专业：信息工程
姓名：徐晓刚
学院：信电学院
学号：3140102480
日期：May 17, 2017
地点：玉泉 5 舍

Artificial intelligence Homework4

Problem 1

14.14

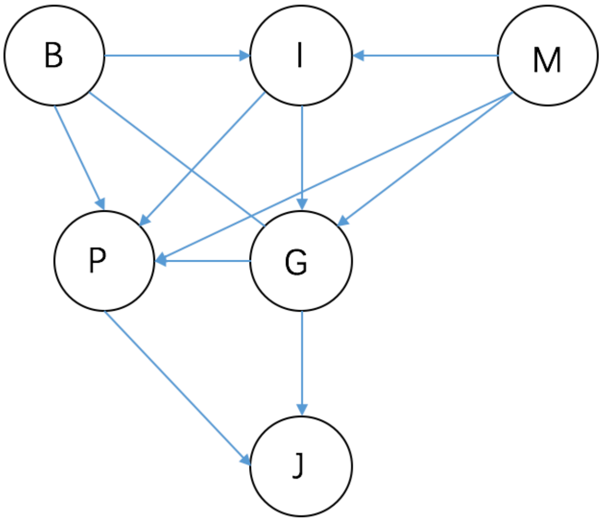
- **a.** For (i), we can see that B,I,M can influence each other so they will never be independent. For (ii), G is the only parent of J, so the network assert (ii). For (iii), G,B,I is the Markov blanket of M, so the network assert (iii).

- **b.** According to the independent relation in the network, we can get the equation:
$\boldsymbol{P(b,i,\neg m,g,j) = P(b)P(\neg m)P(i|b,\neg m)P(g|b,i,\neg m)P(j|g) = 0.2916}$

- **c.** Since B,I,M are fixed true in the evidence we can treat G as having a prior of 0.9 :
$\boldsymbol{P(J|b,i,m) = \alpha \sum_g P(J,g) = \alpha[P(J,g) + P(J,\neg g)]}$
$\boldsymbol{= < 0.81, 0.19 >}$
So the probability of going to jail is 0.81

- **d.**Intuitively, a person cannot be found guilty if not indicted, regardless of whether they broke the law and regardless of the prosecutor. so G is context-specifically independent of B and M given I = false.

- **e.** It's obvious that I and G are parents of P . Because a pardon is more likely if the person is actually innocent and if the prosecutor is politically motivated, so B and M as parents of P. The pardon may not be jailed, so P is a parent of J. The new network can be see as following:



The new network

Problem 2

14.15

- **a.** $P(B|j,m)$
  $= \alpha P(B)\sum_e P(e)\sum_a P(a|b,e)P(j|a)P(m|a)$
  $= \alpha P(B)\sum_e P(e)[0.9\times 0.7\times(\begin{pmatrix}0.95 & 0.29\\ 0.94 & 0.001\end{pmatrix})+0.05\times 0.01\times(\begin{pmatrix}0.05 & 0.71\\ 0.06 & 0.999\end{pmatrix})]$
  $= \alpha P(B)\sum_e P(e)(\begin{pmatrix}0.598525 & 0.183055\\ 0.59223 & 0.0011295\end{pmatrix})$
  $= \alpha(\begin{pmatrix}0.001\\ 0.999\end{pmatrix})\times(\begin{pmatrix}0.59224259\\ 0.001493351\end{pmatrix})$
  $=< 0.284, 0.716 >$

- **b.** There are 7 additions, 16 multiplications, and 2 divisions. The enumeration algorithm has two extra multiplications.

- **c.** If we compute $P(X_1|X_n = true)$ using enumeration, we have to evaluate two complete binary trees, each of depth n-1, so the tota work is $O(2^n)$.

If we use the method of variable elimination instead, the factors never grow beyond two variables. The total work is $O(n)$.

- **d.** Prove:

  First, based on the inductive hypothesis, we can assume that any polytree with n nodes can be evaluated in time proportional to the size of the polytree .

  Next, consider a polytree with n + 1 nodes. To eliminate any leaf node, we have to do work proportional to the size of its CPT. Then, because the network is a polytree, we are left with independent subproblems, one for each parent.

  Each subproblem takes total work proportional to the sum of its CPT sizes, so the total work for n + 1 nodes is proportional to the sum of CPT sizes

Problem 3

15.13

We can see that there are three variables:
S_t, whether the student gets enough sleep;
R_t, whether they have red eyes in the class;
C_t, whether the student sleeps in class. It's obvious that S_t is a parent of S_{t+1}, R_t, C_t
The CPT are given as following:
$P(s_0) = 0.7, P(s_{t+1}|s_t) = 0.8, P(s_{t+1}|\neg s_t) = 0.3,$
$P(r_t|s_t) = 0.2, P(r_t|\neg s_t) = 0.7,$
$P(c_t|s_t) = 0.1, P(c_t|\neg s_t)$
To reformulate as an HMM with a single observation node, simply combine the 2-valued variables "having red eyes"and "sleeping in class"into a single 4-valued variable. The Probability table are as following :

S_0	$P(S_0)$
s	0.7
$\neg s$	0.3

R_t	S_t	$P(R_t\|S_t)$
r	s	0.2
$\neg r$	s	0.8
r	$\neg s$	0.7
$\neg r$	$\neg s$	0.3

S_{t+1}	S_t	$P(S_{t+1}\|S_t)$
s_{t+1}	s_t	0.8
$\neg s_{t+1}$	s_t	0.2
s_{t+1}	$\neg s_t$	0.3
$\neg s_{t+1}$	$\neg s_t$	0.7

C_t	S_t	$P(C_t\|S_t)$
c	s	0.1
$\neg c$	s	0.9
c	$\neg s$	0.3
$\neg c$	$\neg s$	0.7

If we reformulate this model into HMM, we can set the observation as O_t

O_t	S_t	$P(O_t\|S_t)$
r, c	s	0.02
$r, \neg c$	s	0.18
$\neg r, c$	s	0.08
$\neg r, \neg c$	s	0.72
r, c	$\neg s$	0.21
$r, \neg c$	$\neg s$	0.49
$\neg r, c$	$\neg s$	0.09
$\neg r, \neg c$	$\neg s$	0.21

Problem4

a.

Y	$P(Y)$
-1	0.5
$+1$	0.5

X_1	$P(X_1\|Y=-1)$	$P(X_1\|Y=+1)$
0	$1/3$	2/3
1	$2/3$	1/3

X_2	$P(X_2\|Y=-1)$	$P(X_2\|Y=+1)$
0	$2/3$	2/3
1	$1/3$	1/3

b.

In this problem we assume k = 1

Y	$P(Y)$
-1	0.5
$+1$	0.5

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

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

c.

$$\boldsymbol{P(Y, X_1 = 0, X_2 = 0)}$$
$$= \begin{pmatrix} \boldsymbol{P(Y = +1) \times P(X_1 = 0|Y = +1) \times P(X_2 = 0|Y = +1)} \\ \boldsymbol{P(Y = -1) \times P(X_1 = 0|Y = -1) \times P(X_2 = 0|Y = -1)} \end{pmatrix} = \begin{pmatrix} \boldsymbol{0.18} \\ \boldsymbol{0.12} \end{pmatrix}$$
$$\boldsymbol{P(X_1 = 0, X_2 = 0) = \frac{2+1}{6+4} = 0.3}$$
$$\boldsymbol{P(Y|X_1 = 0, X_2 = 0) =< 0.6, 0.4 >}$$

d.

$$\text{As } \boldsymbol{k \to \infty}$$
$$\boldsymbol{P(Y, X_1 = 0, X_2 = 0) = \begin{pmatrix} 1/8 \\ 1/8 \end{pmatrix}}$$
$$\boldsymbol{P(X_1 = 0, X_2 = 0) = \frac{2+k}{6+4k} = 1/4}$$
$$\boldsymbol{P(Y|X_1 = 0, X_2 = 0) =< 0.5, 0.5 >}$$

e.

The sets that would enable a linear binary classifier include (v),(viii)

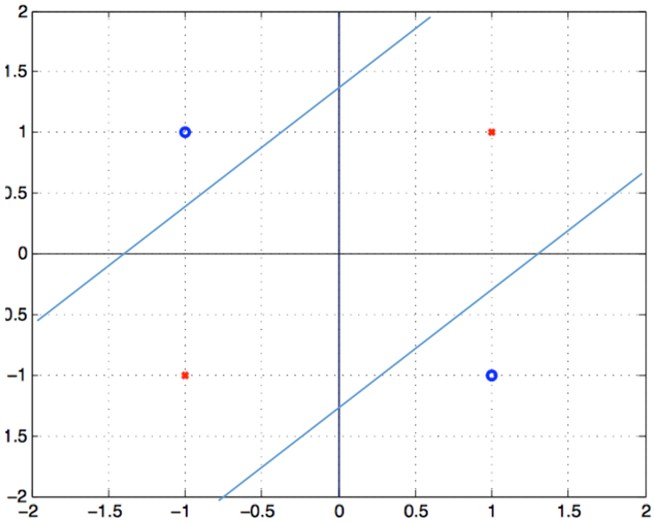
Problem5

We first draw the four examples using axes x1, x2, denoting class by +/-.
Label them (p1, ..., p4) .
we will set a transformation that will make these points linearly separable:

$$\boldsymbol{y_1 = x_1,}$$
$$\boldsymbol{y_2 = (x_1 - x_2)^2}$$

The margin is 2. For the transformation above, the max-margin separator is at y2 = 2.
We get the separating line :

$$\boldsymbol{x_1 - x_2 = \sqrt{2}} \text{ or } \boldsymbol{x_1 - x_2 = -\sqrt{2}}$$


2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
-2
-1.5
-1
-0.5
0
0.5
1
1.5
2