

# 模式识别课程大作业

## 作业要求：

作业提交形式为 PDF 文件，文件命名方式为“学号-姓名-大作业.pdf”，作业可选用一下三种形式书写：(1) LaTeX 编译 PDF 文件(可自定模板，A4 页面，主体部分为小四号字体)；(2) Word 文档书写保存为 PDF 文件；(3) 在纸上书写拍照或者扫描保存为 PDF 文件(请注意书写工整，笔迹清晰)。建议使用 Python、MATLAB、Excel 等工具作图更加方便美观。编程题需要提交源码，并且指出运行环境，源码中请提供简单注释。

7 月 1 日以前[提交到课程网站](#)。

1. 从四类方法中选三类方法，从选定的每类方法中各选一种具体的方法，从给定的数据集中选一个数据集（CIFAR-10, Kaggle Telco Customer Churn Dataset, UCI Wine Quality Dataset）对这三种方法进行测试比较。
  - 第一类方法：线性方法：线性 SVM、Logistic Regression；
  - 第二类方法：非线性方法：Kernel SVM、决策树；
  - 第三类方法：集成学习：Bagging、Boosting；
  - 第四类方法：神经网络：自选结构。

加分项包括：

- 同类方法比较多个模型（如神经网络的 ResNet, ViT, 是否 pretrain 等）；
- 公正全面的算法比较，如训练和测试集的合理划分，量化指标比较和可视化比较，性能和计算资源开销的比较；
- 扎实的实验分析，如 bad case analysis，性能差距产生的原因，不同种类算法优劣的分析等；
- 优美的格式，文字的易读性（假定你的读者是没学过这门课程的人，为他们做些算法的简要介绍），合理的篇幅分配；
- 英文撰写会略为加分，但如果对驾驭文字表达没有充分自信，那么还是建议以中文撰写，加分项以前几项为主。

本次大作业需各位同学组成 5-6 人团队协同完成，要求不同队伍的数据集和算法组合不能完全相同，一名同学只能参与一个队伍，完成组队和选题的小组请在[腾讯文档链接](#)中顺序登记。此外，每份作业文件的第一节内容请简要说明所选数据集和具体算法，并罗列团队的人员分工安排，并且将自己所负责的内容部分加粗，例如，李四的作业文件里分工安排部分应如下：

- 张三，负责算法一实现和测试；
- **李四，负责算法二实现和测试；**
- 王五，负责算法三实现和测试；
- 小明，负责文档前三节的撰写；
- 小芳，负责文档后两节的撰写。