

作业要求

一、数据导入

从 UCI 网站 (<http://archive.ics.uci.edu/ml/index.php>) 下载 Breast Cancer Wisconsin 数据集，并用 Python 实现数据的读取。

二、数据表述

对 Breast Cancer Wisconsin 数据集进行数据描述，包括样本数、特征数、特征信息、类别信息等。

三、缺失值处理

分析 Breast Cancer Wisconsin 数据集中数据缺失情况并处理缺失值。

四、数据标准化

将各维度的特征数据标准化为均值为 0，方差为 1 的数据。

五、构造训练集和测试集

利用 sklearn 库实现留出法和 K 折交叉验证，构造随机的训练集和测试集。

六、模型训练与测试

利用 sklearn 库实现 Logistic 回归模型和决策树模型，完成训练和测试过程。

七、性能度量

分别计算 Logistic 回归模型和决策树模型的精度、混淆矩阵、查准率、查全率、F1 指标，绘制 ROC 曲线并计算 AUC 值。