

机器学习算法实践

题目一： 机器学习分类算法： 利用机器学习算法完成糖尿病数据集的分类预测。

数据介绍：该数据集包含 768 个人的 8 个指标，最后一列 1 代表是糖尿病患者，0 位非糖尿病患者（该题未单独划分训练集测试集，请自行划分训练）。

要求：至少使用三种分类算法实现分类预测，编程语言不限（python、R、matlab 都可以）。建议使用 Jupyter 软件记录详细的实现代码（最好添加注释、中间图、中间输出结果等），比较不同算法的预测指标（准确率、精准率、召回率、AUC 曲线、F1 Score 等）。对结果进行解释，阐述算法优缺点。

题目二： 机器学习聚类算法： 利用机器学习算法完成蛋白表达模式的聚类预测。

数据介绍：该数据集包含 7 个不同时间点的 54 个蛋白的表达丰度数据，每一行代表一个蛋白，第一列代表每一个样本。（没有参考的聚类类别数目，根据自己的探索提出聚类类别数目并阐述证据）

要求：至少使用两种分类算法实现蛋白的聚类分析，编程语言不限（python、R、matlab 都可以）。建议使用 Jupyter 软件记录详细的实现代码（最好添加注释、中间图、中间输出结果等）。对结果进行解释，阐述算法优缺点。

请于 3 月 25 日之前完成作业并发送到助教邮箱（kongyan@sjtu.edu.cn）。有疑问可以在微信群询问。