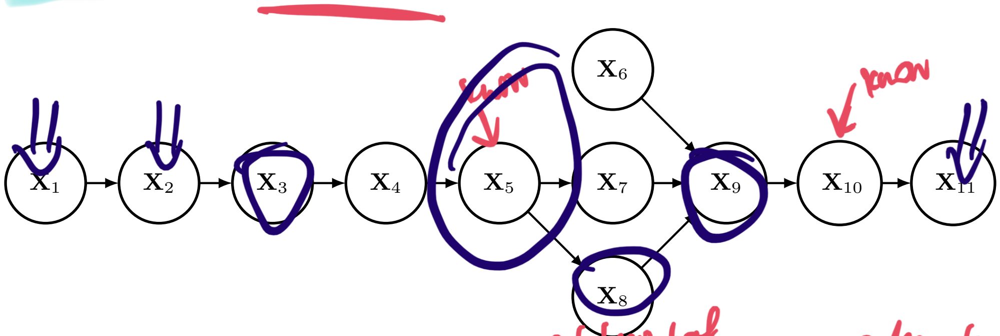
SINGAPORE UNIVERSITY OF TECHNOLOGY AND DESIGN

50.007 Machine Learning, Fall 2021
Homework 5

Due Thursday 9 December 2021, 5pm

This homework will be graded by Zhang Qi

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: $\{1, 2\}$.



x_2:

	1	2
1		
2		

We have to ensure that at least of the inbetween nodes take on the diff values

→ To ensure that X_1 & X_6 are independent

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 $\mathbf{X}_5$ and $\mathbf{X}_{10}$? *(10 points)*

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 $\mathbf{X}_3$, $\mathbf{X}_8$ and $\mathbf{X}_9$ can take 5 different values: $\{1, 2, 3, 4, 5\}$, and all other nodes can only take 4 different values: $\{1, 2, 3, 4\}$? *(10 points)*

We know 11 nodes 11 θs.

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:

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

(8 points)

(b) Calculate the following conditional probability:

$$P(\mathbf{X}_5 = 2 | \mathbf{X}_2 = 1, \mathbf{X}_{11} = 2, \mathbf{X}_1 = 1)$$

(12 points)

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

	X_1	
	1	2
	0.5	0.5

	X_2	
X_1	1	2
1	0.2	0.8
2	0.3	0.7

	X_3	
X_2	1	2
1	0.3	0.7
2	0.3	0.7

	X_4	
X_3	1	2
1	0.1	0.9
2	0.5	0.5

	X_5	
X_4	1	2
1	0.5	0.5
2	0.6	0.4

	X_6	
	1	2
	0.6	0.4

			X_9	
X_6	X_7	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

	X_7	
X_5	1	2
1	0.2	0.8
2	0.3	0.7

	X_8	
X_5	1	2
1	0.8	0.2
2	0.7	0.3

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

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

Question 4. 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_7 and X_9 respectively. *(10 points)*

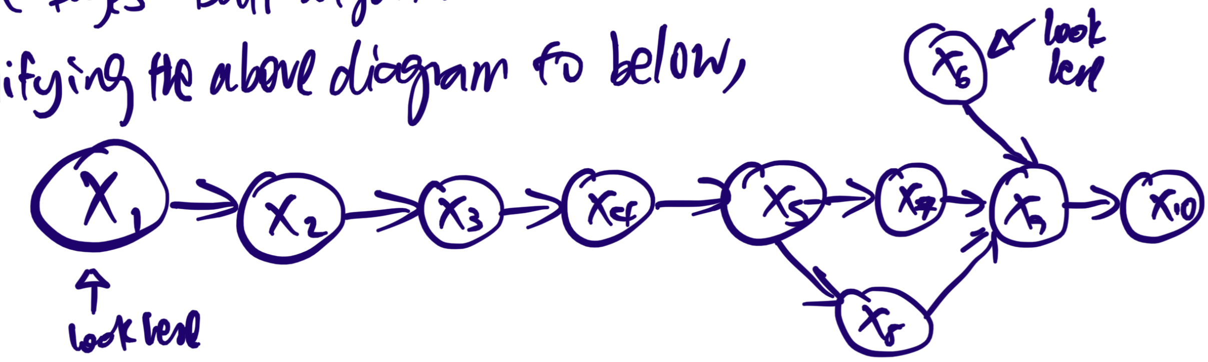
X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}	X_{11}
1	1	2	2	2	1	1	1	2	1	1
1	2	1	1	2	1	1	1	1	1	2
2	2	2	1	2	2	1	1	1	2	1
1	1	2	1	2	1	1	2	1	2	2
1	2	1	1	1	1	2	2	2	1	1
2	2	1	2	1	2	2	1	1	1	2
2	1	2	2	1	2	1	2	2	2	1
2	2	2	1	2	1	2	2	1	2	2
1	1	1	1	2	2	1	1	1	1	1
1	1	1	1	2	1	1	1	2	1	2
1	2	1	2	2	1	2	1	1	1	2
2	2	1	2	1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

Question 1

Aim: X_1 & X_6 independent of each other?

We use the Bayes' Ball algorithm seen below:

Simplifying the above diagram to below,



From this diagram, there is a closed path. Hence, we can say that the ball cannot find the path from X_1 to X_6 or from X_6 to X_1 as none of the X variables are active.

This is because with none of the observed X variables,

Path 1 from X_1 to X_6:

Closed path

Path 2 from X_1 to X_6:

Closed path

Given that X_9 is not observed.

$\therefore$ The ball cannot flow to X_6.

$\therefore$ X_1 and X_6 are independent.

Part 2: With X_5 and X_{10} observed,

With X_5 known, the path is active. Hence, X_1 only need to depend on X_5

With X_5 and X_{10} known, two nodes are active along the path. Hence, X_1 and X_6 are independent

Question 2

Part 1

For the same network were each node takes the value of $\{1, 2\}$, for the number of free parameters, we exclude that of the last column in each node.

$\therefore$

$$\text{N-of free parameters} = \underbrace{(2-1)(1)}_{X_1} + \underbrace{8(2-1)(2)}_{\substack{X_2, X_3, X_4, X_5,\\ X_7, X_8, X_{10}, X_{11}}} + \underbrace{(2-1)(1)}_{X_6} + \underbrace{(2-1)}_{X_9}(2^3)$$

8 nodes

X_9 depends on X_6, X_7, X_8

X_6	X_7	X_8	1	2
1	1	1		
1	1	2		
1	2	1		
1	2	2		

This is just for my sanity check

$$= 1 + 16 + 1 + 8$$

$$= 26 \ \#$$

Same thing as Part 1, given new observation set of

① X_3, X_8 and X_9 taking values $\{1, 2, 3, 4, 5\}$

② The rest taking $\{1, 2, 3, 4\}$

Continue on next page!

For X_3, X_8 and X_9,

N.o of free parameters for X_3 $= (5-1)(4) = 16$

N.o of free parameters for X_8 $= (5-1)(4) = 16$

N.o of free parameters for X_9 $= (5-1)\underbrace{(4^2)}_{\text{depends on } 6,7}\underbrace{(5)}_{\text{depends on } 8} = 320$

For the others,

A — N.o of free parameters $= (4-1) = 3$ ← for X_1 and X_6 each

B — N.o of free parameters $= (4-1)(4) = 12$ ← for X_2, X_5, X_7, X_{11} each

C — N.o of free parameters $= (4-1)(5) = 15$ ← X_4, X_{10} each

Total n. of effective Parameters $= 2\,C + 4\,B + 2\,A + 16 + 16 + 320$

$= 30 + 48 + 6 + 16 + 16 + 320$

$= 436$ #

Question 3

(a) $$P(X_3=1 \mid X_4=2) = \frac{P(X_3=1, X_4=2)}{\sum_{b\in\{1,2\}} P(X_3=b, X_4=2)}$$

$$P(X_3=b, X_4=2)$$

$$= \sum_{a\in\{1,2\}} \sum_{d\in\{1,2\}} P(X_1=a, X_2=d, X_3=b, X_4=2)$$

$$= \sum_{a\in\{1,2\}} \sum_{d\in\{1,2\}} P(X_1=a)\, P(X_2=d \mid X_1=a)\, P(X_3=b \mid X_2=d,$$
$$P(X_2=d)\, P(X_4=2 \mid X_3=b)\, P(X_3=b)$$

$$= P(X_3=b)\, P(X_4=2 \mid X_3=b)$$ ← we can trace and know that X_4 is only dependent on X_3 being observed.

$$\therefore \sum_{b\in\{1,2\}} P(X_3=b, X_4=2)$$
$$= [P(X_3=1)\, P(X_4=2 \mid X_3=1)] + [P(X_3=2) \cdot P(X_4=2 \mid X_3=2)]$$

next Page ....

$$P(X_3 = b) = \sum_{z \in \{1,2\}} \sum_{a \in \{1,2\}} P(X_1 = z)\, P(X_2 = a) \cdot P(X_3 = b \mid X_2 = a)$$

$$P(X_2 \mid X_1 = z)$$

we can use the idea of the open & close paths to determine independence!

$$= \sum_{a \in \{1,2\}} P(X_3 = b \mid X_2 = a) \cdot P(X_2 = a)$$

$$\therefore P(X_3 = 1) = 0.3\,[(0.2 \times 0.5) + (0.3 \times 0.5)] + 0.3\,[(0.8 \times 0.5) + (0.7 \times 0.5)]$$

$$= 0.3$$ ← we can see that X_3 does need to involve X_1 and X_2 in this case.

Using the same discovery,

$$P(X_3 = 2) = 0.7$$

$$\therefore \sum_{b = \{1,2\}} P(X_3 = b, X_4 = 2) = 0.3 \times (0.9) + 0.7 \times (0.5) = \frac{31}{50} = 0.62$$

$$P(X_3 = 1, X_4 = 2) = 0.3\,[0.9] = 0.27$$

$$\therefore P(X_3 = 1 \mid X_4 = 2) = \frac{0.27}{0.62} = \frac{27}{62} \#$$

(b) $P(X_5=2 \mid X_2=1, X_{11}=2, X_1=1)$

Same thing using Bayes theorem,

$$= \frac{P(X_5=2, X_2=1, X_{11}=2, X_1=1)}{P(X_2=1, X_{11}=2, X_1=1)}$$

We identify the close and open paths and split accordingly:

X_5 is given in terms of observation

We also identify that it's a chain. Hence, we can apply the chain ppty stated in the lecture notes!

Since X_5 is an ob ed event,

$$P(X_6,\ldots,X_{11} \mid X_1, X_2, X_3, X_4, X_5)$$

$$= P(X_6,\ldots,X_{11} \mid X_5)$$

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

$$= \sum_{\substack{X_3, X_4, \\ X_6, X_7 \\ X_8, X_9, X_{10}}} P(X_1,\ldots,X_5) \left(P(X_6,\ldots,X_{11} \mid X_5) \right)$$

next page....

$$P(X_1, X_2, X_5, X_{11})$$

$$= \sum_{X_3, X_4} P(X_1, \ldots, X_5) \cdot \sum_{\substack{X_6, X_7, \\ X_8, X_9, \\ X_{10}}} P(X_6, \ldots, X_{11} \mid X_5)$$

let's eyeball and use the standing similarity in the table of values of $X_{10} \Rightarrow X_1$ to X_9 are independent of X_{10} we employ the same trick as part a.

$$= \left(\sum_{X_3, X_4} P(X_5 \mid X_4, \ldots, X_1) \cdot P(X_4 \mid X_3, X_2, X_1) \cdot P(X_2 \mid X_1) \cdot P(X_1) \right) \cdot \left(\cdot \sum_{X_{10}} P(X_{11}, X_{10}) \right)$$

$$= \left(\sum_{X_3, X_4} P(X_5 \mid X_4) \cdot P(X_4 \mid X_3) \cdot P(X_2 \mid X_1) \cdot P(X_1) \cdot \right)$$

← for each node has their set of values.

$$\sum_{X_{10}} P(X_{11} / X_{10}) \, P(X_{10})$$

$$= \sum_{X_3, X_4} P(X_5 \mid X_4) \cdot P(X_4 \mid X_3) \cdot P(X_3) \cdot P(X_2 \mid X_1) \cdot P(X_1)$$

$$\cdot \sum_{X_{10}} P(X_{11} \mid X_{10}) \cdot P(X_{10})$$

$$P(X_1 = 1, X_2 = 1, X_5 = 2, X_{11} = 2)$$

Continue next page!

Note: The method is adapted from Ang Song Gee.

$$P(X_1=1, X_2=1, X_5=2, X_{11}=2)$$

$$= \begin{bmatrix} P(X_5=2 \mid X_4=1) \cdot P(X_4=1 \mid X_3=1) \cdot P(X_3=1) + \\ P(X_5=2 \mid X_4=1) \cdot P(X_4=1 \mid X_3=2) \cdot P(X_3=2) + \\ P(X_5=2 \mid X_4=2) \cdot P(X_4=2 \mid X_3=1) \cdot P(X_3=1) + \\ P(X_5=2 \mid X_4=2) \cdot P(X_4=2 \mid X_3=2) \cdot P(X_3=2) \end{bmatrix} \cdot$$

$$P(X_2=1 \mid X_1=1) \cdot P(X_1=1) \cdot$$

$$\begin{bmatrix} P(X_{11}=2 \mid X_{10}=1) \cdot P(X_{10}=1) + \\ P(X_{11}=2 \mid X_{10}=2) \cdot P(X_{10}=2) \end{bmatrix}$$

$$= \begin{bmatrix} 0.5 \cdot 0.1 \cdot 0.3 + \\ 0.5 \cdot 0.5 \cdot 0.7 + \\ 0.4 \cdot 0.9 \cdot 0.3 + \\ 0.4 \cdot 0.5 \cdot 0.7 \end{bmatrix} \cdot 0.2 \cdot 0.5 \cdot \begin{bmatrix} 0.3 \cdot 0.8 + \\ 0.2 \cdot 0.2 \end{bmatrix}$$

$$= 0.012264$$

$$P(X_2, X_{11}, X_1) = P(X_1) \cdot P(X_2 \mid X_1) \cdot \sum_{X_{10}} P(X_{11} \mid X_{10}) \cdot P(X_{10})$$

Same thing,

$$P(X_2=1, X_{11}=2, X_1=1) = P(X_1=1) \cdot P(X_2=1 \mid X_1=1) \cdot \sum_{X_{10}} P(X_{11}=2 \mid X_{10}) \cdot P(X_{10})$$

$$= 0.5 \cdot 0.2 \cdot [0.3 \cdot 0.8 + 0.2 \cdot 0.2]$$

$$= 0.028$$

Next page

$$\therefore P(X_5=2 \mid X_1=1, X_2=1, X_{11}=2)$$

$$= \frac{0.012264}{0.028} = 0.438$$

Question 4

For node X_7, we notice the following:

Direct connection is node X_5

To form the probability table, we calculate the maximum likelihood parameter using the following formula to calculate each probability table of nodes X_7 and X_9 individually.

$$\hat{\theta}_i(x_i \mid x_{pa_i}) = \frac{\text{Count}((x_i, X_{pa_i})) \text{ in } D}{\text{Count}((X_{pa_i}) \text{ in } D)}, \quad x_i \in \{1, \dots, r_i\}$$

$$\hat{\theta}_7(x_7 \mid x_5) = \frac{\text{Count}((x_7, X_5)) \text{ in } D}{\text{Count}((X_5) \text{ in } D)}$$

$X_5 \Rightarrow X_7$ } what we are looking at.

$$\hat{\theta}_7(x_7=1 \mid X_5=1) = \frac{1}{4}$$

$$\hat{\theta}_7(x_7=2 \mid X_5=2) = \frac{2}{8}$$

$$= \frac{1}{4}$$

$$\hat{\theta}_7(x_7=1 \mid X_5=2) = \frac{6}{8} = \frac{3}{4}$$

$$\hat{\theta}_7(x_7=2 \mid x_5=1) = \frac{3}{4}$$

$\therefore$ For X_7

	X_7	
X_5	1	2
1	0.25	0.75
2	0.75	0.25

For X_9, we need to take into account X_6, X_7 and X_8

$$\hat{\theta}_9(X_9=1 \mid X_6=1, X_7=1, X_8=1)$$

$$= \frac{1}{3}$$

$$\hat{\theta}_9(X_9=1 \mid X_6=1, X_7=1, X_8=2)$$

$$= \frac{1}{1} = 1$$

$$\hat{\theta}_9(X_9=1 \mid X_6=1, X_7=2, X_8=1)$$

$$= \frac{1}{1} = 1$$

$$\hat{\theta}_9(X_9=1 \mid X_6=1, X_7=2, X_8=2)$$

$$= \frac{1}{2}$$

$$\hat{\theta}_9(X_9=1 \mid X_6=2, X_7=1, X_8=1)$$

$$= \frac{1+1}{2} = 1$$

$$\hat{\theta}_9(X_9=1 \mid X_6=2, X_7=1, X_8=2)$$

$$= \frac{0}{1} = 0$$

$$\hat{\theta}_9(X_9=1 \mid X_6=2, X_7=2, X_8=1)$$

$$= \frac{1+1}{1+1} = 1$$

$$\hat{\theta}_9(X_9=1 \mid X_6=2, X_7=2, X_8=2)$$

$$= \frac{0}{1} = 0$$

$$\hat{\theta}_9(X_9=2 \mid X_6=1, X_7=1, X_8=1)$$

$$= \frac{1+1}{1+1+1} = \frac{2}{3}$$

$$\hat{\theta}_9(X_9=2 \mid X_6=1, X_7=1, X_8=2)$$

$$= \frac{0}{1} = 0$$

$$\hat{\theta}_9(X_9=2 \mid X_6=1, X_7=2, X_8=1)$$

$$= \frac{0}{1} = 0$$

$$\hat{\theta}_9(X_9=2 \mid X_6=1, X_7=2, X_8=2)$$
$$= \frac{1}{1+1} = \frac{1}{2}$$

$$\hat{\theta}_9(X_9=2 \mid X_6=2, X_7=1, X_8=1)$$
$$= \frac{0}{1+1} = 0$$

$$\hat{\theta}_9(X_9=2 \mid X_6=2, X_7=1, X_8=2)$$
$$= \frac{1}{1} = 1$$

$$\hat{\theta}_9(X_9=2 \mid X_6=2, X_7=2, X_8=1)$$
$$= \frac{0}{1} = 0$$

$$\hat{\theta}_9(X_9=2 \mid X_6=2, X_7=2, X_8=2)$$
$$= \frac{1}{1} = 1$$

			X_9	
X_6	X_7	X_8	1	2
1	1	1	$1/3$	$\frac{2}{3}$
1	1	2	1	0
1	2	1	1	0
1	2	2	$\frac{1}{2}$	$\frac{1}{2}$
2	1	1	1	0
2	1	2	0	1
2	2	1	1	0
2	2	2	0	1

entries
$2^3 \times 2$
$= 16$