OUR Project Parameter-Efficient Training through Efficient Joint Sparse and Low-Rank Adaptation

Project Title

Parameter-Efficient Training through Efficient Joint Sparse and Low-Rank Adaptation

Project Description

With the advent of deep learning and large language models, which have delivered impressive results for a large number of machine learning tasks, models with hundred of millions, billions or more parameters have been become main stream. While the hardware and energy requirements of a full-training process prevent state-of-the-art deep learning models to be trained on consumer hardware, it is possible to “optimize” a pre-trained model to excel for a particular task on consumer hardware via Parameter-Efficient Fine-Tuning (PEFT).

Among the most popular techniques for PEFT, low-rank adaptation and sparse adaptation has merged in the last two years. The recent paper “RoSA: Accurate Parameter-Efficient Fine-Tuning via Robust Adaptation” (https://arxiv.org/pdf/2401.04679) has shown that the performance of methods that jointly optimize sparse and low-rank adapters outperform either of the two pure methods. In this projcet, we build on recent advances on low-rank fine-tuning and training within our research group based on quadratic differentiable rank regularizers and plan to develop algorithms for PEFT that obtain improved performance over RoSA and other PEFT methods given the same parameter budget.

Number of Openings

2

Requirements

A good grasp of calculus, linear algebra and a strong interest in machine learning together with good coding skills (Python) are a requirement of this project. Furthermore, familiarity with one at least one of the following areas is required: - Continuous optimization algorithms - Sparse models in machine learning and/or statistics (e.g., Lasso, sparse logistic regression, etc.) - Low-rank models in machine learning and/or statistics (e.g., LoRA, (robust/sparse) principal component analysis, low-rank matrix completion, singular value decomposition) - Iteratively reweighted least squares algorithms

Preferences

Familiarity and knowledge with more than one of the areas listed under “Requirements” is preferred.

Training

No specific requirements, training takes places during the research project meetings and in discussions with the faculty and graduate student mentor.

Anticipated Student Learning Outcomes

- Familiarization with the state-of-the-art in finetuning of large machine learning models/ LLMs, which is an extremely popular technique in modern AI
- Experience with the development of new deep learning models
- Experience in contributing to ongoing AI/machine learning research
- If successful, co-authorship in a conference publication at a top-tier AI/ML conference can arise from this project.

Mentoring Plan

The student will be guided through the research process in weekly meetings with the mentoring faculty as well as in communication with a Ph.D. student, who serves as a co-mentor. Communication will occur additionally asynchronously through a Slack channel, during which smaller questions can be addressed. If the research project turns out to be successful, mentoring in academic writing and the preparation of research manuscripts will be provided.

Conflicts

Availability on Wednesdays for weekly project meetings with faculty and Ph.D. student mentor.