In-context learning vs few-shot learning: any difference?

Supervisor(s):

Yingzhen Li (Department of Computing, yingzhen.li@imperial.ac.uk)

Project description:

Few-shot learning (or meta-learning) trains a machine learning model with a dataset of "context dataset" (which contains a few datapoints of (x,y) pairs), so that the model, given new context data in deployment time, quickly learns to produce predictions for future queries based on the context data. Few-shot learning is a well-established research field, and models based on neural processes and Transformers have been shown to be successful in this regard. On the other hand, recent empirical studies in large language models (LLMs) also seem to suggest their capability of so called "in-context learning". In-context learning has the same goal as few-shot learning: to enable models for quick learning of the context data in deployment time for responding future queries. However, the context and the query are not structured in a regular dataset/datapoint form, rather they are presented as a sequence, e.g., with language modelling, they would be presented as "China's capital is Beijing. The US's capital is Washington DC. The UK's capital is [to be answered]." The principles of few-shot learning and in-context learning are very similar, but it remains an open question about why in-context learning is possible, and what are the additional properties of in-context learning as compared with few-shot learning.

This project will investigate the similarities and differences between in-context learning and few-shot learning, under a general machine learning setting beyond text data. In particular, we will spend time on the following studies:
1. Build simulated datasets (e.g., "polygon shapes" and Raven's Progressive Matrices) to accommodate the test of both few-shot learning and in-context learning. We will mimic the properties of language modelling but also ensure a clear separation between meta-training and meta-testing and avoid data leakage, to eliminate memorisation as a potential explanation for LLM's in-context learning ability.
2. Evaluate Neural processes and Transformers' ability for both few-shot learning and in-context learning on the simulated dataset. Specifically, by comparing their performances, this will determine the difficulty level comparison between few-shot learning and in-context learning, and it will also reveal the additional assumptions (if any) needed for in-context learning to succeed.
3. Evaluate Neural processes and Transformers' ability for both few-shot learning and in-context learning on real-world datasets, to test if the conclusions from step 2 carries over to realistic scenarios.

Timeline (tentative):

Oct 2024 - Dec 2024: Literature review and reproduction of selected papers.
Jan 2024 - Mar 2024: Initial experiments on the simulated "polygon shapes" datasets.
Apr 2024 - July 2024: Further experiments on abstract visual reasoning tasks (e.g., Raven's Progressive Matrices).
Aug 2024: Finishing experiments and thesis writing.

Minimum viable thesis:

Description of the simulated datasets and evaluations on them. Documenting down the experimental difference between in-context learning setting and few-shot learning setting by using Transformers.

Required background & skills:

Student suitable for this project would have strong mathematical analysis skills. They should feel very comfortable in derivations with basic probability & statistics, linear algebra, and calculus. They should also have experience with existing deep learning frameworks (e.g. Tensorflow or Pytorch). Having hands-on experience with Transformers and/or Few-shot learning will be a plus factor.

Representative References:

Brown et al. Language Models are Few-Shot Learners. arXiv:2005.14165
Jha et al. The Neural Process Family: Survey, Applications and Perspectives. arXiv:2209.00517
Fin et al. Model-Agnostic Meta-Learning for Fast Adaptation of Deep Networks. ICML 2017
Aykurek et al. What learning algorithm is in-context learning? Investigations with linear models. ICLR 2023
Barrett et al. Measuring abstract reasoning in neural networks. ICML 2018