Barry Tee Wei Cong 1001549 01.112 Machine Learning

HW5.

Date No.

1. Without knowing the actual value of any node, X_1 and X_6 are independent of each other. There is no open path from X_1 to X_6 as justified by Bayes' Ball Algorithm.

Knowing node X_7 and X_{10} values, X_1 and X_6 are dependent on each other. There is an open path: $X_1 - X_2 - X_3 - X_4 - X_5 - X_8 - X_9 - X_{10} - X_9 - X_6$ using Bayes' Ball Algorithm.

Date No.

2. If all nodes can take 2 different values, number of free parameters is:

$1(X_1) + 2\times1(X_2) + 2\times1(X_3) + 2\times1(X_4) + 2\times1(X_5) + 1(X_6) + 2\times1(X_7) + 2\times1(X_8)$
$+ 2\times2\times2\times1(X_9) + 2\times1(X_{10}) + 2\times1(X_{11})$
$= 1 + 2 + 2 + 2 + 2 + 1 + 2 + 2$
$+ 8 + 2 + 2$
$= 26$

If X_3 and X_9 take 4 different values, all other nodes take 3 diff values, number of free parameters is:

$2(X_1) + 3\times2(X_2) + 3\times3(X_3) + 4\times2(X_4) + 3\times2(X_5) + 2(X_6) + 3\times2(X_7) + 3\times2(X_8)$
$+ 3\times3\times3\times3(X_9) + 4\times2(X_{10}) + 3\times2(X_{11})$
$= 2 + 6 + 9 + 8 + 6 + 2 + 6 + 6$
$+ 81 + 8 + 6$
$= 140$

Date No.

3a $P(X_3=2 \mid X_4=1)$

$= \frac{P(X_3=2, X_4=1)}{P(\quad X_4=1)}$

$= \frac{P(X_3=2) \cdot P(X_4=1 \mid X_3=2)}{P(\quad X_4=1)}$

$= \frac{0.7 \cdot 0.5}{P(X_4=1 \mid X_3=1)P(X_3=1) + P(X_4=1 \mid X_3=2)P(X_3=2)}$, X_3 independent of x_2 based on table values.

$= \frac{0.7 \cdot 0.5}{0.1 \cdot 0.3 + 0.5 \cdot 0.7}$

$= 0.92$

b. $P(X_5=2 \mid X_3=1, X_{11}=2, X_1=1)$

$= P(X_5=2 \mid X_3=1)$, x_{10} independent of x_9 based on table values.

$= \frac{P(X_5=2, X_3=1)}{P(\quad X_3=1)}$ ↓

$= \frac{P(X_3=1)P(X_5=2 \mid X_4)P(X_4 \mid X_3=1)}{0.3}$ x_{11} independent of x_5

$= \frac{P(X_3=1)P(X_4=1 \mid X_3=1)P(X_5=2 \mid X_4=1) + P(X_3=1)P(X_4=2 \mid X_3=1)P(X_5=2 \mid X_4=2)}{0.3}$ X_3 independent of X_2

↓

$= \frac{0.3 \cdot 0.1 \cdot 0.5 + 0.3 \cdot 0.9 \cdot 0.4}{0.3}$ X_2 independent of x_1

$= \frac{0.123}{0.3}$ ↓

X_5 independent of x_1

$= 0.41$

but dependent on x_3.

Scanned by CamScanner

Date No.

4a. $\theta_{x_2(1),x_3(1)} = \frac{count(x_2=1, x_3=1)}{count(x_2=1)} = \frac{2}{5}$

$\theta_{x_2(1),x_3(2)} = 1-\frac{2}{5} = \frac{3}{5}$

$\theta_{x_2(2),x_3(1)} = \frac{count(x_2=2, x_3=1)}{count(x_2=2)} = \frac{5}{7}$

$\theta_{x_2(2),x_3(2)} = 1-\frac{5}{7} = \frac{2}{7}$

	x_3	
x_2	1	2
1	2/5	3/5
2	5/7	2/7

111 $\theta_{x_6(1),x_7(1),x_8(1),x_9(1)} = \frac{count(x_6=1, x_7=1, x_8=1, x_9=1)}{count(x_6=1, x_7=1, x_8=1)} = \frac{2}{3}$

$\theta_{x_6(1),x_7(1),x_8(1),x_9(2)} = 1-\frac{2}{3} = \frac{1}{3}$

112 $\theta_{x_6(1),x_7(1),x_8(2),x_9(1)} = \frac{count(x_6=1, x_7=1, x_8=2, x_9=1)}{count(x_6=1, x_7=1, x_8=2)} = \frac{1}{1} = 1$

$\theta_{x_6(1),x_7(1),x_8(2),x_9(2)} = 1-1=0$

121 $\theta_{x_6(1),x_7(2),x_8(1),x_9(1)} = \frac{count(x_6=1, x_7=2, x_8=1, x_9=1)}{count(x_6=1, x_7=2, x_8=1)} = 0$

$\theta_{x_6(1),x_7(2),x_8(1),x_9(2)} = 1-0=1$

122 $\theta_{x_6(1),x_7(2),x_8(2),x_9(1)} = \frac{count(x_6=1, x_7=2, x_8=2, x_9=1)}{count(x_6=1, x_7=2, x_8=2)} = \frac{1}{2}$

$\theta_{x_6(1),x_7(2),x_8(2),x_9(2)} = 1-\frac{1}{2} = \frac{1}{2}$

211 $\theta_{x_6(2),x_7(1),x_8(1),x_9(1)} = \frac{count(x_6=2, x_7=1, x_8=1, x_9=1)}{count(x_6=2, x_7=1, x_8=1)} = \frac{1}{2}$

$\theta_{x_6(2),x_7(1),x_8(1),x_9(2)} = 1-\frac{1}{2} = \frac{1}{2}$

212 $\theta_{x_6(2),x_7(1),x_8(2),x_9(1)} = \frac{count(x_6=2, x_7=1, x_8=2, x_9=1)}{count(x_6=2, x_7=1, x_8=2)} = \frac{0}{1} = 0$

$\theta_{x_6(2),x_7(1),x_8(2),x_9(2)} = 1-0=1$

221 $\theta_{x_6(2),x_7(2),x_8(1),x_9(1)} = \frac{count(x_6=2, x_7=2, x_8=1, x_9=1)}{count(x_6=2, x_7=2, x_8=1)} = \frac{0}{1} = 0$

$\theta_{x_6(2),x_7(2),x_8(1),x_9(2)} = 1-0=1$

222 $\theta_{x_6(2),x_7(2),x_8(2),x_9(1)} = \frac{count(x_6=2, x_7=2, x_8=2, x_9=1)}{count(x_6=2, x_7=2, x_8=2)} = \frac{1}{1} = 1$

$\theta_{x_6(2),x_7(2),x_8(2),x_9(2)} = 1-1=0$

			x_9	
x_6	x_7	x_8	1	2
1	1	1	2/3	1/3
1	1	2	1	0
1	2	1	0	1
1	2	2	1/2	1/2
2	1	1	1/2	1/2
2	1	2	0	1
2	2	1	0	1
2	2	2	1	0

Date No.

4b.

	X_{11}	
X_{10}	1	2
1	$4/8 = 1/2$	$4/8 = 1/2$
2	$2/4 = 1/2$	$2/4 = 1/2$

X_{11} independent of X_{10} → Removing X_{10} does not affect the Bayesian Network Log Likelihood

However, the new Bayesian network needs less free variables. This leads to higher BIC score

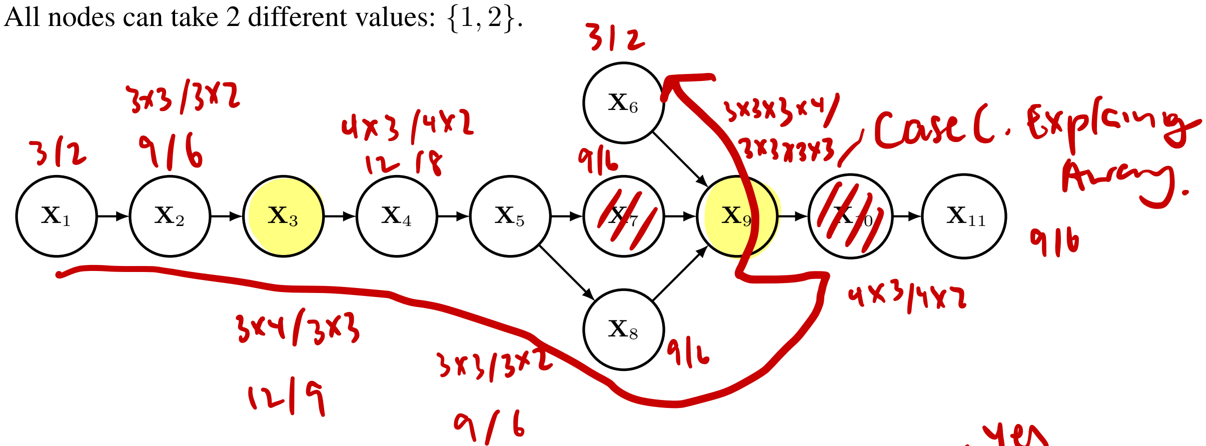
SINGAPORE UNIVERSITY OF
TECHNOLOGY AND DESIGN

01.112 Machine Learning, Spring 2018
Homework 5

Due 20 Apr 2018, 5pm

This homework will be graded by Allan Jie

In this homework, we would like to look at the Bayesian Networks. You are given a Bayesian network as below. All nodes can take 2 different values: $\{1, 2\}$.



- yes

Question 1. Without knowing the actual value of any node, are node $\mathbf{X}_1$ and $\mathbf{X}_6$ independent of each other? What if we know the value of node $\mathbf{X}_7$ and $\mathbf{X}_{10}$? *(5 points)*

-Dependent. $X_1 - X_2 - X_3 - X_4 - X_5 - X_8 - X_9 - X_{10} - X_9 - X_6$

Question 2. What is the *effective* number of parameters needed to for this Bayesian network? What would be the *effective* number of parameters for the same network if node $\mathbf{X}_3$ and $\mathbf{X}_9$ can take 4 different values: $\{1, 2, 3, 4\}$, and all other nodes can only take 3 different values: $\{1, 2, 3\}$? *(5 points)*

Question 3. If we have the following probability tables for the nodes. Compute the following probabilities. Clearly write down all the necessary steps.

(a) Calculate the following conditional probability:

$$P(\mathbf{X}_3 = 2 | \mathbf{X}_4 = 1)$$

(6 points)

(b) Calculate the following conditional probability:

$$P(\mathbf{X}_5 = 2 | \mathbf{X}_3 = 1, \mathbf{X}_{11} = 2, \mathbf{X}_1 = 1)$$

(9 points)

(*Hint: find a short answer. The values in some of the probability tables may reveal some useful information.*)

	$\mathbf{X}_1$	
	1	2
	0.5	0.5

	$\mathbf{X}_2$	
$\mathbf{X}_1$	1	2
1	0.2	0.8
2	0.3	0.7

	$\mathbf{X}_3$	
$\mathbf{X}_2$	1	2
1	0.3	0.7
2	0.3	0.7

	$\mathbf{X}_4$	
$\mathbf{X}_3$	1	2
1	0.1	0.9
2	0.5	0.5

	$\mathbf{X}_5$	
$\mathbf{X}_4$	1	2
1	0.5	0.5
2	0.6	0.4

	$\mathbf{X}_6$	
	1	2
	0.6	0.4

			$\mathbf{X}_9$	
$\mathbf{X}_6$	$\mathbf{X}_7$	$\mathbf{X}_8$	1	2
1	1	1	0.8	0.2
1	1	2	0.1	0.9
1	2	1	0.9	0.1
1	2	2	0.7	0.3
2	1	1	0.3	0.7
2	1	2	0.2	0.8
2	2	1	0.2	0.8
2	2	2	0.9	0.1

	$\mathbf{X}_7$	
$\mathbf{X}_5$	1	2
1	0.2	0.8
2	0.3	0.7

	$\mathbf{X}_8$	
$\mathbf{X}_5$	1	2
1	0.8	0.2
2	0.7	0.3

	$\mathbf{X}_{10}$	
$\mathbf{X}_9$	1	2
1	0.8	0.2
2	0.8	0.2

	$\mathbf{X}_{11}$	
$\mathbf{X}_{10}$	1	2
1	0.7	0.3
2	0.8	0.2

Question 4.

(a) Now, assume we do not have any knowledge about the probability tables for the nodes in the network, but we have the following 12 observations/samples. Find a way to estimate the probability tables associated with the nodes $\mathbf{X}_3$ and $\mathbf{X}_9$ respectively. *(6 points)*

$\mathbf{X}_1$	$\mathbf{X}_2$	$\mathbf{X}_3$	$\mathbf{X}_4$	$\mathbf{X}_5$	$\mathbf{X}_6$	$\mathbf{X}_7$	$\mathbf{X}_8$	$\mathbf{X}_9$	$\mathbf{X}_{10}$	$\mathbf{X}_{11}$
1	1	2	2	2	1	1	1	1	1	1
1	2	1	1	2	1	1	1	1	1	2
2	2	2	1	2	2	1	1	2	2	1
1	1	2	1	2	1	1	2	1	2	2
1	2	1	1	1	1	2	2	2	1	1
2	2	1	2	1	2	2	1	1	1	2
2	1	2	2	1	2	1	2	2	2	1
2	2	2	1	2	1	2	2	1	2	2
1	1	1	1	2	2	1	1	1	1	1
1	1	1	1	2	1	1	1	2	1	2
1	2	1	2	2	1	2	1	2	1	2
2	2	1	2	1	2	2	2	1	1	1

(b) Based on the above observations, you would like to find a good Bayesian network structure to model the data. You started with the initial structure shown on the previous page, and decided to delete the edge between $\mathbf{X}_{10}$ and $\mathbf{X}_{11}$. Is the resulting new structure (after deleting the single edge between $\mathbf{X}_{10}$ and $\mathbf{X}_{11}$ from the original graph) better than the original structure in terms of BIC score? Clearly explain the reason. *(9 points)*

(Hint: Try to find a short answer.)

Question 3. If we have the following probability tables for the nodes. Compute the following probabilities. Clearly write down all the necessary steps.

(a) Calculate the following conditional probability:

$$P(\mathbf{X}_3 = 2|\mathbf{X}_4 = 1)$$

(6 points)

(b) Calculate the following conditional probability:

$$P(\mathbf{X}_5 = 2|\mathbf{X}_3 = 1, \mathbf{X}_{11} = 2, \mathbf{X}_1 = 1)$$

(9 points)

(*Hint: find a short answer. The values in some of the probability tables may reveal some useful information.*)

	$\mathbf{X}_1$ = 1	$\mathbf{X}_1$ = 2
	0.5	0.5

$\mathbf{X}_1$	$\mathbf{X}_2$ = 1	$\mathbf{X}_2$ = 2
1	0.2	0.8
2	0.3	0.7

$\mathbf{X}_2$	$\mathbf{X}_3$ = 1	$\mathbf{X}_3$ = 2
1	0.3	0.7
2	0.3	0.7

$\mathbf{X}_3$	$\mathbf{X}_4$ = 1	$\mathbf{X}_4$ = 2
1	0.1	0.9
2	0.5	0.5

$\mathbf{X}_4$	$\mathbf{X}_5$ = 1	$\mathbf{X}_5$ = 2
1	0.5	0.5
2	0.6	0.4

	$\mathbf{X}_6$ = 1	$\mathbf{X}_6$ = 2
	0.6	0.4

$\mathbf{X}_6$	$\mathbf{X}_7$	$\mathbf{X}_8$	$\mathbf{X}_9$ = 1	$\mathbf{X}_9$ = 2
1	1	1	0.8	0.2
1	1	2	0.1	0.9
1	2	1	0.9	0.1
1	2	2	0.7	0.3
2	1	1	0.3	0.7
2	1	2	0.2	0.8
2	2	1	0.2	0.8
2	2	2	0.9	0.1

$\mathbf{X}_5$	$\mathbf{X}_7$ = 1	$\mathbf{X}_7$ = 2
1	0.2	0.8
2	0.3	0.7

$\mathbf{X}_5$	$\mathbf{X}_8$ = 1	$\mathbf{X}_8$ = 2
1	0.8	0.2
2	0.7	0.3

$\mathbf{X}_9$	$\mathbf{X}_{10}$ = 1	$\mathbf{X}_{10}$ = 2
1	0.8	0.2
2	0.8	0.2

$\mathbf{X}_{10}$	$\mathbf{X}_{11}$ = 1	$\mathbf{X}_{11}$ = 2
1	0.7	0.3
2	0.8	0.2

$$P(X_3, X_4) = \sum_{X_1 X_2 X_5 X_6 X_7 X_8 X_9 X_{10} X_{11}} P(X_1, X_2, \dots, X_{11})$$

$$= \sum_{X_1 X_2 X_5 X_6 X_7 X_8 X_9 X_{10} X_{11}} P(X_1) P(X_2|X_1) P(X_3|X_2) P(X_4|X_3) P(X_5|X_4) P(X_6) P(X_7|X_5) P(X_8|X_5) P(X_9|X_6 X_7 X_8) P(X_{10}|X_9) P(X_{11}|X_{10})$$

$$= \sum_{X_1} P(X_1) \sum_{X_2 X_5 X_6} P(X_2|X_1) P(X_3|X_2) P(X_4|X_3) P(X_5|X_4) \sum_{X_6} P(X_6) \sum_{X_7} P(X_7|X_5) \sum_{X_8} P(X_8|X_5) \sum_{X_9} P(X_9|X_6 X_7 X_8) \sum_{X_{10}} P(X_{10}|X_9) \sum_{X_{11}} P(X_{11}|X_{10})$$

$$= \sum_{X_2 X_5 X_6} P(X_2|X_1) P(X_3|X_2) P(X_4|X_3) P(X_5|X_4)$$

Question 3. If we have the following probability tables for the nodes. Compute the following probabilities. Clearly write down all the necessary steps.

(a) Calculate the following conditional probability:

$$P(\mathbf{X}_3 = 2 | \mathbf{X}_4 = 1)$$

(6 points)

(b) Calculate the following conditional probability:

$$P(\mathbf{X}_5 = 2 | \mathbf{X}_3 = 1, \mathbf{X}_{11} = 2, \mathbf{X}_1 = 1)$$

(9 points)

(Hint: find a short answer. The values in some of the probability tables may reveal some useful information.)

$\mathbf{X}_1$ = 1	$\mathbf{X}_1$ = 2
0.5	0.5

$\mathbf{X}_1$	$\mathbf{X}_2$ = 1	$\mathbf{X}_2$ = 2
1	0.2	0.8
2	0.3	0.7

$\mathbf{X}_2$	$\mathbf{X}_3$ = 1	$\mathbf{X}_3$ = 2
1	0.3	0.7
2	0.3	0.7

$\mathbf{X}_3$	$\mathbf{X}_4$ = 1	$\mathbf{X}_4$ = 2
1	0.1	0.9
2	0.5	0.5

$\mathbf{X}_4$	$\mathbf{X}_5$ = 1	$\mathbf{X}_5$ = 2
1	0.5	0.5
2	0.6	0.4

$\mathbf{X}_6$ = 1	$\mathbf{X}_6$ = 2
0.6	0.4

$\mathbf{X}_6$	$\mathbf{X}_7$	$\mathbf{X}_8$	$\mathbf{X}_9$ = 1	$\mathbf{X}_9$ = 2
1	1	1	0.8	0.2
1	1	2	0.1	0.9
1	2	1	0.9	0.1
1	2	2	0.7	0.3
2	1	1	0.3	0.7
2	1	2	0.2	0.8
2	2	1	0.2	0.8
2	2	2	0.9	0.1

$\mathbf{X}_5$	$\mathbf{X}_7$ = 1	$\mathbf{X}_7$ = 2
1	0.2	0.8
2	0.3	0.7

$\mathbf{X}_5$	$\mathbf{X}_8$ = 1	$\mathbf{X}_8$ = 2
1	0.8	0.2
2	0.7	0.3

$\mathbf{X}_9$	$\mathbf{X}_{10}$ = 1	$\mathbf{X}_{10}$ = 2
1	0.8	0.2
2	0.8	0.2

$\mathbf{X}_{10}$	$\mathbf{X}_{11}$ = 1	$\mathbf{X}_{11}$ = 2
1	0.7	0.3
2	0.8	0.2

(a) $P(X_3=2 | X_4=1) = \dfrac{P(X_3=2, X_4=1)}{P(X_4=1)} = \dfrac{P(X_3=2)\,P(X_4=1 | X_3=2)}{P(X_4=1)}$

$= \dfrac{(0.7).(0.5)}{P(X_4=1|X_3=1)\,P(X_3=1) + P(X_4=1|X_3=2)\,P(X_3=2)}$

X_3 independent of X_2 based on table

$= \dfrac{(0.7)\ (0.5)}{(0.1)(0.3) + (0.5)(0.7)}$

$= 0.92$

(b) $P(X_5=2 | X_3=1, X_{11}=2, X_1=1)$

$= P(X_5=2 | X_3=1)$

X_{10} independent of X_9

$\rightarrow X_{11}$ independent of X_5

X_3 independent of X_2

$\rightarrow X_2$ independent of X_1

$\rightarrow X_5$ independent of X_1

$= \dfrac{P(X_5=2, X_3=1)}{P(X_3=1)}$

$= \dfrac{P(X_3=1)\,P(X_5=2|X_4)\ P(X_4|X_3=1)}{0.3}$

$= \dfrac{P(X_3=1)P(X_4=1|X_3=1)\,P(X_5=2|X_4=1) + P(X_3=1)\,P(X_4=2|X_3=1)\,P(X_5=2|X_4=2)}{0.3}$

$= \dfrac{0.3 \cdot 0.1 \cdot 0.5 + 0.3 \cdot 0.9 \cdot 0.4}{0.3} = \dfrac{0.123}{0.3} = 0.41$

	X_1	
	1	2
	0.5	0.5

X_1	X_2 = 1	2
1	0.2	0.8
2	0.3	0.7

X_2	X_3 = 1	2
1	0.3	0.7
2	0.3	0.7

X_3	X_4 = 1	2
1	0.1	0.9
2	0.5	0.5

X_4	X_5 = 1	2
1	0.5	0.5
2	0.6	0.4

	X_6	
	1	2
	0.6	0.4

X_6	X_7	X_8	X_9 = 1	2
1	1	1	0.8	0.2
1	1	2	0.1	0.9
1	2	1	0.9	0.1
1	2	2	0.7	0.3
2	1	1	0.3	0.7
2	1	2	0.2	0.8
2	2	1	0.2	0.8
2	2	2	0.9	0.1

X_5	X_7 = 1	2
1	0.2	0.8
2	0.3	0.7

X_5	X_8 = 1	2
1	0.8	0.2
2	0.7	0.3

X_9	X_{10} = 1	2
1	0.8	0.2
2	0.8	0.2

X_{10}	X_{11} = 1	2
1	0.7	0.3
2	0.8	0.2

Question 4.

(a) Now, assume we do not have any knowledge about the probability tables for the nodes in the network, but we have the following 12 observations/samples. Find a way to estimate the probability tables associated with the nodes X_3 and X_9 respectively. *(6 points)*

X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}	X_{11}
1	1	2	2	2	1	1	1	1	1	1
1	2	1	1	2	1	1	1	1	1	2
2	2	2	1	2	2	1	1	2	2	1
1	1	2	1	2	1	1	2	1	2	2
1	2	1	1	1	1	2	2	2	1	1
2	2	1	2	1	2	2	1	1	1	2
2	1	2	2	1	2	1	2	2	2	1
2	2	2	1	2	1	2	2	1	2	2
1	1	1	1	2	2	1	1	1	1	1
1	1	1	1	2	1	1	1	2	1	2
1	2	1	2	2	1	2	1	2	1	2
2	2	1	2	1	2	2	2	1	1	1

$$\theta_{x_3}(1) = \frac{\text{Count}(X_2=1, X_3=1)}{\text{Count}(X_2=1)} = \frac{2}{5}$$

$$\theta_{X_3}(2) = \frac{\text{Count}(X_2=1, X_3=2)}{\text{Count}(X_2=1)} = \frac{3}{5}$$

$$\theta_{x_3}(1) = \frac{\text{Count}(X_2=2, X_3=1)}{\text{Count}(X_2=2)} = \frac{5}{7}$$

$$\theta_{x_3}(2) = \frac{\text{Count}(X_2=2, X_3=2)}{\text{Count}(X_2=2)} = \frac{2}{7}$$

X_2	X = 1	2
1	$\frac{2}{5}$	$\frac{3}{5}$
2	$\frac{5}{7}$	$\frac{2}{7}$

	X_1 = 1	X_1 = 2
	0.5	0.5

X_1	X_2 = 1	X_2 = 2
1	0.2	0.8
2	0.3	0.7

X_2	X_3 = 1	X_3 = 2
1	0.3	0.7
2	0.3	0.7

X_3	X_4 = 1	X_4 = 2
1	0.1	0.9
2	0.5	0.5

X_4	X_5 = 1	X_5 = 2
1	0.5	0.5
2	0.6	0.4

	X_6 = 1	X_6 = 2
	0.6	0.4

X_5	X_7 = 1	X_7 = 2
1	0.2	0.8
2	0.3	0.7

X_5	X_8 = 1	X_8 = 2
1	0.8	0.2
2	0.7	0.3

X_6	X_7	X_8	X_9 = 1	X_9 = 2
1	1	1	0.8	0.2
1	1	2	0.1	0.9
1	2	1	0.9	0.1
1	2	2	0.7	0.3
2	1	1	0.3	0.7
2	1	2	0.2	0.8
2	2	1	0.2	0.8
2	2	2	0.9	0.1

X_9	X_{10} = 1	X_{10} = 2
1	0.8	0.2
2	0.8	0.2

X_{10}	X_{11} = 1	X_{11} = 2
1	0.7	0.3
2	0.8	0.2

Question 4.

(a) Now, assume we do not have any knowledge about the probability tables for the nodes in the network, but we have the following 12 observations/samples. Find a way to estimate the probability tables associated with the nodes $\mathbf{X}_3$ and $\mathbf{X}_9$ respectively. *(6 points)*

X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}	X_{11}
1	1	2	2	2	1	1	1	1	1	1
1	2	1	1	2	1	1	1	1	1	2
2	2	2	1	2	2	1	1	2	2	1
1	1	2	1	2	1	1	2	1	2	2
1	2	1	1	1	1	2	2	2	1	1
2	2	1	2	1	2	2	1	1	1	2
2	1	2	2	1	2	1	2	2	2	1
2	2	2	1	2	1	2	2	1	2	2
1	1	1	1	2	2	1	1	1	1	1
1	1	1	1	2	1	1	1	2	1	2
1	2	1	2	2	1	2	1	2	1	2
2	2	1	2	1	2	2	2	1	1	1

111 $\theta_{x9}(1) = \dfrac{\text{count}(X_6=1, X_7=1, X_8=1, X_9=1)}{\text{count}(X_6=1, X_7=1, X_8=1)} = \dfrac{2}{3}$

$\theta_{x9}(2) = \dfrac{\text{count}(X_6=1, X_7=1, X_8=1, X_9=2)}{\text{count}(X_6=1, X_7=1, X_8=1)} = \dfrac{1}{3}$

112 $\theta_{x9}(1) = \dfrac{\text{count}(X_6=1, X_7=1, X_8=2, X_9=1)}{\text{count}(X_6=1, X_7=1, X_8=2)} = \dfrac{1}{1} = 1$

$\theta_{x9}(2) = \dfrac{\text{count}(X_6=1, X_7=1, X_8=2, X_9=2)}{\text{count}(X_6=1, X_7=1, X_8=2)} = \dfrac{0}{1} = 0$

121 $\theta_{x9}(1) = \dfrac{\text{count}(X_6=1, X_7=2, X_8=1, X_9=1)}{\text{count}(X_6=1, X_7=2, X_8=1)} = \dfrac{0}{1} = 0$

$\theta_{x9}(2) = \dfrac{\text{count}(X_6=1, X_7=2, X_8=1, X_9=2)}{\text{count}(X_6=1, X_7=2, X_8=1)} = \dfrac{1}{1} = 1$

122 $\theta_{x9}(1) = \dfrac{\text{count}(X_6=1, X_7=2, X_8=2, X_9=1)}{\text{count}(X_6=1, X_7=2, X_8=2)} = \dfrac{1}{2}$

$\theta_{x9}(2) = \dfrac{\text{count}(X_6=1, X_7=2, X_8=2, X_9=2)}{\text{count}(X_6=1, X_7=2, X_8=2)} = \dfrac{1}{2}$

$\theta_{x9}(1) = \dfrac{\text{count}(X_6=2, X_7=1, X_8=1, X_9=1)}{\text{count}(X_6=2, X_7=1, X_8=1)} = \dfrac{1}{2}$

$\theta_{x9}(2) = \dfrac{\text{count}(X_6=2, X_7=1, X_8=1, X_9=2)}{\text{count}(X_6=2, X_7=1, X_8=1)} = \dfrac{1}{2}$

$\theta_{x9}(1) = \dfrac{\text{count}(X_6=2, X_7=1, X_8=2, X_9=1)}{\text{count}(X_6=2, X_7=1, X_8=2)} = \dfrac{0}{1} = 0$

$\theta_{x9}(2) = \dfrac{\text{count}(X_6=2, X_7=1, X_8=2, X_9=2)}{\text{count}(X_6=2, X_7=1, X_8=2)} = \dfrac{1}{1} = 1$

$\theta_{x9}(1) = \dfrac{\text{count}(X_6=2, X_7=2, X_8=1, X_9=1)}{\text{count}(X_6=2, X_7=2, X_8=1)} = \dfrac{0}{1} = 0$

$\theta_{x9}(2) = \dfrac{\text{count}(X_6=2, X_7=2, X_8=1, X_9=2)}{\text{count}(X_6=2, X_7=2, X_8=1)} = \dfrac{1}{1} = 1$

$\theta_{x9}(1) = \dfrac{\text{count}(X_6=2, X_7=2, X_8=2, X_9=1)}{\text{count}(X_6=2, X_7=2, X_8=2)} = \dfrac{1}{1} = 1$

$\theta_{x9}(2) = \dfrac{\text{count}(X_6=2, X_7=2, X_8=2, X_9=2)}{\text{count}(X_6=2, X_7=2, X_8=2)} = \dfrac{0}{1} = 0$

X_6	X_7	X_8	X_9 = 1	X_9 = 2
1	1	1	2/3	1/3
1	1	2	1	0
1	2	1	0	1
1	2	2	1/2	1/2
2	1	1	1/2	1/2
2	1	2	0	1
2	2	1	0	1
2	2	2	1	0

	X_1 = 1	2
	0.5	0.5

X_1	X_2 = 1	2
1	0.2	0.8
2	0.3	0.7

X_2	X_3 = 1	2
1	0.3	0.7
2	0.3	0.7

X_3	X_4 = 1	2
1	0.1	0.9
2	0.5	0.5

X_4	X_5 = 1	2
1	0.5	0.5
2	0.6	0.4

	X_6 = 1	2
	0.6	0.4

X_6	X_7	X_8	X_9 = 1	2
1	1	1	0.8	0.2
1	1	2	0.1	0.9
1	2	1	0.9	0.1
1	2	2	0.7	0.3
2	1	1	0.3	0.7
2	1	2	0.2	0.8
2	2	1	0.2	0.8
2	2	2	0.9	0.1

X_5	X_7 = 1	2
1	0.2	0.8
2	0.3	0.7

X_5	X_8 = 1	2
1	0.8	0.2
2	0.7	0.3

X_9	X_{10} = 1	2
1	0.8	0.2
2	0.8	0.2

X_{10}	X_{11} = 1	2
1	0.7	0.3
2	0.8	0.2

Question 4.

(a) Now, assume we do not have any knowledge about the probability tables for the nodes in the network, but we have the following 12 observations/samples. Find a way to estimate the probability tables associated with the nodes $\mathbf{X}_3$ and $\mathbf{X}_9$ respectively. *(6 points)*

X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}	X_{11}
1	1	2	2	2	1	1	1	1	1	1
1	2	1	1	2	1	1	1	1	1	2
2	2	2	1	2	2	1	1	2	2	1
1	1	2	1	2	1	1	2	1	2	2
1	2	1	1	1	1	2	2	2	1	1
2	2	1	2	1	2	2	1	1	1	2
2	1	2	2	1	2	1	2	2	2	1
2	2	2	1	2	1	2	2	1	2	2
1	1	1	1	2	2	1	1	1	1	1
1	1	1	1	2	1	1	1	2	1	2
1	2	1	2	2	1	2	1	2	1	2
2	2	1	2	1	2	2	2	1	1	1

(b) Based on the above observations, you would like to find a good Bayesian network structure to model the data. You started with the initial structure shown on the previous page, and decided to delete the edge between $\mathbf{X}_{10}$ and $\mathbf{X}_{11}$. Is the resulting new structure (after deleting the single edge between $\mathbf{X}_{10}$ and $\mathbf{X}_{11}$ from the original graph) better than the original structure in terms of BIC score? Clearly explain the reason. *(9 points)*

(Hint: Try to find a short answer.)

X_{10}	1	2
1	4/8	4/8
2	2/4	2/4

X_{11} independent of X_{10}

Hence removing X_{10} does not affect the Bayesian network log-likelihood

However, the new Bayesian network needs less free variables which leads to a higher BIC score.