1.题目

**决策树练习**

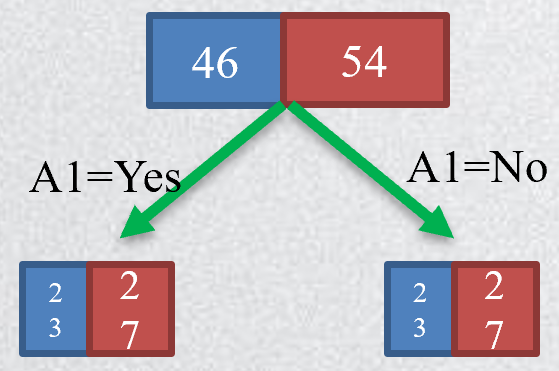
2.数据集

参考任务描述。

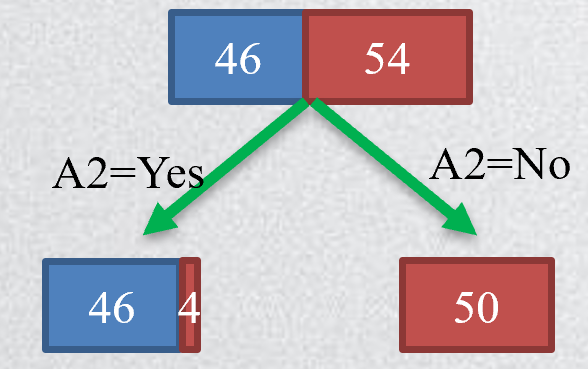
3.任务描述

（1）在决策树的构建过程中，需要选择第1个决策属性，

请计算A1属性的信息增益（蓝色表示一个类别，红色表示另一个类别；框里的数值表示样本数量；当A1=yes，数据划分出一个子集，包含两个类别的样本，数量见框内数值；当A1=no，数据划分出另一个子集，包含两个类别的样本，数量见框内数值）



请计算A2属性的信息增益（对下图的解释同上）



请问应该选择哪个属性对数据集进行划分？（自己用纸笔完成，拍照，以图片形式嵌入到ipynb文件中(Markdown Cell)，图片和ipynb文件打包提交）；

（2）示例代码在参考文献[2]（分析提供的示例代码，增加注释的Cell）；

（3）自行选择参考文献[3]的一个数据集，使用scikit learn库的Decision Tree进行分类，并且可视化最后生成的决策树。

4.成果提交

以ipynb形式提交。

5.参考文献

[1] <https://sefiks.com/2017/11/20/a-step-by-step-id3-decision-tree-example/>

[2] <https://medium.com/@lope.ai/decision-trees-from-scratch-using-id3-python-coding-it-up-6b79e3458de4>

请自行通过VPN访问

[3] <https://archive.ics.uci.edu/ml/datasets.php>

[4] Visualize Decision Tree with Python Sklearn Library. <https://vitalflux.com/visualize-decision-tree-python-sklearn-library/>