**群体智能实验**

**任课教授：蔡宏民、魏凤凤**

**2024-2025下学期**

**实验题目：基于UCI大规模数据集的支持向量机（SVM）分类实践**

**实验目的**：

1. 深度掌握支持向量机（SVM）的核心原理、算法运行逻辑以及核函数在模型中的应用机制，通过实际操作熟练完成SVM分类器的搭建与参数调整。
2. 熟练运用数据预处理、特征工程等技术处理UCI大规模数据集，增强在大规模数据环境下的数据清洗、转换与特征优化能力。
3. 学会运用多种评估指标对SVM分类器进行全面性能评估，能够对实验结果进行深入分析，总结实验过程中的问题与解决策略，提升机器学习问题的综合解决能力。

**实验要求**：

1. **数据处理与准备**

* 从UCI公开数据集平台（[Home - UCI Machine Learning Repository](https://archive.ics.uci.edu/)）筛选一个具有代表性的大规模数据集，完成数据集的下载、解压，并导入至Python实验环境。
* 对数据集开展系统性预处理：使用统计学方法检测并处理异常值；针对缺失值，依据数据特征选择多重填补法、K近邻填补法等进行填充；利用正则表达式、文本编码转换等手段处理非结构化或半结构化数据。
* 采用归一化（Min-Max Scaling）或标准化（Standard Scaler）方法对数据特征进行缩放，同时，依据数据集规模和特点，合理运用分层抽样、留一法等策略划分训练集、验证集（可选，用于超参数调优）和测试集。

1. **SVM 模型构建与优化**

* 基于Scikit-learn库或其他机器学习框架，完成SVM分类器的搭建，依次尝试线性核函数、多项式核函数、径向基核函数（RBF）等，观察不同核函数下模型的训练时长与分类效果。
* 运用网格搜索结合交叉验证、随机搜索、贝叶斯优化等调参技术，对SVM模型的超参数进行优化，记录不同超参数组合下模型在验证集上的性能表现，确定最优模型参数。

1. **模型训练、评估与分析**

* 使用划分好的训练集对优化后的SVM分类器进行训练，记录训练过程中的关键信息（如损失函数变化曲线、训练耗时），并保存训练好的模型。
* 利用测试集对训练好的模型进行预测，计算并输出评估指标，绘制可视化图表，直观展示模型分类效果。
* 对比分析不同核函数、超参数设置下模型性能差异，探究数据集规模、特征维度对SVM分类器性能的影响，撰写详细的实验报告，内容涵盖实验背景、数据处理流程、模型构建与优化过程、实验结果分析及结论等。

截止时间：2025年7月13日24时前，发送至邮箱fengfengwei@scut.edu.cn

华南理工大学

《机器学习》课程实验报告

实验题目：

姓名： 学号：

班级： 指导教师：

|  |
| --- |
| **实验概述** |
| 【实验目的及要求】  实验目的：  实验要求：  【实验环境】 |
| **实验内容** |
| 【实验过程】  一、实验步骤：  二、实验数据：  三、实验主要过程： |
| **小结** |
|  |
| **指导教师评语及成绩** |
| 评语：  成绩：           指导教师签名：                                                 批阅日期： |