第三次作业

提交 DDL：2021 年 12 月 6 日 0 时

徐薪 519021910726

作业完成形式有三种：

（1）你可以手写自己的解答并拍照，再将照片整理成一份 word/pdf 文件并提交。

（2）你可以使用 word 文档进行编辑，最后提交 word/pdf 文件。

（3）你可以使用 latex 进行编辑，最后提交 pdf 文件。

如果你没有在 DDL 之前提交作业，请及时在微信群里联系助教进行补交。如果对作业有任何问题，你可以在从微信里询问助教谢瑜璋，或者发邮件到 constantjxyz@sjtu.edu.cn。

1 本次作业可能用到的知识点

本次作业可能会用到以下知识点：

（1）命题逻辑（propositional logic）、谓词逻辑（first-order logic）的定义、语义

（2）逻辑证明的假言推理规则（Modus Ponens）、归结证明（resolution）。

（3）逻辑推理的前向链接算法（forward chaining）、后向链接算法（backward chaining）。

（3）贝叶斯网络节点联合概率表示，马尔科夫毯（Markov Blanket）的定义，贝叶斯网络中的条件独立（conditional independence）关系

（4）贝叶斯网络中的精确推理、消元法

（5）贝叶斯网络近似推理的采样方法

（6）HMM 模型的建模方式、概率推导

2 第一题

2-CNF 表达式类似于 3-CNF，其中每个子句 (clause) 里含有两个文字 (literals)，例如：

$$(a \vee b) \wedge (\neg a \vee c) \wedge (\neg b \vee d) \wedge (\neg c \vee g) \wedge (\neg d \vee g)$$

（1）运用反证法，证明上面的表达可以推导出 g (the above sentence entails g)，可以参考课件的归结证明 resolution 部分。

Proof. If the equation above is not equal to g,$(a \vee b) \wedge (\neg a \vee c) \wedge (\neg b \vee d) \wedge (\neg c \vee g) \wedge (\neg d \vee g) \wedge \neg g \neq \emptyset$. But according to resolution inference rule(CNF),$(a \vee b) \wedge (\neg a \vee c) \wedge (\neg b \vee d) \wedge (\neg c \vee g) \wedge (\neg d \vee g) \wedge \neg g = (b \vee c) \wedge (\neg b \vee d) \wedge (\neg c \vee g) \wedge (\neg d \vee g) \wedge \neg g = (c \vee d) \wedge (\neg c \vee g) \wedge (\neg d \vee g) \wedge \neg g = (d \vee g) \wedge (\neg d \vee g) \wedge \neg g = g \wedge \neg g \Rightarrow \emptyset$.
So, the assumption is wrong, and the original equation above can result in g. $\square$

（2）对于 2-CNF 问题，假设现在我们有 n 个不同的符号。如果我们规定每个子句要求用**不同**的符号组成，那么我们可以用这 n 个符号组成多少种**语义不同（semantic distinct）**的子句（clause）？如果我们规定每个子句可以用**相同**的符号组成，那么我们可以用这 n 个符号组成多少种**语义不同**的子句？

Solution. If request for same symbols among different clauses, we suppose $a = \binom{n}{2}$, which means the posible number of groups with two symbols. With the meaning of $\neg$, one group can have four

clauses. So, there are $4a = 4\binom{n}{2}$ semantic distinct clauses.
If request for different symbols among different clauses, there are $\frac{n!}{2^{\lfloor n/2 \rceil}}$ groups with no symbol the same. With the meaning of $\neg$, one group can have four clauses. So, there are $\frac{4n!}{2^{\lfloor n/2 \rceil}}$ semantic distinct clauses. □

3 第二题

我们建立了一个新的数学空间，在这个空间里有以下公理：

- 1. $0 \leq 3$
- 2. $7 \leq 9$
- 3. $\forall x \quad x \leq x$
- 4. $\forall x \quad x \leq x + 0$
- 5. $\forall x \quad x + 0 \leq x$
- 6. $\forall x, y \quad x + y \leq y + x$
- 7. $\forall w, x, y, z \quad x \leq y \wedge w \leq z \Rightarrow w + x \leq y + z$
- 8. $\forall x, y, z \quad x \leq y \wedge y \leq z \Rightarrow x \leq z$

我们希望用以上原子语句进行推理，得到 $7 \leq 3 + 9$。注意在推理的过程中，我们只能使用以上 8 条公理，不能使用现实数学中的各种运算。

（1）假如我们使用反向链接算法（见课件上 backward chaining 部分）。我们可以得到以下推理过程，请完善推理过程。

Goal G0: $7 \leq 3 + 9$, resolve with axiom 8 and $\{x0/7, z0/(3+9), y0/(7+0)\}$
*/*Use axiom 8, and substitute the (x, y, z) in axiom with (7, (7+0), (3+9)*/*
*/*To achieve Goal G0, we need to find a intermediate (7+0) and achieve Goal G1 and G2.*/*
  Goal G1: $7 \leq 7 + 0$, resolve with (a) axiom4________. Goal G1 Succeeds.
  Goal G2: $7 + 0 \leq 3 + 9$, resolve with (b) axiom7________.
*/*To achieve Goal G2, we need to find a intermediate and achieve Goal G3 and G4.*/*
    Goal G3: $7 + 0 \leq n$, resolve with (c) axiom6________. Goal G3 Succeeds.
    Goal G4: $n \leq 3 + 9$, resolve with (d) axiom7________.
*/*To achieve Goal G4, we need to find a intermediate and achieve Goal G5 and G6.*/*
      Goal G5: $0 \leq 3$, resolve with axiom 1. Goal G5 Succeeds.
      Goal G6: $7 \leq 9$, resolve with axiom 2. Goal G6 Succeeds.
    Goal G4 succeeds.
  Goal G2 succeeds.
Goal G0 succeeds.

(2) 假如我们使用前向链接算法（见课件上 forward chaining 部分），我们可以怎样推理得到结论？请写出推理过程。

Solution. We can first order these axioms by the number of preconditions. So the order is: 1,2,3,4,5,6,8,7. And we will reason from the first one:
$1.0 \leq 3$ and $2.7 \leq 9 \Rightarrow 7.\forall w, x, y, z \quad x \leq y \wedge w \leq z \Rightarrow w + x \leq y + z$. So, we have $0 + 7 \leq 3 + 9$.
Since we have $0 + 7$, according to $6.\forall x, y \quad x + y \leq y + x$, we have $0 + 7 \leq 7 + 0$.

Since we have $7+0$, according to $4.\forall x \quad x \le x+0$, we have $7 \le 7+0$.
Since we have $7+0$, according to $5.\forall x \quad x+0 \le x$, we have $7+0 \le 7$.
Since we have $7+0$, according to $6.\forall x,y \quad x+y \le y+x$, we have $7+0 \le 0+7$.
Since we have $7 \le 7+0$ and $7+0 \le 0+7$, according to $8.\forall x,y,z \quad x \le y \wedge y \le z \Rightarrow x \le z$, we have $7 \le 0+7$.
Since we have $7 \le 0+7$ and $0+7 \le 3+9$, according to $8.\forall x,y,z \quad x \le y \wedge y \le z \Rightarrow x \le z$, we have $7 \le 3+9$.
So, finally, we have made it. □

4 第三题

给定下图1所示的贝叶斯网络。网络中有 (B, A, E, J, M) 五个变量。

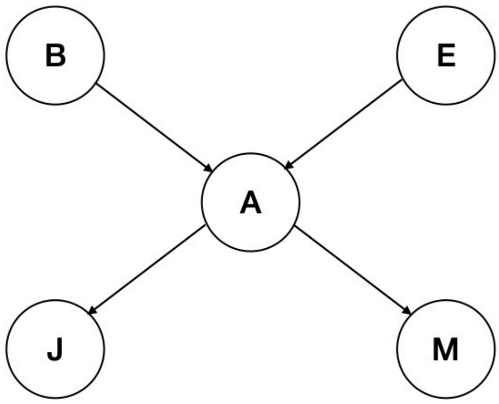


图 1: 第三题的贝叶斯网络

（1）根据给定的贝叶斯网络对联合概率 $P(B, E, A, J, M)$ 进行因子分解。

Solution. $P(B, E, A, J, M) = P(B)P(E)P(A|B, E)P(J|A)P(M|A)$. □

（2）对于变量 B 而言，我们给定哪一个或者两个变量的值，能够使得该变量条件独立于贝叶斯网络中的其他变量？

Solution. We should fix two parameters A and E. □

（3）如果我们给定变量 A 的值，其他的变量 (B, E, J, M) 之间有哪些是条件独立的？如果我们不给定变量 A 的值呢？注意，贝叶斯网络间两个节点条件独立可以写为 $X_i \perp X_j$。

Solution. When A is fixed:$B \perp M$, $B \perp J$, $E \perp J$, $E \perp M$, $M \perp J$. When A isn't fixed: $B \perp E$. □

（4）我们希望求解贝叶斯网络中 $J = j, M = m, B = b$ 的概率，即 $P(B = b_0, J = j_0, M = m_0)$。请根据联合概率的表达式列出计算 $P(B = b_0, J = j_0, M = m_0)$ 的表达式。除此以外，请列出使用消元法求解 $P(B = b_0, J = j_0, M = m_0)$ 的过程。为了方便，我们要求消元的顺序是

$M \to J \to A \to E \to B$。

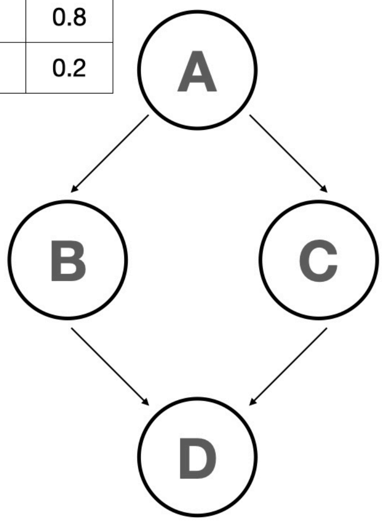
Solution. $P(B=b_0, J=j_0, M=m_0) = \sum_{a\in A}\sum_{e\in E} P(B=b_0)P(E=e)P(A=a|B=b_0, E=e)P(J=j_0|A=a)P(M=m_0|A=a)$.
$P(B,E,A,J,M) = P(B)P(E)P(A|B,E)P(J|A)P(M|A) = P(B)P(E)P(A|B,E)P(J|A)\phi_M(A) = P(B)P(E)P(A|B,E)\phi_J(A)\phi_M(A) = P(B)P(E)\phi_A(B,E) = P(B)\phi_E(B) = \phi(B)$. □

5 第四题

给定如下图2所示的贝叶斯网络模型。我们希望从给定的模型中通过取样的方式进行一些概率的估算，且我们规定采样时该模型各节点的拓扑顺序为 $(A \to B \to C \to D)$

A	P(A)
+a	0.8
-a	0.2

A	B	P(B|A)
+a	+b	0.8
+a	-b	0.2
-a	+b	0.5
-a	-b	0.5



B	C	D	P(D|B,C)
+b	+c	+d	0.3
+b	+c	-d	0.7
+b	-c	+d	0.1
+b	-c	-d	0.9
-b	+c	+d	0.2
-b	+c	-d	0.8
-b	-c	+d	0.9
-b	-c	-d	0.1

A	C	P(C|A)
+a	+c	0.7
+a	-c	0.3
-a	+c	0.1
-a	-c	0.9

图 2: 第四题的贝叶斯网络

示例：采用先验采样（prior sample）的方法生成样本。给定采样过程中的随机数为 $(0.31 \to 0.58 \to 0.04 \to 0.94 \to 0.67 \to 0.49 \to 0.37 \to 0.42)$。

- 首先我们取随机数 $r = 0.31$，而 $r = 0.31 < P(+a) = 0.8$，因此对于节点 A 我们取 $+a$
- 接着随机数 $r = 0.58 < P(+b|+a) = 0.8$，因此对于节点 B 我们取 $+b$
- 随机数 $r = 0.04 < P(+c|+a) = 0.7$，因此对于节点 B 我们取 $+c$
- 随机数 $r = 0.94 > P(+d|+b,+c) = 0.3$，因此对于节点 D 我们取 $-d$。这样我们通过一次采样得到了一个样本 $(+a,+b,+c,-d)$
- 随机数 $r = 0.67 < P(+a) = 0.8$，因此对于节点 A 我们取 $+a$
- 随机数 $r = 0.49 < P(+b|+a) = 0.8$，因此对于节点 B 我们取 $+b$

- 随机数 $r = 0.37 < P(+c|+a) = 0.7$，因此对于节点 C 我们取 $+c$
- 随机数 $r = 0.42 > P(+d|+b, +c) = 0.3$，因此对于节点 D 我们取 $-d$。这样我们通过一次取样获得了一个新的样本 $(+a, +b, +c, -d)$
- 采用了八个随机数，进行了两次采样，得到了两个样本，均为 $(+a, +b, +c, -d)$

（1）采用拒绝采样的方法（rejection sampling）计算 $P(-d|-b)$，请参照模型上方的示例写出采样过程（包括最后进行了几次采样，得到了怎样的样本）。采样时的随机数为 $(0.31 \to 0.58 \to 0.04 \to 0.94 \to 0.67 \to 0.49 \to 0.37 \to 0.42)$，且规定当随机数 $r < P(+a)$ 时采样 $+a$, 当 $r \geq P(+a)$ 时采样 $-a$，随机数用完以后即废弃、当所有随机数用完整个算法停止。

Solution. Use rejection sampling and calculate $P(-d|-b)$.

- 首先我们取随机数 $r = 0.31$，而 $r = 0.31 < P(+a) = 0.8$，因此对于节点 A 我们取 $+a$
- 接着随机数 $r = 0.58 < P(+b|+a) = 0.8$，因此对于节点 B 我们取 $+b$, 与证据不符，返回
- 随机数 $r = 0.04 < P(+a) = 0.8$，因此对于节点 A 我们取 $+a$
- 随机数 $r = 0.94 > P(+b|+a) = 0.8$，因此对于节点 B 我们取 $-b$。与证据相符，继续
- 随机数 $r = 0.67 < P(+c|+a) = 0.7$，因此对于节点 C 我们取 $+c$
- 随机数 $r = 0.49 > P(+d|-b, +c) = 0.2$，因此对于节点 D 我们取 $-d$, 从而我们得到一个样本：$(+a, -b, +c, -d)$.
- 随机数 $r = 0.37 < P(+a) = 0.8$，因此对于节点 A 我们取 $+a$
- 随机数 $r = 0.42 < P(+b|+a) = 0.8$，因此对于节点 B 我们取 $+b$。与证据不符，返回
- 采用了八个随机数,进行了三次采样,得到了一个样本,为 $(+a, -b, +c, -d)$. 所以此时 $P(-d|-b) = 1$.

□

（2）采用似然采样的方法（likelihood weighting sampling）计算 $P(-d|-b)$，请参照模型上方的示例写出采样过程（包括最后进行了几次采样，得到了怎样的样本）。采样时的随机数为 $(0.31 \to 0.58 \to 0.04 \to 0.94 \to 0.67 \to 0.49)$，且规定当随机数 $r < P(+a)$ 时采样 $+a$, 当 $r \geq P(+a)$ 时采样 $-a$，随机数用完以后即废弃、当所有随机数用完整个算法停止。

Solution. Use likelihood weighting sampling and calculate $P(-d|-b)$.

- 首先我们取随机数 $r = 0.31$，而 $r = 0.31 < P(+a) = 0.8$，因此对于节点 A 我们取 $+a$
- 由于 B 是证据，所以对于节点 B 我们取 $-b$,$w = 1 * 0.8 = 0.8$.
- 接着随机数 $r = 0.58 < P(+c|+a) = 0.7$，因此对于节点 C 我们取 $+c$.
- 随机数 $r = 0.04 < P(+d|-b, +c) = 0.2$，因此对于节点 D 我们取 $+d$, 这样我们得到一个样本 $(+a, -b, +c, +d, 0.8)$.
- 随机数 $r = 0.94 > P(+a) = 0.8$，因此对于节点 A 我们取 $-a$。
- 由于 B 是证据，所以对于节点 B 我们取 $-b$,$w = 1 * 0.8 = 0.8$.

- 随机数 $r = 0.67 < P(+c|+a) = 0.7$，因此对于节点 C 我们取 $+c$
- 随机数 $r = 0.49 > P(+d|-b,+c) = 0.2$，因此对于节点 D 我们取 $-d$, 从而我们得到一个样本：$(-a,-b,+c,-d\ 0.8)$.
- 采用了六个随机数,进行了两次采样,得到了两个样本,为 $(+a,-b,+c,+d\ 0.8)$ 和 $(-a,-b,+c,-d\ 0.8)$. 所以此时 $P(-d|-b) = 0.8/(0.8+0.8) = 0.5$.

□

（3）采用吉布斯采样的方法（Gibbs sampling）计算 $P(-d|-b)$，请参照模型上方的示例写出采样过程（包括最后进行了几次采样，得到了怎样的样本）。采样时的随机数为 $(0.31 \to 0.58 \to 0.04 \to 0.94 \to 0.67 \to 0.49)$，采样初时刻的节点取值初始化为 $(+a,-b,+c,+d)$，且规定当随机数 $r < P(+a)$ 时采样 $+a$, 当 $r \geq P(+a)$ 时采样 $-a$，随机数用完以后即废弃、当所有随机数用完整个算法停止。

Solution. Use Gibbs sampling and calculate $P(-d|-b)$.

- 首先我们取随机数 $r = 0.31$，而 $r = 0.31 < P(+a) = 0.8$，因此对于节点 A 我们取 $+a$, 样本变为 $(+a,-b,+c,+d)$.
- 接着随机数 $r = 0.58 < P(+b|+a) = 0.8$,因此对于节点 B 我们取 $+b$,样本变为 $(+a,+b,+c,+d)$.
- 随机数 $r = 0.04 < P(+c|+a) = 0.2$，因此对于节点 C 我们取 $+c$, 这样我们得到一个样本 $(+a,+b,+c,+d)$.
- 随机数 $r = 0.94 > P(+d|+b,+c) = 0.3$,因此对于节点 D 我们取 $-d$, 样本变为 $(+a,+b,+c,-d)$.
- 随机数 $r = 0.67 < P(+a) = 0.8$，因此对于节点 A 我们取 $+a$, 样本变为 $(+a,+b,+c,-d)$.
- 随机数 $r = 0.49 < P(+b|+a) = 0.8$，因此对于节点 B 我们取 $+b$, 样本变为 $(+a,+b,+c,-d)$.
- 采用了六个随机数,进行了两次采样,得到了六个样本,为 $(+a,-b,+c,+d)$,$(+a,+b,+c,+d)$，$(+a,+b,+c,+d)$，$(+a,+b,+c,-d)$，$(+a,+b,+c,-d)$，$(+a,+b,+c,-d)$. 所以 $P(-d|-b) = 0$.

□

6 第五题

给定一个隐马尔可夫模型（HMM）。

（1）运用 HMM 的建模方式，用条件概率的计算方式推导并化简 $P(x_1,...,x_t,y_{t-1}=s_v,y_t=s_j)$。

提示：一般来说，我们会对 HMM 模型有以下的建模方式（与课件上一致)——x_t 表示 t 时刻的观察状态，y_t 表示 t 时刻的隐藏状态, 隐藏状态可能取值为 $\{1,2,3,\cdots,M\}$。初始状态的概率（start probabilities）表示为 $\{\pi_1,\cdots,\pi_M\}$，隐藏状态之间从状态 i 转化为状态 j 的转移概率（transition probabilities）为 $a_{i,j}$，隐藏状态 y_j 与观察状态 x_i 之间的发散概率（emission probabilities)为 $b_j(x_i)$。在使用前向算法推导序列概率的时候,设定的前向因子 $\alpha_t^i = P(x_1,...,x_t,y_t=s_i)$。

Solution. $P(x_1,...,x_t,y_{t-1}=s_v,y_t=s_j) = \alpha_{t-1}^v P(y_t=s_j|x_1,x_2,\ldots x_{t-1},y_{t-1}=s_v)P(x_t|x_1,x_2,\ldots,x_{t-1},y_{t-1}= s_v,y_t=s_j) = \alpha_{t-1}^v P(y_t=s_j|y_{t-1}=s_v)P(x_t|y_t=s_j) = \alpha_{t-1}^v a_{v,j} b_j(x_t)$. □

（2）我们给定的 HMM 模型如下图3所示。假定图中的圆圈表示 A, B, C 三种可能的隐藏状态，而圆圈中的数字表示特定隐藏状态下的观察状态以及其发射概率（emission probabilities）。圆圈之间的箭头表示隐藏状态的转移以及相应的转移概率（transition probabilities）。我们用 $p_{y=A}^t$ 表示在 t 时刻，隐状态为 A 的概率，用 $p_{x=1}^t$ 表示在 t 时刻，观察状态为 1 的概率。为了简便，我们规定初始的隐状态为 A，即 $p_A^1 = 1$。在每个时刻，隐状态会通过转移概率确定下个时刻的状态。

(a) 请用图中给定的数值与符号计算 $p_{y=C}^3$。

Solution. 隐藏状态转移矩阵为

$$\begin{bmatrix} A1 & 0 & A2 \\ B2 & B1 & 0 \\ C2 & C1 & 0 \end{bmatrix}$$

所以在 $t=3$ 时的隐藏状态概率矩阵为

$$\begin{bmatrix} A1 & 0 & A2 \\ B2 & B1 & 0 \\ C2 & C1 & 0 \end{bmatrix}^2 * [1,0,0]^T = \left[A1^2 + A2C2, (A1+B1)B2, A1C2 + B2C1\right]^T$$

所以，$p_{y=C}^3 = A1C2 + B2C1.$ □

(b) 假设在 t 时刻，$(p_{y=B}^t, p_{y=C}^t)$ 分别为 (b_0, c_0), 请用图中给定的数值与符号计算 $p_{x=2}^t$。

Solution. 在 t 时刻，$p_{y=A}^t = 1 - b_0 - c_0, p_{x=2}^t = p_{y=A}^t * b_A(x=2) = 0.2(1 - b_0 - c_0).$ □

7 作业反馈

点击访问链接https://www.wjx.cn/vj/tMDNNZ5.aspx或者扫描下方的二维码就可以反馈意见。

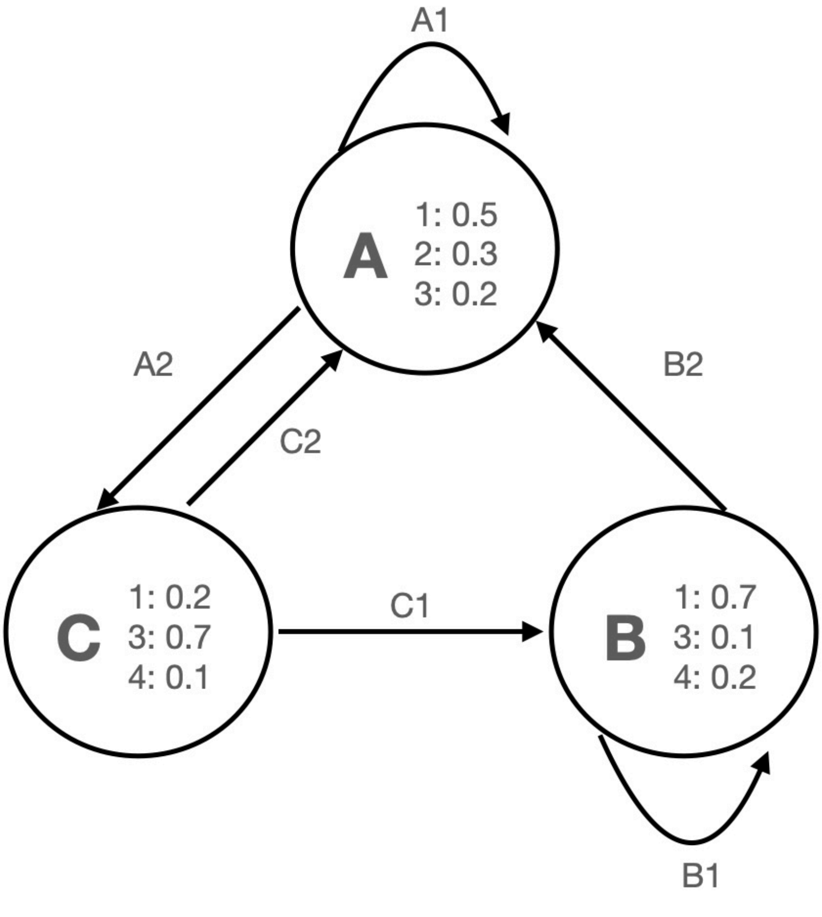


图 3: 第四题第二小问的 HMM 模型

图 4: 作业调查问卷