PRML-2023Fall 期末论文选题&注意事项

期末论文占总成绩的40%。

支持**1 - 2人组队**，评分标准主要依据论文质量（可以是有趣的题目上的尝试或者完成度很高的项目），根据组员工作量微调。

DDL: **2024/1/3 23:59**

我们为同学们提供一些可能的选题，主要涉及语言模型、视觉和图神经网络，这些题目有些可能比较困难，**如果选择助教提供的题目，请提前联系助教**贺正夫，我们会根据任务提前沟通，不论选题难易，我们希望评分标准在工作量和智力活动上尽可能公平，即我们不会因为在困难的问题上进展较少而做出惩罚（因此，即使尝试失败了也应当记录并分析）。如果同学们有自己感兴趣的选题，也请**登记在**这个文档里，并附上100字以内的简介。

我们会为同学们提供共8张3090显卡的计算资源，同学们也可以使用自己的计算资源。登录方式（需要连ifudan wifi或开easyconnect VPN）：ssh -p 20000 root@10.176.52.114 -i <path_to_id_rsa_file>，密钥是一个id_rsa.txt文件，助教在群里公布。

资源使用方法和规则：下周作业正式开始后更新。

语言模型

1. 自回归语言模型知道汤姆克鲁斯的妈妈是谁，但是却不知道这位女士最有名的儿子叫什么。截止出题这一天，世界上最强的语言模型GPT4也在这些特定的问题上几乎全军覆没。这说明愚蠢的Transformer学习了A和B的某种双射关系，却并不能从一端反推到另一端。这个问题或许很难定论为什么，但是有人认为Bert作为双向语言模型可以缓解这一问题。你怎么想？你能否在（一个比较大的）BERT上复现这些实验，来佐证你的观点？

   https://arxiv.org/abs/2309.12288

2. 自回归语言模型在学习算术任务的过程中，被发现一种“顿悟”现象，模型的训练loss降到很低之后，test loss仍然极高，就这样继续训下去，你不会从这两个指标中看到什么端倪，但是突然test loss就骤降了。你能否在模加法这一任务上 (a + b = c (mod p))对p = 53复现这个实验？（这种现象不是一定会出现的，比较需要调参）

https://arxiv.org/abs/2201.02177

3. MOSS是这样被训练出来的，我们在大规模语料上预训练一个模型，让它具有“预测下一个词”的能力。然后我们用很多的指令数据（例如“<用户>: 说出三个F开头的国家. <MOSS>: France, Finland, Fiji”）和多轮对话数据(例如"<用户>: 你觉得我这件衣服好看嘛？ <MOSS>: 作为一个语言模型，我没有主观blah blah blah <用户>: 你真没用。<MOSS>: 对不起，我没能帮到你blah blah blah ")对模型进行有监督微调，一般而言，我们理解预训练阶段模型获取很多知识，在有监督微调阶段模型可以学习到阅读指令和礼貌地回复，学会后者对前者是有要求的。出于好奇，我们想知道如果一个很小的模型（比如GPT2-Small）用指令数据微调，他能学会什么，包括模型训练&行为分析。

   https://arxiv.org/abs/2203.02155

4. Transformer内部有很多可以解释的feature，更多的是不能解释的feature。具体而言，我们重点关注GPT2-Small中的MLP，你能否在任意一层找到任意一个neuron，使得其对某个类型的token作出响应，你需要从两个角度评估这种响应：专一性和敏感性，即这个neuron被激活时，当前的token是否总具有你猜测的性质？以及当具有这种性质的token出现时，这个neuron是否总被激活？

5. Transformer很伟大，有很多人说要取代他成为语言模型的基础设施，但是它的地位目前来看很难撼动！一些不错的尝试包括State Space Models, H3, Hyena, RWKV, RetNet, Monarch等等等等等。用你的fdunn实现其中的一个的小版本模型，并与torch结果对照，验证其正确性。

   https://arxiv.org/abs/2305.13048

   https://arxiv.org/abs/2307.08621

6. 其他经过讨论认为是语言模型范畴内的题目