

本次作业仍采用作业一使用的乳腺癌预测数据集。

1、读取数据并将数据集划分为训练集和测试集，使用随机森林方法进行预测，并输出在测试集上的准确率，绘制 ROC 曲线并获得 AUC 指标。

2、采用决策树桩（深度为 1 的决策树，即每次只采用一个特征的一个阈值作为划分依据）作为基本分类器，探究随基本分类器个数变化，随机森林在测试集上的准确率的变化，画出变化曲线并进行分析。

编写程序完成上述要求，并写一份 500 字左右的实验报告（pdf 格式，包括程序实现说明，实验结果及所画图像，实验结果分析），2021 年 5 月 20 日前将源代码与实验报告提交到

chenty@stu.pku.edu.cn