Artificial Intelligence, Spring 2017

Homework 4 – PGM

Xinglu Wang 3140102282 ISEE 1403, ZJU

1 Problem 1

a. Statements in (ii), (iii) are asserted by BN structure.

- **(i)** Not asserted by BN, $\mathbf{P}(B, I, M) = \mathbf{P}(B)\mathbf{P}(M)\mathbf{P}(I|B, M)$ instead.
- **(ii)** True assertion. According to Causal Reasoning Pattern, $J \perp I | G$, so $\mathbf{P}(J|G) = \mathbf{P}(J|G, I)$.
- **(iii)** True assertion. M's Markov blanket is G, B, I, so $\mathbf{P}(M|G, B, I) = \mathbf{P}(M|G, B, I, J)$.

b. $P(b, i, \neg m, g, j) = P(b)P(\neg m)P(i|b, \neg m)P(g|b, i, \neg m)P(j|g) = \underline{0.2916}$

c. Select Entries consistent with $(b, i, m) = (t, t, t)$.

B	I	M	$P(G)$
t	t	t	$.9$

$$\mathbf{P}(J|b, i, m) \propto \phi_0(J)$$
$$= \sum_G \phi_1(G)\phi_2(J, G)$$

Sum Out

$$= \begin{array}{cc} G = t & G = f \\ \hline 0.81 & 0.09 \\ 0 & 0.1 \end{array}$$

$$=< 0.81, 0.19 >$$

Normalization $\mathbf{P}(J = j|b, i, m) = {}^{\phi_0(J=j)}/_{\sum_J \phi_0(J)} = 0.81$

2 Problem 2

a. Use variable elimination algorithm:

$$\begin{aligned}
\mathbf{P}(B|j, m) &\propto \sum_{A,E} \mathbf{P}(B, A, E|j, m) \\
&= \sum_{A,E} \mathbf{P}(B)\mathbf{P}(E)\mathbf{P}(A|B, E)\mathbf{P}(j|A)\mathbf{P}(m|A) \\
&= \sum_E \mathbf{P}(B)\mathbf{P}(E) \sum_A \mathbf{P}(A|B, E)\mathbf{P}(j|A)\mathbf{P}(m|A) \\
&= \mathbf{P}(B) \sum_E \mathbf{P}(E) \left[.63 \times \begin{array}{c|cc} & B = t & B = f \\ \hline E = t & 0.95 & 0.29 \\ E = f & 0.94 & 0.01 \end{array} + .005 \times \begin{array}{c|cc} & B = t & B = f \\ \hline E = t & .05 & .71 \\ E = f & .06 & .999 \end{array} \right] \\
&= \mathbf{P}(B) \sum_E \mathbf{P}(E) \begin{array}{c|cc} & B = t & B = f \\ \hline E = t & .598525 & .183055 \\ E = f & .59223 & .0011295 \end{array} \\
&= \mathbf{P}(B) \sum_E \mathbf{P}(E)\phi_1(B, E) \\
&= \mathbf{P}(B)\phi_2(B) \\
&= \begin{array}{c|c} B = t & .00059224259 \\ B = f & .0014918576 \end{array}
\end{aligned}$$

After normalizing, $\mathbf{P}(B|j, m) \approx < .284, .716 >$

b. Variable elimination has 7 additions, 16 multiplications, and 2 divisions.

If we use enumeration,

$$\begin{aligned}\mathbf{P}(B|j,m) &\propto \textstyle\sum_{A,E}\mathbf{P}(B,A,E|j,m)\\ &= \textstyle\sum_{A,E}\mathbf{P}(B)\mathbf{P}(E)\mathbf{P}(A|B,E)\mathbf{P}(j|A)\mathbf{P}(m|A)\\ &= \textstyle\sum_{A,E}\phi_1(A,B,E)\\ &= \textstyle\sum_{A}\phi_2(B,E)\\ &= \phi_3(B)\end{aligned}$$

There are 7 additions, 22 multiplications, and 2 divisions.
c. Using enumeration,

$$\begin{aligned}\mathbf{P}(\mathbf{X}_1|x_n^t) &= \textstyle\sum_{\mathbf{X}_2\cdots\mathbf{X}_{n-1}}\mathbf{P}(\mathbf{X}_1)\mathbf{P}(\mathbf{X}_2|\mathbf{X}_1)\cdots\mathbf{P}(x_n^t|\mathbf{X}_{n-1})\\ &= \textstyle\sum_{\mathbf{X}_2\cdots\mathbf{X}_{n-1}}\phi_1(\mathbf{X}_1,\cdots\mathbf{X}_{n-1})\end{aligned}$$

The largest condition probability table is of size 2^{n-1}, thus the complexity is $O(x^n)$

Using variable elimination, the largest conditional probability table can concern 2 variables at most, since $Parent(\mathbf{X}_i) = \mathbf{X}_{i-1}$. Thus, for each step, complexity is $O(4)$ and there are $n-1$ steps. Thus, the overall complexity is $O(4(n-1)) = O(n)$.
d. We prove it by induction. Base case: when $n = 1$, complexity is $O(n)$.

Assume *a polytree network with n nodes has $O(n)$ complexity*. We need to prove a polytree network with $n+1$ nodes has $O(n+1)$ complexity.

We will choose a leaf node, since the variable ordering should be consistent with the network structure. Eliminate this node is constant complexity dependent on the size of CPT concerning $\mathbf{X}_{n+1}$ and $Parent(\mathbf{X}_{n+1})$ and we can write as $O(1)$ in the meaning of complexity. Thus, the overall complexity is $O(x) + O(1) = O(x+1)$. Thus, our assumption – *a polytree network with n nodes has $O(n)$ complexity* is true.

3 Problem 3

Model as Dynamic Bayesian Network, ref. to fig 1 on the following page. The transition probability is given by:

$$\begin{aligned}P(s_0) &= 0.7\\ P(s_{t+1}|s_t) &= 0.8\\ P(s_{t+1}|\neg s_t) &= 0.3\\ P(r_t|s_t) &= 0.2\\ P(r_t|\neg s_t) &= 0.7\\ P(c_t|s_t) &= 0.1\\ P(c_t|\neg s_t) &= 0.3\end{aligned}$$

Model as Hidden Markov Model, ref to fig 2 on the next page. The variable R_t and C_t is combined as RC_t. The conditional probability table becomes

$RC_t =$	t,t	t,f	f,t	f,f
$S_t = t$	0.02	0.18	0.08	0.72
$S_t = f$	0.21	0.49	0.09	0.21

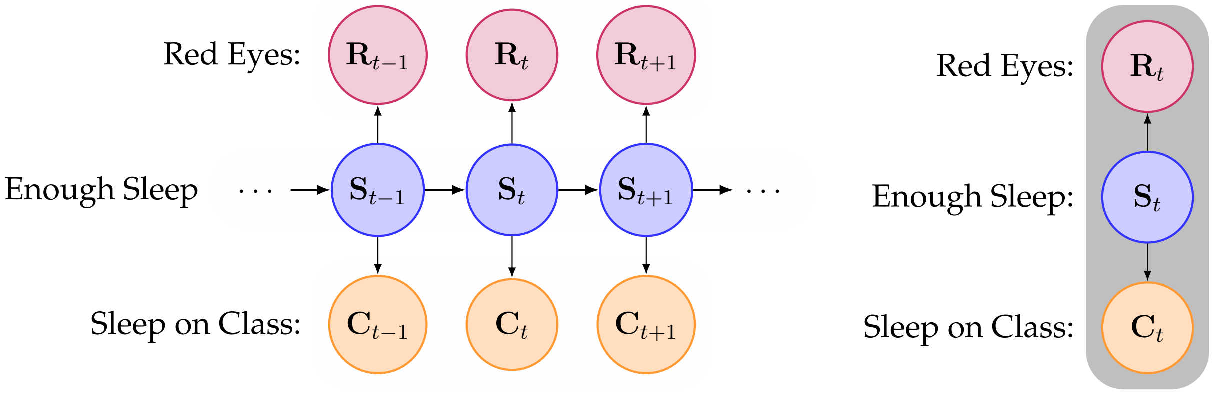

Figure 1: Dynamic Bayesian Network. **Right**: Express by plate model

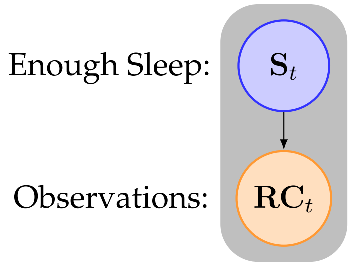

Figure 2: Hidden Markov Model, expressed by plate model

4 Problem 4

a. Using frequency to estimate probability:

Y	$\mathbf{P}(Y)$
-1	1/2
1	1/2

X_1	$\mathbf{P}(X_1\|Y=-1)$	$\mathbf{P}(X_1\|Y=1)$
0	1/3	2/3
1	2/3	1/3

X_2	$\mathbf{P}(X_2\|Y=-1)$	$\mathbf{P}(X_2\|Y=1)$
0	2/3	2/3
1	1/3	1/3

b. Add k (Laplace) smoothing, with $k = 1$,

$$\mathbf{P}(X_i|Y) = \frac{N_i + k}{N + k \times 2} \qquad (i = 1, 2)$$

Y	$\mathbf{P}(Y)$
-1	1/2
1	1/2

X_1	$\mathbf{P}(X_1\|Y=-1)$	$\mathbf{P}(X_1\|Y=1)$
0	2/5	3/5
1	3/5	2/5

X_2	$\mathbf{P}(X_2\|Y=-1)$	$\mathbf{P}(X_2\|Y=1)$
0	3/5	3/5
1	2/5	2/5

c.

$$\begin{aligned}\mathbf{P}(Y|X_1=0,X_2=0) &= {}^{\mathbf{P}(Y,X_1=0,X_2=0)}/_{\mathbf{P}(X_1=0,X_2=0)} \\ &\propto \mathbf{P}(Y)\mathbf{P}(X_1=0|Y)\mathbf{P}(X_2=0|Y) \\ &\text{(Since the evidence } \mathbf{P}(X_1=0,X_2=0) \text{ is just a normalization factor)} \\ &=< {}^1/_2 \times {}^2/_5 \times {}^3/_5, {}^1/_2 \times {}^3/_5 \times {}^3/_5 >\end{aligned}$$

After normalization, $\mathbf{P}(Y|X_1=0,X_2=0) =< 0.4, 0.6 >$.

d. When $k \to \infty$, the prior is dominant.

$$\mathbf{P}(X_i|Y) = \lim_{k\to\infty} \frac{N_i + k}{N + k \times 2} \qquad (i = 1, 2) = 1/2$$

All value of table in (b). will be 1/2. Thus, $\mathbf{P}(Y|X_1=0,X_2=0) =< 0.5, 0.5 >$

e. The following feature set can consist a linear binary classifier.

v. with decision boundary $abs(X_1 - X_2) = 1/2$

vii. with decision boundary(not optimal) $2\max(X_1, X_2) - X_1 - X_2 = \epsilon$, where $\epsilon \to 0^+$

viii. with decision boundary $I(X_1 - X_2) = 1/2$

5 Problem 5

In space of X_1, X_1X_2, the margin is $X_1X_2 = 0$, draw the separating line back to origin Euclidean input space, there are two lines $X_1 = 0$ and $X_2 = 0$.