

注意事项及截止时间

1. 所有项目均需**通过Blackboard平台提交**。
2. **需要且仅需要提交两份文件**：
 - 实验报告，pdf格式。内容详见项目要求。
 - 实验代码等压缩包，zip格式。源程序及实验图表等放入同**一个**顶层文件夹，之后再打包文件夹，注意压缩包中有且仅有一个顶层文件夹。
 - **命名要求**：两份提交文件命名为学号姓名加后缀，文件夹命名为学号姓名，例如：报告（11956001张三.pdf），文件夹（11956001张三），文件夹打包后的文件（11956001张三.zip）
3. 所有提交**必须严格按照2中格式提交，否则该项project扣40分**。
4. 所有报告须在**期末考试前一天晚24时前**提交（以提交邮件时间为准），迟交的按0分处理。

数据描述

该数据集有train和test分为两个文件，表示为两个 $N \times (d+1)$ 的矩阵， N 为样本数， d 为特征数，矩阵最后一列是样本的类别（1或-1，即这是一个二类问题）。数据集有10个特征，其中train有8285个样本（因此矩阵有8285行，11列），用于训练；test有2072个样本。数据存为.data文件，可使用python读入，或用文本编辑器打开查看。

项目要求

- 请各自独立完成，鼓励讨论、交流。
 - ◆ **严禁抄袭。一经发现，按0分处理。**
 - ◆ **若提交的代码明显无法运行，同样按照0分处理。**
- 提交报告一份，该报告需至少包含以下内容：
 1. 个人信息（姓名、学号）
 2. 数据预处理方法（若未采用则可不写）
 3. 算法介绍
 - 采用的分类模型及其训练方法¹
 - 分类器输出后期处理方法（若未采用可不写）
 4. 实验
 - 实验采用的分类器性能评价标准²
 - 整体实验方法及步骤³
 - 分类器训练算法的参数调整步骤
 - 实验结果（整理成标准的图表）

¹ 若使用了软件包，请简要介绍之；若完全为自己编程实现也请注明，并将源程序随报告一同提交。自己编程实现算法将获得加分。

² 使用多种评判标准将获得加分。

³ 请注意validation和testing两个概念的区别，并比较validation和testing的结果是否一致。