Assignment 3

1. 贝叶斯规则应用

已知脑膜炎会导致患者 70% 的几率出现僵硬的脖子. 如果任何患者患有脑膜炎的先验概率为 $\frac{1}{50000}$，且任何患者出现僵硬的脖子的概率为 1%，求一位出现僵硬脖子的患者实际患有脑膜炎的概率.

首先，我们设两个事件： A 表示患者患有脑膜炎， B 表示患者出现僵硬的脖子.

由题意得：

- $P(A) = \frac{1}{50000}$：任意患者患有脑膜炎的先验概率.
- $P(B|A) = 0.7$：如果患者患有脑膜炎，则患者出现僵硬的脖子的概率.
- $P(B) = 0.01$：任意患者出现僵硬的脖子的概率.

要求 $P(A|B)$，即求患者出现僵硬的脖子的情况下，患者实际患有脑膜炎的概率.

代入贝叶斯公式得：

$$P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B)} = \frac{0.7 \cdot \frac{1}{50000}}{0.01} = 0.0014 = \boxed{0.14\%}$$

因此，出现僵硬脖子的患者实际患有脑膜炎的概率是 0.14%.

2. 贝叶斯网络推理 - 陷阱问题

在探索一个未知地区的过程中，系统遇到三个方格，每个方格可能含有陷阱. 系统检测到每个方格的附近有风的迹象，风的存在表明至少一个邻近方格可能有陷阱. 假设每个方格有陷阱的先验概率为 20%，已知：

1. 陷阱引起风的条件概率是 0.75
2. 无陷阱时引起风的概率是 0.25

要求：使用贝叶斯推理，计算每个方格含有陷阱的概率.

首先，我们作出如下假设：

- 设 T_i 表示第 i 个方格含有陷阱（$i \in \{1, 2, 3\}$）.
- $P(T_i) = 0.2$：每个方格含有陷阱的先验概率.
- 设 W 表示系统检测到风的迹象.
- $P(W|T_i) = 0.75$：如果第 i 个方格有陷阱，则检测到风的概率.
- $P(W|\neg T_i) = 0.25$：如果第 i 个方格没有陷阱，则检测到风的概率.

我们要求的是 $P(T_i|W)$，即检测到风的情况下，第 i 个方格含有陷阱的概率. 可以使用贝叶斯定理：

$$P(T_i|W) = \frac{P(W|T_i)P(T_i)}{P(W)}$$

首先，计算 $P(W)$，即检测到风的总概率. 我们需要考虑风的迹象可能由以下几种情况引起，也即有一个或多个方格有陷阱，或者所有方格都没有陷阱. 对于每个方格，我们考虑独立性，风的迹象可以由其中至少一个方格有陷阱引起. 具体计算如下：

风迹象的总概率可以分解为：

$$P(W) = P(W|\text{至少一个方格有陷阱}) \cdot P(\text{至少一个方格有陷阱}) + P(W|\text{没有方格有陷阱}) \cdot P(\text{没有方格有陷阱})$$

其中，

$$P(\text{至少一个方格有陷阱}) = 1 - P(\text{没有一个方格有陷阱})$$

$$P(\text{没有一个方格有陷阱}) = (1 - 0.2)^3 = 0.8^3 = 0.512$$

因此，

$$P(\text{至少一个方格有陷阱}) = 1 - 0.512 = 0.488$$

如果至少一个方格有陷阱，风的概率就是 $1 - P(\text{没有一个方格引起风})$

$$P(W|\text{至少一个方格有陷阱}) = 1 - (1 - 0.75)^3 = 1 - 0.25^3 = 1 - 0.015625 = 0.984375$$

$P(W|\text{没有方格有陷阱})$：
$$P(W|\text{没有方格有陷阱}) = 0.25$$

因此，
$$P(W) = 0.984375 \cdot 0.488 + 0.25 \cdot 0.512 = 0.480390625 + 0.128 = 0.608390625$$

应用贝叶斯公式代入得:

$$P(T_i|W) = \frac{0.75 \cdot 0.2}{0.608390625} = \frac{0.15}{0.608390625} \approx 0.2465 = \boxed{24.65\%}$$

综上，在检测到风的情况下，每个方格含有陷阱的概率约为 24.65%.

3. 贝叶斯网络条件概率表

假设你正在使用贝叶斯网络来建模一个简单的医疗诊断系统，该系统旨在根据病人的症状判断其是否患有某种疾病. 患病的先验概率为 5%.

- 变量 "Disease" 表示病人是否患有疾病，"Yes" 和 "No".
- 变量 "Symptom" 表示病人是否展示特定的症状，"Present" 和 "Absent".

要求:
填写以下表格，并解释如何使用这个 CPT 来计算一个病人展示症状时实际患病的概率.

	Symptom=Present	Symptom=Absent
Disease=Yes	$P(S\|D=Yes)=80\%$	$P(S\|D=Yes)=20\%$
Disease=No	$P(S\|D=No)=10\%$	$P(S\|D=No)=90\%$

我们使用贝叶斯定理来计算在病人展示症状时其实际患病的概率

$$P(D=\text{Yes} \mid S=\text{Present}) = \frac{P(S=\text{Present} \mid D=\text{Yes}) \cdot P(D=\text{Yes})}{P(S=\text{Present})}$$

首先有:

$$P(S=\text{Present} \mid D=\text{Yes}) = 0.80$$

$$P(D=\text{Yes}) = 0.05$$

因此,

$$P(S=\text{Present} \mid D=\text{Yes}) \cdot P(D=\text{Yes}) = 0.80 \cdot 0.05 = 0.04$$

接下来使用全概率公式:

$$P(S=\text{Present}) = P(S=\text{Present} \mid D=\text{Yes}) \cdot P(D=\text{Yes}) + P(S=\text{Present} \mid D=\text{No}) \cdot P(D=\text{No})$$

其中,

$$P(S=\text{Present} \mid D=\text{No}) = 0.10$$

$$P(D=\text{No}) = 0.95$$

所以,

$$P(S=\text{Present}) = (0.80 \cdot 0.05) + (0.10 \cdot 0.95) = 0.04 + 0.095 = 0.135$$

最后，代入得:

$$P(D=\text{Yes} \mid S=\text{Present}) = \frac{0.04}{0.135} \approx 0.296 = \boxed{29.6\%}$$

因此，当病人展示症状时，实际患病的概率约为 29.6%.

4. 概率计算

已知变量 A 和 B 的取值只能为 0 或 1，$A \perp\!\!\!\perp B$，且 $P(A=1)=0.65$，$P(B=1)=0.77$，C 的取值与 A 和 B 有关，具体关系如下表所示.

A	B	$P(C=1\|A,B)$
0	0	0.1
0	1	0.99
1	0	0.8
1	1	0.25

求 $P(A=1|C=0)$.

我们要计算 $P(A=1 \mid C=0)$. 为此，我们需要使用贝叶斯定理:

$$P(A=1 \mid C=0) = \frac{P(C=0 \mid A=1) \cdot P(A=1)}{P(C=0)}$$

首先，计算 $P(C=0 \mid A=1)$ 和 $P(C=0)$.

根据题目中给出的条件概率表，我们有:

$P(C=0 \mid A=1,B=0)=1-P(C=1 \mid A=1,B=0)=1-0.8=0.2$

$P(C=0 \mid A=1,B=1)=1-P(C=1 \mid A=1,B=1)=1-0.25=0.75$

由于 A 和 B 相互独立，故有:

$P(C=0 \mid A=1)=P(C=0 \mid A=1,B=0)\cdot P(B=0)+P(C=0 \mid A=1,B=1)\cdot P(B=1)$

又

$P(B=0)=1-P(B=1)=1-0.77=0.23$

$P(B=1)=0.77$

因此:

$P(C=0 \mid A=1)=0.2\cdot 0.23+0.75\cdot 0.77=0.046+0.5775=0.6235$

接下来，我们需要计算 $P(C)=0$ 我们已知:

$P(C=0)=P(C=0 \mid A=0)\cdot P(A=0)+P(C=0 \mid A=1)\cdot P(A=1)$

首先，计算 $P(C=0 \mid A=0)$:

$P(C=0 \mid A=0,B=0)=1-P(C=1 \mid A=0,B=0)=1-0.1=0.9$

$P(C=0 \mid A=0,B=1)=1-P(C=1 \mid A=0,B=1)=1-0.99=0.01$

因此有:

$P(C=0 \mid A=0)=0.9\cdot P(B=0)+0.01\cdot P(B=1)=0.9\cdot 0.23+0.01\cdot 0.77=0.207+0.0077=0.2147$

然后，使用先验概率 $P(A=0)$ 和 $P(A=1)$ 计算 $P(C=0)$:

$P(A=0)=1-P(A=1)=1-0.65=0.35$

所以:

$P(C=0)=P(C=0 \mid A=0)\cdot P(A=0)+P(C=0 \mid A=1)\cdot P(A=1)$

$=0.2147\cdot 0.35+0.6235\cdot 0.65$

$=0.075145+0.405275=0.48042$

最后，代入贝叶斯定理计算 $P(A=1 \mid C=0)$:

$$\begin{aligned}P(A=1 \mid C=0)&=\frac{P(C=0 \mid A=1)\cdot P(A=1)}{P(C=0)}\\&=\frac{0.6235\cdot 0.65}{0.48042}=\frac{0.405275}{0.48042}\\&\approx \boxed{0.8435}\end{aligned}$$

5. **朴素贝叶斯**

基于朴素贝叶斯算法的医疗诊断系统，诊断病人是否患有某疾病.

1. 已知某疾病与 ABCD 四个基因突变标记有关，每个基因突变标记都可以是阳性 P 或阴性 S
2. 已有以下概率:

- 基因突变标记 A 为阳性的条件概率: $P(A=P|Disease=Yes)=0.8$, $P(A=P|Disease=No)=0.1$
- 基因突变标记 B 为阳性的条件概率: $P(B=P|Disease=Yes)=0.6$, $P(B=P|Disease=No)=0.2$
- 基因突变标记 C 为阳性的条件概率: $P(C=P|Disease=Yes)=0.4$, $P(C=P|Disease=No)=0.1$
- 基因突变标记 D 为阳性的条件概率: $P(D=P|Disease=Yes)=0.7$, $P(D=P|Disease=No)=0.3$

3. 已知患病概率 $P(Disease=Yes)=0.01$，不患病概率 $P(Disease=No)=0.99$.
4. 一个病人的基因突变标记检测结果如下: A 阳性 B 阴性 C 阳性 D 阳性

要求: 使用朴素贝叶斯分类器，计算这个病人患有该疾病的概率.

要计算一个病人在给定基因突变标记检测结果下患有某疾病的概率，我们使用朴素贝叶斯分类器. 具体地，我们需要计算 $P(Disease=Yes \mid A=P,B=S,C=P,D=P)$.

由贝叶斯定理:

$$P(Disease = Yes \mid A = P, B = S, C = P, D = P) = \frac{P(A = P, B = S, C = P, D = P \mid Disease = Yes) \cdot P(Disease = Yes)}{P(A = P, B = S, C = P, D = P)}$$

朴素贝叶斯假设各个基因突变标记是条件独立的，因此：

$$P(A = P, B = S, C = P, D = P \mid Disease = Yes) = P(A = P \mid Disease = Yes) \cdot P(B = S \mid Disease = Yes) \cdot P(C = P \mid Disease = Yes) \cdot P(D = P \mid Disease = Yes)$$

类似地，

$$P(A = P, B = S, C = P, D = P \mid Disease = No) = P(A = P \mid Disease = No) \cdot P(B = S \mid Disease = No) \cdot P(C = P \mid Disease = No) \cdot P(D = P \mid Disease = No)$$

现在，我们需要计算 $P(A = P, B = S, C = P, D = P)$，用全概率公式表达：

$$P(A = P, B = S, C = P, D = P) = P(A = P, B = S, C = P, D = P \mid Disease = Yes) \cdot P(Disease = Yes) + P(A = P, B = S, C = P, D = P \mid Disease = No) \cdot P(Disease = No)$$

代入上述公式.

$$P(A = P \mid Disease = Yes) = 0.8$$

$$P(B = S \mid Disease = Yes) = 1 - P(B = P \mid Disease = Yes) = 1 - 0.6 = 0.4$$

$$P(C = P \mid Disease = Yes) = 0.4$$

$$P(D = P \mid Disease = Yes) = 0.7$$

因此，

$$P(A = P, B = S, C = P, D = P \mid Disease = Yes) = 0.8 \cdot 0.4 \cdot 0.4 \cdot 0.7 = 0.0896$$

$$P(A = P \mid Disease = No) = 0.1$$

$$P(B = S \mid Disease = No) = 1 - P(B = P \mid Disease = No) = 1 - 0.2 = 0.8$$

$$P(C = P \mid Disease = No) = 0.1$$

$$P(D = P \mid Disease = No) = 0.3$$

因此，

$$P(A = P, B = S, C = P, D = P \mid Disease = No) = 0.1 \cdot 0.8 \cdot 0.1 \cdot 0.3 = 0.0024$$

$$P(A = P, B = S, C = P, D = P) = 0.0896 \cdot 0.01 + 0.0024 \cdot 0.99 = 0.000896 + 0.002376 = 0.003272$$

$$\begin{aligned} P(Disease = Yes \mid A = P, B = S, C = P, D = P) &= \frac{0.0896 \cdot 0.01}{0.003272} \\ &= \frac{0.000896}{0.003272} \\ &\approx 0.2738 \\ &= \boxed{27.38\%} \end{aligned}$$

综上，在给定基因突变标记检测结果下，这个病人患有该疾病的概率约为 27.38%.