

分类器设计

本次作业占期末总成绩 10%

现有已知正负标签(± 1)的训练样本 6821 个, 其中正样本 3417 个, 负样本 3404 个, 每个样本具有 107 维的特征, 样本特征矩阵和对应的标签保存在了 `data.mat` 中的 `train_X(6821 \times 107)` 和 `train_Y(1 \times 6821)` 中。请基于该训练数据完成分类器模型的训练、调参和选择。

另外, 在 `data.mat` 文件中, 还有未知标签样本 1705 个: `test_X(1705 \times 107)`, 请使用上述过程中得到的分类器预测这些样本标签。

要求:

1. 可以采用任何你觉得合理的方法对数据进行分析和学习, 请在报告中给出项目流程图, 描述你的数据分析和模型训练步骤、给出估计的预测错误率。提交相关程序代码。
2. 提交未知样本的标签; 将未知样本的预测标签命名为 `test_Y(1 \times 1705)`, 正样本为+1, 负样本为-1, 保存为 `test_Y.mat`, 连同报告提交。该项作为关键评分标准, 请务必按照格式要求提交。

评分标准:

1. 报告 60 分
2. 性能评价 40 分

根据在未知标签数据上的预测准确率(准确率=预测正确的标签数目/1705)来进行评分

$$\text{性能得分} = \frac{\text{你的准确率} - \text{最劣准确率}}{\text{最优准确率} - \text{最劣准确率}} \times 40$$

*最优准确率和最劣准确率分别是所有提交作业同学中的最高准确率和最低准确率。

截止日期:

6 月 22 日(毕业班同学), 6 月 30 日(非毕业班同学)