实验截图包括代码和答案

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

1

ID3方法构造决策树使用的依据[？]

C4.5方法构造决策树使用的依据[？]

CART方法构造决策树使用的依据[？]

A最大信息增益

B最大信息增益比

C最大基尼系数

答

ABC

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

2 试分析预剪枝和后剪枝的优缺点

答：

预剪枝：

训练时间开销降低，测试时间开销降低；

过拟合风险降低，欠拟合风险增加；

泛化能力弱于后剪枝；

后剪枝：

训练时间开销增加，测试时间开销降低；

过拟合风险降低，欠拟合风险基本不变；

泛化能力优于预剪枝

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

3 请根据前面的实验内容实现香农熵计算函数 entropy()。。

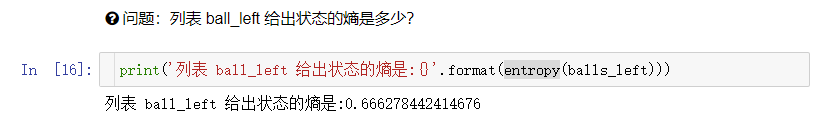
答



%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

4 列表 ball\_left 给出状态的熵是多少？

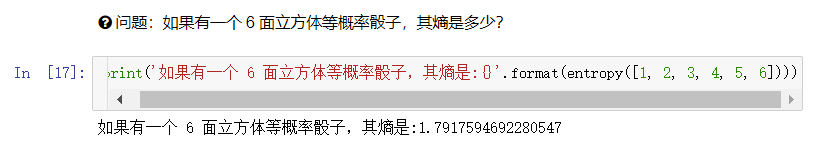
答



%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

5 如果有一个 6 面立方体等概率骰子，其熵是多少？

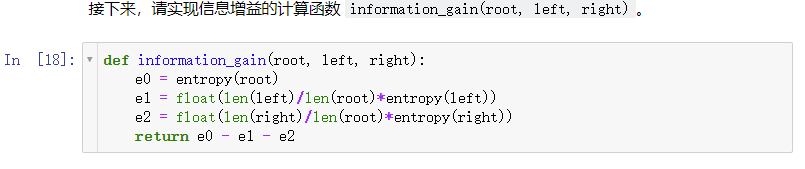
答



%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

6 请实现信息增益的计算函数 information\_gain(root, left, right)。？

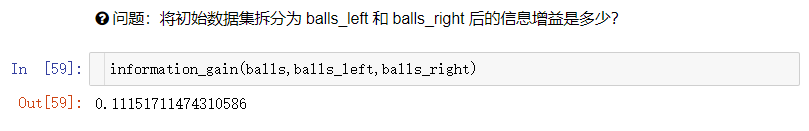
答



%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

7 将初始数据集拆分为 balls\_left 和 balls\_right 后的信息增益是多少？。

答



%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

8 按挑战要求构建决策树，并输出其在测试集上的准确度？

答



%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

9 使用 GridSearchCV 网格搜索对决策树进行调参并返回最佳参数。

挑战规定，决策树参数 random\_state = 17，GridSearchCV 参数 cv=5，并对 max\_depth 参数在 [8,10][8,10] 范围进行网格搜索。

答



%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

10 构建上面最佳参数决策树，并输出其在测试集上的准确度？

答

