.5 pts]
- b. Question-Answering: squad_v2, F1, METEOR, BLEU, ROUGE, exact-match [[Read this!](#), and [this too!](#)] [17.5 pts]

8. Calculate the number of parameters in the model after fine-tuning. Does it remain the same as the pre-trained model? [05 pts]

9. Push the fine-tuned model to 😊. [2.5 pts]

10. Write appropriate comments and rationale behind:

- a. Poor/good performance. [5 pts]
- b. Understanding from the number of parameters between pretraining and fine-tuning of the model. [5 pts]

Feel free to download the pretrained models.

Points Split: 5+10+10+2.5+10+10+17.5+17.5+5+2.5+5+5 = 100

Submission

1. Submit your code (GitHub) or colab notebook with proper comments to [this link](#).
 - a. Ensure the individual contribution is appropriately added (OTHERWISE PENALTY OF 10 MARKS).
 - b. **Please push the pretrained/fine-tuned models to 😊. Add a final section where you pull the pre-train/fine-tuned model (from hugging face) and compute the metrics over the public models. This will be used for evaluation—negative marks (-10 Marks) as there will be verification on the trained models.**

Expectations from the team:

1. Properly divide the team into sub-groups and distribute your tasks equally.
2. Negative marks for documentation and justifications!
3. Write the contributions or tasks completed by each team member. Scores might be different among team members if the tasks are not equally distributed.

TAs to Contact

1. Himanshu Beniwal

2. Progyan Das
3. Disha Suthar
4. Jayesh Malaviya
5. Shiraj Sawant

References

If required, please feel free to take help from the following references:

1. [[HF's Notebooks](#)] Huggingface's notebooks
2. [[Stackoverflow](#)] Number of parameters
3. [[perplexity on SO](#)] Perplexity
4. [[transformers-util](#)] If low on resources.
5. [[SO](#)] Adding layers on top of bert.

FAQs

1. *We will add clarifications to doubts here. Please check periodically, as someone might have already asked the doubt and it would be appended here.*