



一些 AI/ML 学习中重要却被忽视的问题

关于《机器学习》课程学习与能力的底层逻辑



以下问题定语可以加上 AI 或 ML 但答案逻辑其实适用所有工科学科

- 企业靠什么赚钱？实验室靠什么申请经费？商业计划书？科研项目申请书？开题报告？
- 企业、高校导师招聘招生期望的是什么？他们希望你们**简历**上看到什么？为什么企业和导师不只看你们的考试成绩招聘？**复试/面试**他们在考查什么？
- 如何给工程师定薪？如何给自己找工作定位？
- 如何达成上述目标/期望？获得分数和获得能力的差别在哪里？为什么动手很重要？为什么细节很重要？
- 如何选适合自己的教材？教材章节和未来查找最新学术文献联系和区别？如何阅读最新学术文献？
- 如何寻求他人（**大模型**、老师、助教、同学、领导、同事）帮助？如何沟通**交流**技术问题？



企业靠什么赚钱？实验室靠什么申请经费？ 商业计划书？科研项目申请书？开题报告？

- 通用人工智能非常重要(意义)。国内外许多企业都在做(进展)，但他们的研究都遇训练推理成本高昂的困难(问题)。Deepseek 通过架构创新(MoE+MLA)、训练方法突破(GRPO+混合精度)、低成本优化(数据蒸馏+软硬件协同)及开源生态(创新点)。
- (别人和我)的研究证明我的想法是可行的(可行性分析)，而且我以前做研究也还可以(历史业绩)，现在条件不错(足够的数据、算力、人才)，大概花上几个亿(研发经费，人员配置)，我就能让公司估值翻 10 倍。通过极致的Infra优化(如H800集群算力利用率提升至行业顶尖水平)，降低单位token成本，实现高达70%-85%的模型利润率。
- 什么东西非常重要(意义)。国内外许多学者都在做(进展)，但他们的研究都遇到一个困难(问题)。我有一个思路(创新点)有望解决这个问题。一旦这个问题解决了，那意义大的不得了(本项目意义)。
- 我要解决的这个问题是…(研究目标)，具体要研究…(内容)，其中最核心的是…(科学问题)。方案是这样的…(技术路线)
- (别人和我)的研究证明我的想法是可行的(可行性分析)，而且我以前做研究也还可以(研究基础)，现在条件不错(工作条件)，大概花上几十万(经费预算)，我就能为科学(国家)做点贡献。



企业、高校导师招聘招生期望的是什么？他们希望你们简历上看到什么？为什么企业和导师不只看你们的考试成绩招聘？**复试/面试**他们在考查什么？

- 选人标准差异极小，不存在你能找到同等企业工作，却无法申请/考入到同等学校升学资格的情况
- Google 将 NASA 视为重要的人才竞争对手，其核心逻辑在于 “争夺顶尖人才的机构就是最大竞争对手”
- 低难度的重复工作没有意义，无法形成创新
 - 顶级期刊会议论文一作+++
 - 知名Github 开源项目日常 commit 记录++
 - 高额奖金竞赛获奖+获奖技术方案分享++
 - 技术博客、专业知识分享、论文解读+
 - 一般期刊会议论文一作+
 - 人人都会干的项目但充满细节有改进尝试+
 - 人人都会干的项目/实验-
 - 简历上除了空话啥也没有---
 - 内容造假一问三不知或忘了（**企业拉黑、研究生复试不及格**）---

如何给工程师定薪？如何给自己找工作定位？



廣東工業大學
GUANGDONG UNIVERSITY OF TECHNOLOGY

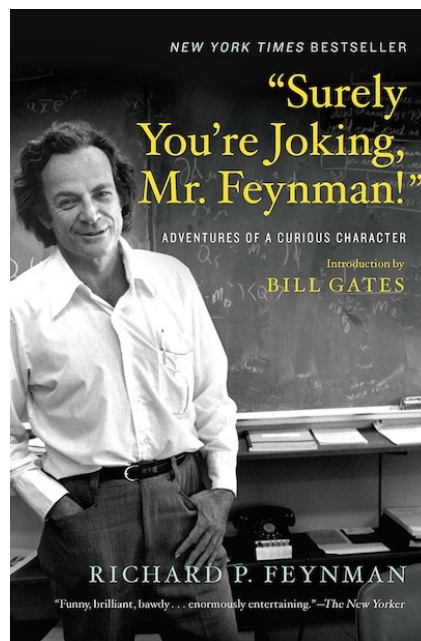
- 你专业方向的顶级期刊会议论文（如 CCF-A 类）未开源
- 7 天可以进行开源复现 50 万年薪以上
- 1 个月可以进行开源复现 30 万年薪以上
- 3-6个月可以进行开源复现 18 万年薪以上



如何达成上述目标/期望？获得分数和获得能力的差别在哪里？为什么动手很重要？为什么细节很重要？

- **Richard Feynman: What I cannot **create**, I do not **understand****

- 明确概念（Concept）
- 模拟教学（Teach）
- 回顾修正（Review）
- 简化输出（Simplify）



- 用 C++/CUDA实现卷积神经网络 CNN？
- 古代人如何测量地球半径？

- 简化特例、枚举
- 推理、总结规律
- 抽象一般化
- 修正完善

- 两人 7 局 4 胜，有多少种胜负比赛情况？
- 40个人围成一圈，从第一个人开始报数，每轮报奇数的人将被淘汰，直到只剩一个人。求最后一个人一开始站第几位？



如何选适合自己的教材？教材章节和未来查找最新学术文献联系和区别？如何阅读最新学术文献？

- 细节越多越能帮助你理解
- 教材往往是一系列论文的合集，为了帮助学生阅读，不断增加细节和先验知识，survey <讲义 <教材
- 不假设读者有先验知识>有代码=同一内容厚的>简洁、薄的
- Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton. "Imagenet classification with deep convolutional neural networks." *Advances in neural information processing systems* 25 (2012).

3.5 Overall Architecture

Now we are ready to describe the overall architecture of our CNN. As depicted in Figure 2, the net contains eight layers with weights; the first five are convolutional and the remaining three are fully-connected. The output of the last fully-connected layer is fed to a 1000-way softmax which produces a distribution over the 1000 class labels. Our network maximizes the multinomial logistic regression objective, which is equivalent to maximizing the average across training cases of the log-probability of the correct label under the prediction distribution.

The kernels of the second, fourth, and fifth convolutional layers are connected only to those kernel maps in the previous layer which reside on the same GPU (see Figure 2). The kernels of the third convolutional layer are connected to all kernel maps in the second layer. The neurons in the fully-connected layers are connected to all neurons in the previous layer. Response-normalization layers follow the first and second convolutional layers. Max-pooling layers, of the kind described in Section 3.4, follow both response-normalization layers as well as the fifth convolutional layer. The ReLU non-linearity is applied to the output of every convolutional and fully-connected layer.

The first convolutional layer filters the $224 \times 224 \times 3$ input image with 96 kernels of size $11 \times 11 \times 3$ with a stride of 4 pixels (this is the distance between the receptive field centers of neighboring

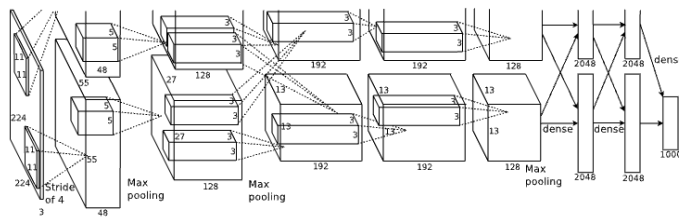


Figure 2: An illustration of the architecture of our CNN, explicitly showing the delineation of responsibilities between the two GPUs. One GPU runs the layer-parts at the top of the figure while the other runs the layer-parts at the bottom. The GPUs communicate only at certain layers. The network's input is 150,528-dimensional, and the number of neurons in the network's remaining layers is given by 253,440–186,624–64,896–64,896–43,264–4096–4096–1000.

neurons in a kernel map). The second convolutional layer takes as input the (response-normalized and pooled) output of the first convolutional layer and filters it with 256 kernels of size $5 \times 5 \times 48$. The third, fourth, and fifth convolutional layers are connected to one another without any intervening pooling or normalization layers. The third convolutional layer has 384 kernels of size $3 \times 3 \times 256$ connected to the (normalized, pooled) outputs of the second convolutional layer. The fourth convolutional layer has 384 kernels of size $3 \times 3 \times 192$, and the fifth convolutional layer has 256 kernels of size $3 \times 3 \times 192$. The fully-connected layers have 4096 neurons each.



如何寻求他人（大模型、老师、助教、同学、领导、同事）帮助？如何沟通交流技术问题？

- 学会提问：How
- 交流采用总分结构，先一个短句子总起。
- 总：我在完成aa 作业的时候遇到了bb 的问题（学会总结）
 - 1. 我的思路是cc，预期结果是dd（总起的进一步展开）
 - 2. 我试着搜索了关键字ee/询问了大模型 ee，进行了操作ff 但没有什么收获（学会自己先尝试）
 - 3. 下面的图片gg 是可能有用的报错信息和我的一些想法（学会截图）
 - 4. 请问我可能哪里出了问题？（最后才是一个问句）



以为自己掌握和实际动手存在巨大差距
实际动手并与他人交流是学习的捷径，过拟合题库不是

“在群里提问有点丢脸”

学习不是什么羞耻的事情

“反正 × 助教人好，我就私戳 ta 了”

助教会累，还会秃头

“好挫败，怎么全是 Bug”

搞定它，你就能变强

“群友好卷”

1. 与自我和解
2. 转换思路：卷王是我的学习资源

