

模式识别与计算机视觉：第二次作业

2025 年 3 月 25 日

注意事项

1. 请务必认真阅读所有注意事项。
2. 本作业发布时间 2024.3.25，交作业时间：2024 年 4 月 10 日上午 9:00。此时间之后的提交不再接收，成绩以 0 分计。如确有特殊原因（例如因公出差），请**提前**向任课教师请假，提交相应证明材料后另行安排；如有紧急医疗需求等不可预知的特殊情况，需事后尽早提交正式医院证明等相关证明材料。
3. 请手写或通过 Word/LaTeX 等软件记录答案，回答尽量简洁，一般每次作业的答案（只要答案，不要抄写题目）不超过 3 页为佳。
4. 手写答案的同学可以用拍照、扫描等方式提交电子版，但在保证内容清晰可见的前提下尽量减少文件大小。如果文件超过 1 个，压缩为**单个**文件上传。
5. 请在每次作业的开始部分写上姓名、学号、所属院系。缺少信息的，本次作业总分扣除 10 分。请注意：只有在正式选课名单上的同学，作业才会被批改并计算分数。
6. 建议作业完成后、交作业之前自行拍照或扫描并妥善保存，以备特殊情况时使用（例如认为自己已经交作业了，但系统中没有）。例如，如系统发生错误，可以提供照片或扫描文件以作**证明**。
7. 作业提交通过教学立方进行，请务必在教学立方中注册本课程。

1 习题一（15 分，每小题 5 分）

教材第三章的习题 3.2。

2 习题二 (18 分, 每小题 2 分)

教材第四章的习题 4.2。

3 习题三 (12 分, 每小题 3 分)

教材第四章的习题 4.5。

4 习题四 (10 分)

教材第四章的习题 5.2。你可以根据系统中提供的 Matlab 代码, 自己用 Python 重写相关代码, 以完成本习题。

5 习题五 (15 分, 每小题 3 分)

教材第五章的习题 5.4。

6 习题六 (12 分, 每小题 4 分)

教材第六章的习题 6.3。

7 习题七 (18 分, 每小题 9 分)

模型和数据准备: 从网页https://huggingface.co/timm/vit_tiny_patch16_224.augreg_in21k 下载 ViT-Tiny 模型, 然后从网页https://huggingface.co/datasets/student/CUB_birds_200_2011 下载 CUB 数据集。在 CUB 数据集的 200 个类中, 从每个类的**训练集**中**随机抽取**1张图片, 组成我们将要使用的数据集 \mathcal{S} 。请注意:

- 本题需要一些深度学习的基础, 如不具备相关知识, 请先行进行自学。
- 本题需要事先安装 PyTorch、CUDA 驱动等, 请自行解决 (例如可通过遵循在线 tutorial 尝试安装、请教同学等)。
- 在准备数据集 \mathcal{S} 时, 使用你的学号做为随机数, 从每个类中随机抽取 1 张图像。因此, 每位同学使用的数据集 \mathcal{S} 是**不一样**的, 作业各题运行得到的结果也**不一样**。

- 如果没有条件使用 GPU 卡进行运算，可以使用 CPU 进行计算。这需要在代码中指定。因为本题所需的计算量较小，使用 CPU 计算也可以完成。

本题需要完成下述（简单的）任务。

1. 针对 \mathcal{S} 中的每一张图像，可以将其输入 ViT-Tiny 模型进行前向计算，直至计算得到最后一层的 **CLS token**。**独立**写代码获得集合 \mathcal{S} 中所有图像的 CLS token，组成特征集合 \mathcal{F} 。这里“独立”的意思是不要使用网络上现有的代码，但你可以在学习网络代码后自己重写。本小题需要上传你的上述代码文件，在压缩包中命名为 `extract-cls.py`。
2. 集合 \mathcal{F} 应该包括 200 个向量，每个向量为 d 维（ViT 模型的基础维数）。对其做 PCA 运算，当保存 90% 方差时，保留下来的维度有多少个？相比 d ，占据了多少百分比？
3. （本小题不计分，但**非常鼓励**感兴趣的同学进行尝试）从上述的实验中，你应该能发现：深度学习模型产生的特征（通常称为 activations）实际上是低秩的，这个现象可以用来压缩深度学习模型。相关论文可以从网页<https://ojs.aaai.org/index.php/AAAI/article/view/26304> 下载。你可以参考论文，在上述作业的基础上，自己实现论文中算法的基础部分，并进行实验。