

模式识别与计算机视觉：第一次作业

2025 年 3 月 4 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 2024.3.4, 交作业时间: 2024 年 3 月 20 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 (16 分, 每小题 2 分)

教材第一章的习题 1.1。

2 习题二 (14 分, 每小题 7 分)

若 $X \sim \mathcal{N}(0, 1)$, 证明以下不等式:

(a). 对于任意 $\epsilon > 0$, 有

$$P(X \geq \epsilon) \leq \frac{1}{2}e^{-\epsilon^2/2}.$$

(b). 对于任意 $\epsilon > 0$, 有

$$P(|X| \geq \epsilon) \leq \min \left\{ 1, \sqrt{\frac{2}{\pi}} \frac{e^{-\epsilon^2/2}}{\epsilon} \right\}.$$

提示: 对于 $\mathcal{N}(0, 1)$ 的概率密度函数 $f(x)$, 有 $f'(x) = -xf(x)$.

3 习题三 (25 分, 每小题 5 分)

本题涉及深度学习中的全连接层 (fully connected layer, 或简称为 FC 层) 以及 BN 层 (batch normalization layer) .

- (a). 自行寻找资料, 了解关于 FC 层和 BN 层的知识, 并用**简短**的语句分别描述两个层操作的数学含义.
- (b). 当两个 FC 层直接相连 (即两者中间不包含其他层) 时, 证明在神经网络的前向计算中, 这两个层可以合并为一个 FC 层.
- (c). 这种合并在什么情况下、能够带来什么样的优势?
- (d). 这种合并是否总是能带来优势? 如果答案是否的话, 举出一种情况, 这种连续两个 FC 层合并为一个 FC 层的操作, 反而会带来劣势.
- (e). 当一个 FC 层后面紧跟的是 BN 层时, 证明在神经网络的前向计算中, 可以将这个 FC+BN 层的两层组合替换为一个单独的 FC 层. 这种替换在什么情况下、能够带来什么样的优势?
- (f). 本小题不计分, 但鼓励同学们自己动手, 推导一下在 ResNet50 中, 当卷积层后面紧跟的是 BN 层时, 如何将两者 (在推理时) 替换为一个单独的 FC 层; 并进一步测试模型的运行速度, 获得这种替换对模型加速是否有作用、有多大作用的第一手经验.

4 习题四 (12 分, 每小题 3 分)

在教材第三章中, 我们了解到细节问题 (p43) 对设计一个模式识别系统的影响。现在我们将探讨如何解决以下细节问题 (以教材中的人脸识别为例)

- a) 假设存储在设备中的人脸图像是 100×100 的分辨率, 即 $\mathbf{x} \in \mathbb{R}^{10000}$, 而设备将你的照片拍成 400×400 。请写出两种不同的预处理方式, 使得你的照片能和设备中的照片正常匹配。
- b) 我们假设一共有 n 张照片, 且将每张存储的照片看作一个 100×100 的矩阵。已知两两不相交的 2×2 的像素格内都具有相似的像素值, 如下矩阵示意:

$$\begin{bmatrix} 1 & 1 & 155 & 156 & \dots \\ 1 & 1 & 154 & 155 & \dots \\ 50 & 51 & 254 & 253 & \dots \\ 49 & 50 & 255 & 255 & \dots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix} \quad (1)$$

你有什么办法能降低存储照片的容量开销吗? 存储开销能降低多少?

- c) 教材中提到了不平衡二分类问题 (p46)。我们假设训练集中: A 类有 9900 个样本, B 类有 100 个样本。测试集中: A 类有 5000 个样本, B 类有 5000 个样本。如果我们学习到一个映射 $f(\cdot)$, 它将所有输入的样本都预测为 A 类, 那么我们在训练集上的准确率 acc_{train} 是多少? 在测试集上的准确率 acc_{test} 是多少?
- d) 你可能已经知道计算准确率有两种不同的计算方法: micro 和 macro。请简要描述评价指标计算方法中 micro 和 macro 两种计算方式的区别? 在 c) 中我们计算准确率用到的是 micro 还是 macro 的计算方式? 如果不了解这两者的区别, 请搜索网上资源, 自行了解他们的区别。
- e) 上述问题实际上描述的是一个长尾识别问题 (long-tailed recognition problem)。在这种问题下, 我们在训练集上应当采取哪种计算方式来评估准确率? 请设计一种针对此问题的训练方法, 使得训练集中样本量少的类别 B 能够在测试集上减少误判? 此处只需描述主要思路即可, 无需提供技术细节。

5 习题五 (12 分, 每小题 3 分)

我们考虑近邻分类器问题。给定一个包含 8 个样本的训练集 $S = \{\mathbf{x}_1, \dots, \mathbf{x}_8\}$, 其中 $\mathbf{x}_1 = (0, 0)$, $\mathbf{x}_2 = (0, 1)$, $\mathbf{x}_3 = (0, -1)$, $\mathbf{x}_4 = (-1, 0)$, $\mathbf{x}_5 = (1, 0)$, $\mathbf{x}_6 = (8, 0)$, $\mathbf{x}_7 = (8, 1)$, $\mathbf{x}_8 = (9, 0)$ 。它们的类别分别是 (A, A, A, A, A, B, A, B)

- a) 对于两个测试样本 $\mathbf{z}_1 = (0, -2)$, $\mathbf{z}_2 = (8, 2)$, 运用最近邻分类器 (1-NN), 得到这两个样本的分类结果是什么?
- b) 同样的两个样本 $\mathbf{z}_1, \mathbf{z}_2$, 运用近邻分类器 k-NN, 取 $k=3$ 。得到的两个样本的分类结果是什么?
- c) 分析两次结果不同的原因?
- d) \mathbf{x}_7 是否可能属于类别 B ? 在此情况下 k-NN 相比 1-NN 的优势在何处?

6 习题六 (21 分, 3—8 功能及感想每部分 3 分)

本题为一道编程题: 从零开始构建一个机器学习系统, 请参见 ‘main.ipynb’ 文件中的提示来完成相关的代码 (请自行安装 Jupyter Notebook)。这份工程的功能包括:

- 1. 常见的机器学习数据集的读取过程 (已提供)
- 2. 训练和验证集的划分 (已提供)
- 3. 实现一个 KNN 分类器 (需完成)
- 4. 实现评估指标-准确率 (需完成)
- 5. 根据验证集进行超参数选择 (需完成)
- 6. 实现 5 折交叉验证并进行超参数选择 (需完成)
- 7. 最终确定超参数之后, 完成在测试集上的测试 (需完成)
- 8. 针对不平衡数据集, 实现 precision, recall 和 F1 score 的计算 (需完成)

在完成代码后, 提交时需要 notebook 文件 (包括代码和中间输出结果, notebook 可直接输出成 pdf 或 html), 并谈谈你在这次编程的感想 (可以包括你遇到的问题、收获等等)。