

完成以下两个作业中的一个。

## 作业一：基于贝叶斯网络的简易医疗诊断系统

### (一) 问题描述

请构建一个简化的医疗诊断贝叶斯网络模型。该模型包含一组相关的变量（疾病、症状、基本情况），你的任务：

1. 从给定的数据集中“学习”出这个网络的条件概率表（CPTs）。
2. 实现一个推理算法，根据输入的“证据”（比如观察到的症状），计算出患有某种疾病的概率。

我们定义如下的贝叶斯网络结构：

网络结构图：

```
A[A: 吸烟 IsSmoker] --> B [B: 患肺癌 HasLungCancer];  
C[C: 旅行史 VisitAsia] --> D [D: 患结核病 HasTuberculosis];  
B --> E [E: X 光异常 XRayPositive];  
D --> E;  
B --> F [F: 呼吸困难 HasDyspnea];
```

变量解释：

- A (IsSmoker): 是否吸烟 (Boolean)
- C (VisitAsia): 近期是否有亚洲旅行史 (Boolean)
- B (HasLungCancer): 是否患有肺癌 (Boolean)
- D (HasTuberculosis): 是否患有结核病 (Boolean)
- E (XRayPositive): X 光检查是否呈阳性 (Boolean)
- F (HasDyspnea): 是否有呼吸困难症状 (Boolean)

依赖关系：

- $P(B | A)$ : 患肺癌的概率取决于是否吸烟。
- $P(D | C)$ : 患结核病的概率取决于是否有亚洲旅行史。
- $P(E | B, D)$ : X 光结果取决于是否患有肺癌和/或结核病。
- $P(F | B)$ : 呼吸困难的症状取决于是否患有肺癌。

- $P(A)$  和  $P(C)$  是先验概率。

## (二) 任务要求

### 任务 1：参数学习

1. **输入:** 提供一个模拟生成的病人数据文件 `data.csv`, 包含数百条匿名病人的记录。每行代表一个病人, 每列对应上述一个变量 (A 到 F), 值为 0 或 1。
2. **实现:** 编写一个程序, 读取 `data.csv` 文件, 并根据数据使用最大似然估计来计算网络中每个节点的条件概率表 (CPT)。
  - 例如, 为了计算  $P(E=1 | B=1, D=0)$ , 你需要找到所有  $B=1$  且  $D=0$  的数据行, 然后计算这些行中  $E=1$  的比例。
3. **输出:** 将学习到的所有 CPTs 清晰地打印出来或保存到一个文件 (如 `cpts.json`) 中。

### 任务 2：精确推理

1. **实现:** 基于你在任务一中学到的 CPTs, 实现推理算法。
2. **查询:** 你的程序需要能够回答以下类型的查询：
  - 查询 1：一位病人 X 光呈阳性( $E=1$ ), 并且有呼吸困难症状 ( $F=1$ ), 求他患有肺癌的概率是多少? 即计算  $P(B=1 | E=1, F=1)$ 。
  - 查询 2：一位已知的吸烟者( $A=1$ ), 他 X 光呈阳性的概率是多少? 即计算  $P(E=1 | A=1)$ 。
  - 查询 3：一位吸烟者( $A=1$ ), 近期去过亚洲( $C=1$ ), 并且 X 光呈阳性( $E=1$ ), 求他患有结核病的概率是多少? 即计算  $P(D=1 | A=1, C=1, E=1)$ 。
3. **输出:** 对每个查询, 输出清晰的计算结果 (概率值)。

## (三) 实验报告要求

要求撰写实验报告, 提交源程序。

为了方便实验, 可以使用提供的 Python 脚本生成 `data.csv`。这能确保数据符合我们预设的网络结构。

## 作业二：实现贝叶斯网络近似推理算法

### (一) 问题描述

实现两种基本的近似推理算法——拒绝采样和似然加权，并在一个简单的贝叶斯网络（PPT示例）上进行应用和比较。

网络结构图：

- $C \rightarrow S$
- $C \rightarrow R$
- $S \rightarrow W$
- $R \rightarrow W$

变量解释：

- $C(\text{Cloudy})$ : 是否多云 (Boolean)
- $S(\text{Sprinkler})$ : 洒水器是否开启 (Boolean)
- $R(\text{Rain})$ : 是否下雨 (Boolean)
- $W(\text{WetGrass})$ : 草坪是否湿润 (Boolean)

条件概率表 (CPTs)：

- $P(C=T) = 0.5$
- $P(S | C):$ 
  - $P(S=T | C=T) = 0.10$
  - $P(S=T | C=F) = 0.50$
- $P(R | C):$ 
  - $P(R=T | C=T) = 0.80$
  - $P(R=T | C=F) = 0.20$
- $P(W | S, R):$ 
  - $P(W=T | S=T, R=T) = 0.99$
  - $P(W=T | S=T, R=F) = 0.90$
  - $P(W=T | S=F, R=T) = 0.90$
  - $P(W=T | S=F, R=F) = 0.00$

## (二) 任务要求

实现拒绝采样算法和似然加权算法，算法需要能够估计条件概率  $P(X | e)$ ，其中  $X$  是查询变量， $e$  是证据。使用你实现的两种算法估计  $P(R | S=T, W=T)$ 。

## (三) 实验报告要求

要求撰写实验报告，提交源程序。