**第7章 决策树**

**1.**概述**决策树算法**的思路、关键操作（不超过300字，逻辑要通，**要有小标题**，要排版清晰，请不要网络大幅摘抄）。

**决策树（Decision Tree，DT）**

决策树是一种常用的机器学习算法，可用于处理分类和回归问题，既可以用来高效地对未知的数据进行分类，也可以用来做预测。决策树通过学习可以得到一个树形结构模型，它表示的是对象属性与对象值之间的一种映射关系。算法有 ID3、C4.5、CART 等。已经成功被应用于商业、医学、金融分析、分子生物学领域。

**决策树的基本原理**

**基本原理： 决策树学习算法就是一棵树的构造过程，它通过不断地选择最优特征，并根据该特征对训练数据进行分割，即对特征空间进行划分，使得各个子数据集有一个最好的分类过程。一棵决策树包含一个根节点、若干个内部节点和若干个叶节点，根节点包含样本全集，其他每个节点对应一个属性测试，叶节点对应于决策结果，从根节点到各个叶节点的路径对应了一个判定测试序列。**

**2.利用决策树解决问题（回归或分类）。**

附：Sklearn常用数据集

|  |  |  |
| --- | --- | --- |
| 序号 | 数据集 | 学号尾号 |
| 1 | 鸢尾花数据集：load\_iris（） | 0 1 |
| 2 | 手写数字数据集：load\_digits（） | 2 3 |
| 3 | 乳腺癌数据集load-barest-cancer（） | 4 5 |
| 4 | 糖尿病数据集：load-diabetes（） | 6 7 |
| 5 | 波士顿房价数据集：load-boston（） | 8 |
| 6 | 体能训练数据集：load-linnerud（） | 9 |

利用决策树对以上数据集（学号尾号不同采用不同数据集）进行分析，找出最佳参数下的决策树或随机森林更适合本数据集（准确率更高，不同回归、不同参数下的不同准确率要列出）。

要求：

1）针对当前数据、模型，**详叙数据预处理、模型参数调整过程及效果比对**；

2）代码要列出，重点代码加注释说明，特别是自己调试过程中的自我理解；

3）运行结果要截图，结果要文字说明；

4）注意排版

**3.[附加题].**利用**决策树**解决**自己遇到的问题**。

[解题要求同题2]