

2022-2023 秋季《机器学习方法基础》课程第 2 次作业

本次作业参考资料为：

书籍《Statistical Analysis in Forensic Science》、文献《似然比方法体系在法庭科学中的研究进展》。fig1、fig2 为似然比中的 Neyman-Pearson 引理及其证明。

作业中的第 1、2 题要求在 **10 月 23 日前**上交；第 3 题有一定难度，请同学们合理规划时间，**10 月 30 日前**上交即可。**作业格式要求与第一次一致。**

1. 在一个钓鱼问题中（数据 fishing.xlsx），想知道具备怎么样的条件可以钓到鱼。我们的数据分为训练集和测试集（由标签变量定义），已在训练集上建立了 2 个模型，分别获得了测试集和训练集的预测评分（ $E(Y|X)$ 的估计），要求编写程序，实现以下内容分析，并作分析报告：

- a. 对训练数据和测试数据分别对模型 1 和模型 2 制作 P-R 图，ROC 曲线图，用经验 AUC 估算面积判断整体预测效果。如果按照实验结果给出一个符合实际的预测阈值，请说明这个阈值选取的依据。
- b. 在给出的阈值上，输出混淆矩阵。

2. 阅读文献《似然比方法体系在法庭科学中的研究进展》。论文介绍了两种似然比的计算方法以及似然比模型的几种评价方法。针对文章撰写一篇小报告（300 字以上），报告中至少回答如下问题：

- a. 谈谈你对两种似然比的计算方法的理解，计算公式的实际含义是什么？
- b. 常用的似然比评价图形有直方图、Tippett 图、DET 图和 ECE 图。如何根据这 4 种图像的特征评价似然比方法的效果好坏？

3. 《Statistical Analysis in Forensic Science》4.1-4.2 节中的内容为如何计算似然比。学习该部分内容，并编写程序实现 4.2.1 中的似然比计算方法，用你编写的程序计算墨迹鉴定数据（第一次作业中的数据）的似然比。