**分类问题作业**

|  |
| --- |
| 得分 |
|  |

**一、下表是14天的气象数据和是否打球的对应关系表。气象信息有4个属性，分别是Outlook、Temp、Humidity和Windy，而分类指标是Play。根据ID3和C4.5算法，计算：**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **天气与打球关联关系表** | | | | | |
| **Day** | Outlook | Temp | Humidity | Windy | Play |
| **D1** | Sunny | Hot | High | Weak | No |
| **D2** | Sunny | Hot | High | Strong | No |
| **D3** | Overcast | Hot | High | Weak | Yes |
| **D4** | Rain | Mild | High | Weak | Yes |
| **D5** | Rain | Cool | Normal | Weak | Yes |
| **D6** | Rain | Cool | Normal | Strong | No |
| **D7** | Overcast | Cool | Normal | Strong | Yes |
| **D8** | Sunny | Mild | High | Weak | No |
| **D9** | Sunny | Cool | Normal | Weak | Yes |
| **D10** | Rain | Mild | Normal | Weak | Yes |
| **D11** | Sunny | Mild | Normal | Strong | Yes |
| **D12** | Overcast | Mild | High | Strong | Yes |
| **D13** | Overcast | Hot | Normal | Weak | Yes |
| **D14** | Rain | Mild | High | Strong | No |

1．分别给出Outlook、 Temp、 Humidity、 Windy天气属性的信息增益和信息增益率。

2．画出两种分类模型下的决策树（在纸上画好拍照即可）。

**二、对loan\_data.xlsx中数据进行以下操作（给出代码和运行结果图）：**

1．将数据集进行预处理，将文本型转化为数值型：

（1）年龄：0代表青年，1代表中年，2代表老年；

（2）有工作：0代表否，1代表是；

（3）有自己的房子：0代表否，1代表是；

（4）信贷情况：0代表一般，1代表好，2代表非常好；

（5）类别(是否给贷款)：0代表否，1代表是。;

2. 采用ID3和Cart 算法构建分类模型，分别画出决策树（给出关键代码和决策树截图）。

**三、读取sklearn的Wine数据集，将数据集划分为训练集和测试集，其中测试集占15%，采用支持向量机构建分类模型（给出关键代码和决策树截图）。：**

1．给出分类报告。

2. 每个类别的召回率和精确率是多少？