

数据降维与分类实验

概述

- 利用两种降维技术对葡萄酒数据进行降维，并对降维前后的数据进行分类。

数据说明

- 所给数据集包含两个.csv 文件，分别为红葡萄酒数据（包括 1599 个样本）和白葡萄酒数据（包含 4898 个样本），每个样本含有 11 个特征（文件的前 11 列）：固定酸度、挥发酸度、柠檬酸、残糖、氯化物、游离二氧化硫、总二氧化硫、密度、pH 值、硫酸盐、酒精和专家对此葡萄酒的打分（文件的最后一列）。该数据集的训练数据可以通过 <https://pan.seu.edu.cn:443/link/E2457AD6A5B3CC111AD45E90F0A8030D> 下载。

实验内容

- 对红白葡萄酒数据分别进行 PCA 和 LDA 降维处理，在降维后的数据集上完成基于 logistic 回归分类器的训练和测试。要求比较应用降维技术前后，分类器准确率的变化（对于降维后的数据，可以尝试利用可视化方法展示结果）。
- 基于 MindSpore 平台实现算法，对相同的数据集进行训练，并与不使用 MindSpore 平台实现的算法对比结果（包括但不限于准确率、算法迭代收敛次数等指标），并分析结果中出现差异的可能原因，给出使用 MindSpore 的心得和建议。
- （加分项）使用 MindSpore、sklearn 等平台提供的相似任务数据集（例如，其他的分类任务数据集）测试自己独立实现的算法，并与 MindSpore、sklearn 等平台上的官方实现算法进行对比，进一步分析差异及其成因。

实验要求

- 推荐使用 Python（在独立实现算法时，可采用 Numpy, Pandas, Matplotlib 等基础代码集成库；在使用 MindSpore 平台时，可使用平台提供的代码集成库）。
- 在独立实现算法时，不得使用集成度较高、函数调用式的代码库（如 sklearn, PyTorch, Tensorflow 等）。
- 尽量以相对路径的形式索引数据集，便于我们对代码进行复现。

实验报告格式

- 需要提供完整的可运行代码文件、测试集分类结果文件（降维前后）和实验报告，将以上内容打包压缩，压缩文件命名格式：学号-姓名-xxx 实验。实验报告和代码注释应尽量详细。需要以相对路径的形式索引数据集或文件，便于我们对代码进行复现。
- 实验报告内容参照报告模板，包括问题描述、实现步骤与流程、实验结果与分析、实验的心得体会（谈谈你自己的实现和 MindSpore 实现的差异、你在使用 MindSpore 平台过程中遇到的问题，以及想对平台改进提出的建议）、一个总的心得体会（谈一谈你对这门课程理论及实验的感悟与体会）。
- 代码和报告若有雷同，一律按 0 分处理。
- 若存在疑问，可以联系：pr_seu@163.com