决策树作业

说明：请使用该模板作答，格式转成**PDF**，文件名为**学号\_姓名.pdf**

学号：2017211146 班级：2017211303 姓名：张静雅

1. 请使用最大信息增益算法为课件73页的数据构建决策树，写出计算过程并画出决策树。（30分）
2. 假定数据库有个人，第个人的先验概率，有个问题，假定第个人对第个问题答案为“是”的概率为，请给出给定第个问题条件下，数据集的条件熵的计算公式。（20分）

**以下两道题目，二选一即可。**(50分)

1. 请选择一个你认为有意义或有趣味的领域，收集一个可以用作20问读心游戏的数据集。角色数不小于100个，问题数不小于20个。请写出你选择该领域的理由和数据集的收集方法。（数据集列出角色和问题，角色对问题的答案仅选择20个角色和10个问题即可）
2. 请编程实现题目1，要求代码运行能够直接打印出决策树。代码只能包含一个文件，文件名为**学号\_姓名.py**。编程环境要求如下：

* Python 3.6.10
* python standard library
* numpy == 1.16.2
* scipy == 1.2.1
* pandas == 0.24.2
* networkx == 2.4
* graphviz==0.13.2
* matplotlib==3.2.1