CS 726: Advanced Machine Learning, Spring 2025, Homework 1

Roll: 22B1003

Name: SAKSHAM RATHI

Mode: Credit/Audit/Sit-through Credit

Write all your answers in the space provided. Do not spend time/space giving irrelevant details or details not asked for. Use the marks as a guideline for the amount of time you should spend on a question. You are only allowed to refer your hand-written notes, no one else's notes or textbook.

---

1. Prove that if we factorize $\Pr(\mathbf{x})$ of a DAG as follows

$$\Pr(x_1 \ldots x_n) = \prod_{i=1}^{n} f_i(x_i, x_{\pi_i})$$

where π_i denotes the indices corresponding to the parents of node i and $\sum_{x_i} f_i(x_i, x_{\pi_i}) = 1$ then $f_i(x_i, x_{\pi_i}) = \Pr(x_i \mid x_{\pi_i})$

Let us assume, without of generality that $x_1 \ldots x_n$ are topologically ordered (parents before children) in DAG G.

Then $x_1, x_2, \ldots, x_{i-1} = Pa_G(x_i) \cup ND'(x_i)$

Basically, in this set $\{x_1 \ldots x_{i-1}\}$, some will be parents of x_i and others will be non-descendants.

Also, $\Pr(x_1 \ldots x_n) = \prod_{i=1}^{n} P(x_i \mid x_1 \ldots x_{i-1})$ (from the definition of conditional probability)

$= \prod_{i=1}^{n} P(x_i \mid Pa_G(x_i) \cup ND'(x_i))$ (from above)

$= \prod_{i=1}^{n} P(x_i \mid Pa_G(x_i))$ (In a DAG, this is true by definition as $x_i \perp\!\!\!\perp ND(x_i) \mid Pa(x_i)$)

$= \prod_{i=1}^{n} f(x_i, x_{\pi_i})$ and hence $f(x_i, x_{\pi_i}) = \Pr(x_i \mid x_{\pi_i})$ ..3

* Another solution added next page

2. Let $P(x_1, \ldots, x_4)$ be a distribution defined over binary variables as follows

$$P(x_1, \ldots, x_4) = \frac{1}{Z} e^{x_1 \oplus x_2 \oplus x_3} e^{x_3 \oplus x_4} \tag{1}$$

where $\oplus$ denotes the XOR operation. XOR of two binary variables is 0 when both its arguments are the same and 1 otherwise. The value of the numerator for some of the entries have been filled in. You need to fill in the five missing entries.

$$Pr(x_1 \ldots x_n) = \prod_{i=1}^{n} f_i(x_i, x_{\pi_i})$$

To prove: $f_i(x_i, x_{\pi_i}) = Pr(x_i | x_{\pi_i})$

W.L.O.G, let's assume that $x_1 \ldots x_n$ are in topological order.

$$f_i(x_i, x_{\pi_i}) = u_i$$
$$Pr(x_i | x_{\pi_i}) = v_i$$

Let us use strong induction

$$u_1 = f_1(x_1, x_{\pi_1}) = f(x_1, \phi)$$
$$Pr(x_1 | \phi) = v_1 = Pr(x_1)$$

Hence the base case is satisfied and $u_1 = v_1$

The equation we have

$$Pr(x_1 \ldots x_n) = \prod_{i=1}^{n} f_i(x_i, x_{\pi_i})$$
$$Pr(x_1) = \sum_{x_2} \sum_{x_3} \ldots \sum_{x_n} Pr(x_1 \ldots x_n)$$
$$= \sum_{x_2} \sum_{x_3} \ldots \sum_{x_n} \prod_{i=1}^{n} f_i(x_i, x_{\pi_i})$$
$$= f_1(x_1) \underbrace{\sum_{x_2} f_2(x_2, x_{\pi_2})}_{1} \underbrace{\sum_{x_3} f(x_3, x_{\pi_3})}_{1} \ldots$$

(from question)

$$= f_1(x_1, \phi)$$

Assume, it is true for $i \le k$ and we need to prove for $k+1$ to n

$$u_i = v_i \quad \forall \; i \le k$$
$$Pr(x_1 \ldots x_n) = \prod_{i=1}^{n} Pr(x_i | x_{\pi_i})$$

from a BN's definition

$$u_i = Pr(x_i | x_{\pi_i})$$
$$v_i = f(x_i, x_{\pi_i})$$

we already have $\prod_{i=1}^{k} u_i = \prod_{i=1}^{k} v_i$ (because $u_i = v_i$)

$$LHS = \prod_{i=1}^{n} u_i \qquad RHS = \prod_{i=1}^{n} v_i$$

$$LHS = \left(\prod_{i=1}^{k} u_i\right) u_{k+1} \left(\prod_{i=k+2}^{n} u_i\right)$$

$$\sum_{x_{k+2}} \sum \ldots \sum_{n} \left(\prod_{i=k+2}^{n} u_i\right) = 1$$

because $\sum_{x_i} Pr(x_i | x_{\pi_i}) = 1 \; \forall i$

Similarly, $RHS = \left(\prod_{i=1}^{k} v_i\right)(v_{k+1})\left(\prod_{i=k+2}^{n} v_i\right)$

$$\sum_{x_{k+2}} \ldots \sum_{n} \left(\prod_{i=1}^{k+2} v_i\right) = 1$$

because $\sum_{x_i} f(x_i, x_{\pi_i}) = 1 \; \forall i$ (given)

Since LHS = RHS

$$\Rightarrow \left(\prod_{i=1}^{k} u_i\right) u_{k+1} = \left(\prod_{i=1}^{k} v_i\right) v_{k+1}$$
$$\Rightarrow u_{k+1} = v_{k+1}$$

Hence our induction is complete and we have proved that $u_i = Pr(x_i | x_{\pi_i}) = f(x_i, x_{\pi_i}) = v_i \quad \forall i$

$P(x_1=0,x_2=0) = P(x_1=1,x_2=0) = P(x_1=0,x_2=1) = P(x_1=1,x_2=1) = \frac{1}{4}$

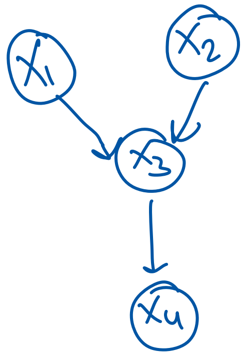
$P(x_1=0) = P(x_1=1) = P(x_2=0) = P(x_2=1) = \frac{1}{2}$ $\Big\} \rightarrow x_1 \perp\!\!\!\perp x_2$

$P(x_3=1) = P(x_3=0) = \frac{1}{2}$ $\quad P(x_4=0) = P(x_4=1) = \frac{1}{2}$

$P(x_3=0 \mid x_1=0, x_2=0) \neq P(x_3=0 \mid x_1=0) \Rightarrow x_3 \not\perp\!\!\!\perp x_2 \mid x_1$

$P(x_3=0 \mid x_1=0, x_2=0) \neq P(x_3=0 \mid x_2=0) \Rightarrow x_3 \not\perp\!\!\!\perp x_1 \mid x_2$

x_1	x_2	x_3	x_4	$ZP(\mathbf{x})$
0	0	0	0	1
0	0	0	1	e
0	0	1	0	e^2
0	0	1	1	e
0	1	0	0	e
0	1	0	1	e^2
0	1	1	0	e
0	1	1	1	1
1	0	0	0	e
1	0	0	1	e^2
1	0	1	0	e
1	0	1	1	1
1	1	0	0	1
1	1	0	1	e
1	1	1	0	e^2
1	1	1	1	e



$(x_4 \perp\!\!\!\perp x_1, x_2 \mid x_3)$

From the nature of the distribution ..1

Eg: $P(x_4=0 \mid x_3=0) = \frac{2+2e}{2+4e+2e^2} = \frac{1}{1+e}$

$P(x_4=0 \mid x_1=x_2=x_3=0) = \frac{1}{1+e}$

Similarly we can show for other variable assignments

(a) Calculate the value of Z

The sum $\frac{1}{Z}\sum_{x_1 \ldots x_4} P(x_1 \ldots x_4)$ should be 1 (sum of probabilities across the distribution should be 1)

Therefore $Z = 1 \times 4 + e \times 8 + e^2 \times 4 = 4(e^2+2e+1) = (2(e+1))^2$

..1

(b) Draw a minimal Bayesian network representing the above distribution using the variable order x_1, x_2, x_3, x_4 to the right of the above table. ..2

(c) Write the CPD for $\Pr(x_1|\text{Pa}(x_1)), \Pr(x_2|\text{Pa}(x_2)), \Pr(x_3|\text{Pa}(x_3))$ in your Bayesian network above. $Pa(x_1) = \phi \quad Pa(x_2) = \phi \quad Pa(x_3) = \{x_1, x_2\}$

$Pr(x_1 \mid Pa(x_1)) =$

x_1	0	1
Probabilities	$\frac{1}{2}$	$\frac{1}{2}$

$Pr(x_2 \mid Pa(x_2)) =$

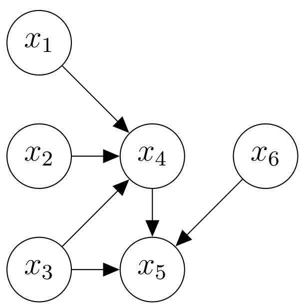
x_2	0	1
P	$\frac{1}{2}$	$\frac{1}{2}$

$Pr(x_3 \mid \{x_1, x_2\}) =$

$x_3 \downarrow$ $x_1x_2 \rightarrow$	00	01	10	11
0	$\frac{1}{1+e}$	$\frac{e}{e+1}$	$\frac{e}{e+1}$	$\frac{1}{e+1}$
1	$\frac{e}{e+1}$	$\frac{1}{e+1}$	$\frac{1}{e+1}$	$\frac{e}{e+1}$

..3

3. Assume a distribution $P(x_1, x_2, \ldots, x_6)$ is represented by the Bayesian network H below:



(a) In the BN H, list all variables that are unconditionally independent of x_6

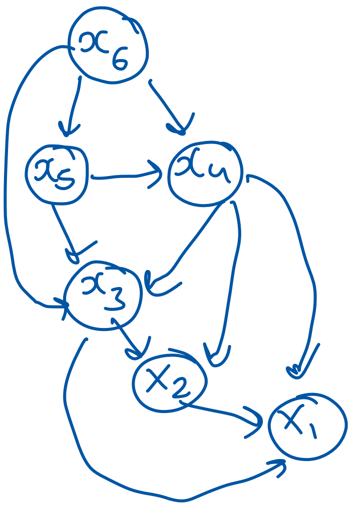
In a BN, we know that $X \perp\!\!\!\perp ND(X) \mid Pa_G(X)$ i.e. X is independent of its non-descendants given its parents. x_6 has no parents, so

$x_6 \perp\!\!\!\perp x_1, x_2, x_3, x_4 \mid \phi$ $\Rightarrow$ variables that are unconditionally independent of $x_6 = x_1, x_2, x_3, x_4$

(x_5 = Descendant)

..1

(b) Using D separation rule on the above network to get answers to yes/no questions on conditional independencies, draw a fresh *minimal and correct* Bayesian network G using the variable order $x_6, x_5, x_4, x_3, x_2, x_1$.



$x_6 \not\perp\!\!\!\perp x_5$ (there is a direct path)

$x_4 \not\perp\!\!\!\perp x_6 \mid x_5$ (There is $x_4 \rightarrow x_5 \leftarrow x_6$ in BN and $x_5 \in Z$)

$x_3 \perp\!\!\!\perp x_6 \mid x_4, x_5$ (Because of the same reason as above)

$x_2 \perp\!\!\!\perp x_5, x_6 \mid x_3, x_4$ → single path: $x_2 \rightarrow x_4 \rightarrow x_5 \leftarrow x_6$ and $x_5 \notin Z$

single path from x_2 to x_5 and one triple and $x_4 \in Z$

$x_1 \perp\!\!\!\perp x_5, x_6 \mid x_2, x_3, x_4$ → spouse

Reasoning similar as above

..3

4. You are given the following statements about the conditional independence about a set of 6 variables A, B, C, D, E, F

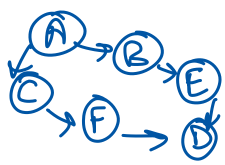
(a) C is independent of B given A

(b) F is independent of $\{A, B\}$ given C

(c) E is independent of $\{A, C, F\}$ given B

(d) D is independent of $\{B, A, C\}$ given E, F

Let us choose the variable ordering for BN: A, B, C, F, E, D



Given these statements state which of the following are true. Give a brief justification for your answer. No marks without the right justification. [You may find it useful to first draw a suitable graphical model that is possible to draw given the above statements.]

(1) Is F is independent of B given A? (2) Is C is independent of E?

Now, we can use d-separation algorithm

(1) $F \perp\!\!\!\perp B \mid A$? F and B two paths (a) $F \rightarrow D \leftarrow E \leftarrow B$ blocked $D \notin Z$, $Z = \{A\}$

(b) $F \leftarrow C \leftarrow A \rightarrow B$ unblocked blocked

$\Rightarrow$ Yes F is independent of B given A.

(2) Is $C \perp\!\!\!\perp E \mid \phi$

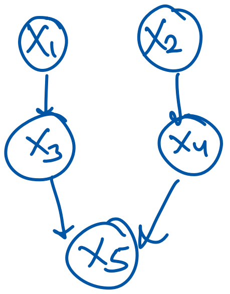
Two paths between C and E: (a) $C \leftarrow A \rightarrow B \rightarrow E$ unblocked path (none of the two triples is blocking this path)

$\Rightarrow$ Hence, $C \not\perp\!\!\!\perp E$.
(no need to check the other path).

(Since, this does not hold in the BN we constructed, we can't say that this will hold for sure)

..4

5. Draw a Bayesian network over five variables $x_1, \ldots, x_5$ assuming the variable order x_1, x_2, x_3, x_4, x_5. For this ordering, assume that the following set of local CIs hold in the distribution: $x_1 \perp\!\!\!\perp x_2$, $x_3 \perp\!\!\!\perp x_2 | x_1$, $x_4 \perp\!\!\!\perp x_1, x_3 | x_2$, $x_5 \perp\!\!\!\perp x_1, x_2 | x_3, x_4$



For every x_i we need to find the smallest subset S of $Q = \{x_1 \ldots x_{i-1}\}$ such that $x_i \perp\!\!\!\perp Q - S \mid S$ and then draw edges from S to x_i.

(i) $X_1 \perp\!\!\!\perp X_2$ (hence no edge)
(ii) $X_3 \perp\!\!\!\perp X_2 | X_1$ (hence $X_1 \rightarrow X_3$ edge)
(iii) $X_4 \perp\!\!\!\perp X_1, X_3 | X_2$ (hence $X_2 \rightarrow X_4$ edge)
(iv) $X_5 \perp\!\!\!\perp X_1, X_2 | X_3, X_4$ hence $X_3 \rightarrow X_5$ and $X_4 \rightarrow X_5$ edges

..2

6. In the above Bayesian network, use only local CIs and the standard conditional probability axioms (2.7 to 2.10 from Chapter 2 of your textbook) to prove that $x_3 \perp\!\!\!\perp x_4$.

Symmetry: $X \perp\!\!\!\perp Y | Z \Rightarrow Y \perp\!\!\!\perp X | Z$ (2.7)
Decomposition: $X \perp\!\!\!\perp Y, W | Z \Rightarrow X \perp\!\!\!\perp Y | Z$ (2.8)
Weak Union: $X \perp\!\!\!\perp Y, W | Z \Rightarrow X \perp\!\!\!\perp Y | W, Z$ (2.9)
Contraction: $(X \perp\!\!\!\perp W | Z, Y)$ & $(X \perp\!\!\!\perp Y | Z) \Rightarrow (X \perp\!\!\!\perp Y, W | Z)$ (2.10)

In a BN, $X_i \perp\!\!\!\perp ND(X_i) \mid Pa(X_i)$

ND = non-descendants
$Pa(X_i)$ = parents of X_i

$X_3 \perp\!\!\!\perp X_2, X_4 | X_1$ ...(1)
$X_4 \perp\!\!\!\perp X_1, X_3 | X_2$ ...(2)
what we have from the BN,

CI (1) and Decomposition give $X_3 \perp\!\!\!\perp X_2 | X_1$ ...(3) $X_3 \perp\!\!\!\perp X_4 | X_1$ ...(4)

similarly, $X_4 \perp\!\!\!\perp X_3 | X_2$ ...(5) $X_4 \perp\!\!\!\perp X_1 | X_2$ ...(6)

From CI(4): $P(X_4 | X_1) = P(X_4 | X_1, X_3)$ ... (7)

Now, from the previous graph, we also have $\underbrace{X_1 \perp\!\!\!\perp X_2}$ (X_1 does not have any parents)

$P(X_1 | X_2) = P(X_1)$

From CI(6), $P(X_1 | X_2, X_4) = P(X_1 | X_2)$
$= P(X_1)$

$\Rightarrow X_1 \perp\!\!\!\perp \{X_2, X_4\} \xrightarrow{\text{Decomposition}} X_1 \perp\!\!\!\perp X_4 \Rightarrow P(X_4 | X_1) = P(X_4)$ ...(8)

Combining (7) and (8), we get

$P(x_4) = P(x_4 | x_1, x_3) \Rightarrow x_4 \perp\!\!\!\perp \{x_1, x_3\} | \phi$

Again using decomposition we get $\Rightarrow x_4 \perp\!\!\!\perp x_3$

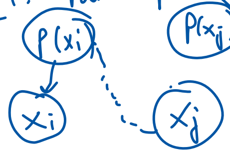
Hence we proved that $x_3 \perp\!\!\!\perp x_4$

..5

7. Prove that in a BN where x_i, x_j are any two nodes with no edge between them, then $x_i \perp\!\!\!\perp x_j | Pa(x_i), Pa(x_j)$. [Use d-separation.]

We need to apply d-separation to the paths between x_i and x_j.

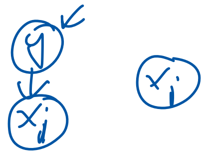
Case-1: Path passes through a node $\in \{Pa(x_i) \cup Pa\{x_j\}\}$.



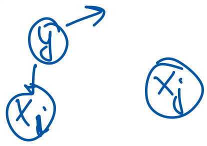
W.L.O.G let the path pass through one of the parents of $P(x_i)$ = call it node y

Case-1.1

edge direction into y then this path is blocked as $y \in Z$

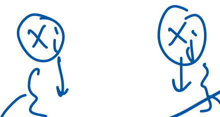


Case 1.2



again blocked because $y \in Z$

Case-2 Path does not pass through any node $\in \{Pa(x_i), Pa(x_j)\}$ then at some node in the path the direction will change and a V-node will be there: $\rightarrow x_k \leftarrow$ and $x_k \notin Z \Rightarrow$ path = blocked



$\therefore$ since, all paths are blocked, we get $(x_i \perp\!\!\!\perp x_j | Pa(x_i) \cup Pa(x_j))$.

edge directions will be like this as the other node is not a parent

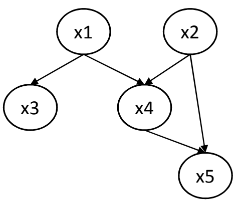
If its descendants $\in Z \Rightarrow$ descendant = parent

$\Rightarrow$ there is a directed cycle in a BN (DAG) $\Rightarrow$ contradiction!

(x_k can lie anywhere on the path and we will get this contradiction).

..5

8. For the Bayesian network G below, perform the following operations



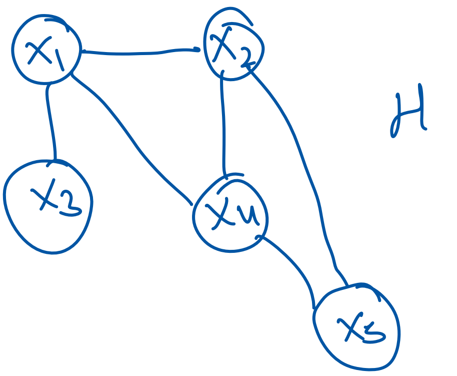
We will use local CI for constructing the UGM

For each vertex x_i, smallest subset U s.t. $x_i \perp\!\!\!\perp V-U-\{x_i\} | U$ in P. we make U the neighbours of x_i in H.

$U = MB(x_i) = Pa(x_i) \cup ch(x_i) \cup spouse(x_i)$

(a) Convert it into a undirected graphical model H.

Variable ordering: x_1, x_2, x_3, x_4, x_5



$MB(x_2) = x_1 \}$ spouse

$MB(x_3) = x_1 \}$ parent

$MB(x_4) = x_1, x_2 \}$ parents

$MB(x_5) = x_4, x_2 \}$ parents

(b) List two CIs that holds in G but do not hold in H.

(i) $x_1 \perp\!\!\!\perp x_2$ $\quad (x_1 \perp\!\!\!\perp \underbrace{ND(x_1)}_{x_2} \mid \underbrace{Pa(x_1)}_{\phi})$

but in H, $x_1 - x_2$ are connected

(ii) $x_2 \perp\!\!\!\perp x_3$ $\quad (x_2 \perp\!\!\!\perp \underbrace{ND(x_2)}_{x_3} \mid \underbrace{Pa(x_2)}_{\phi})$

but in H, x_2 and x_3 have a path in between

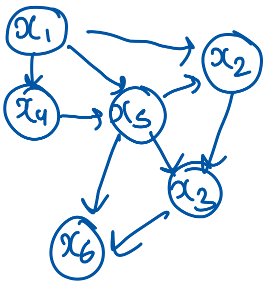
..2

9. For the undirected graphical model H below, perform the following operations

$$\begin{tikzcd} x_1 \arrow[r, dash] \arrow[d, dash] & x_2 \arrow[r, dash] \arrow[d, dash] & x_3 \arrow[d, dash] \\ x_4 \arrow[r, dash] & x_5 \arrow[r, dash] & x_6 \end{tikzcd}$$

For a UGM, if Z separates X and Y then $X \perp\!\!\!\perp Y \mid Z$

(a) Convert it into a BN G using variable order $x_1, x_4, x_5, x_2, x_3, x_6$.



G

We will add nodes one by one and check for CIs from H to draw parent-child relationships

(i) $x_4 \not\perp\!\!\!\perp x_1$ (hence we draw an edge)

(ii) $x_5 \not\perp\!\!\!\perp (x_1, x_4) \mid \phi$, $x_5 \not\perp\!\!\!\perp x_1 \mid x_4$, $x_5 \not\perp\!\!\!\perp x_4 \mid x_1$ (Hence connect to both x_1 & x_4)

(iii) $x_2 \perp\!\!\!\perp x_4 \mid \{x_1, x_5\}$ (others are false)

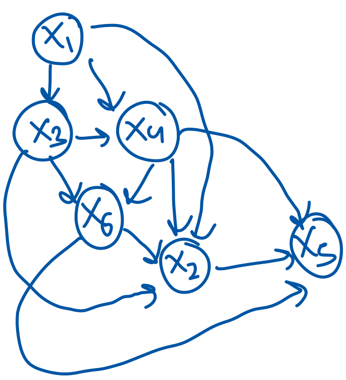
(iv) $x_3 \perp\!\!\!\perp \{x_1, x_4\} \mid \{x_2, x_5\}$ (others are false)

(v) $x_6 \perp\!\!\!\perp \{x_1, x_2, x_4\} \mid \{x_3, x_5\}$ (others are false)

..2

(b) Choose a different variable order that leads to adding more edges in G than in the above ordering.

Ordering chosen : $x_1, x_3, x_4, x_6, x_2, x_5$



Number of edges for this ordering $= 12$

Number of edges in $G = 9$

(i) $x_3 \not\perp\!\!\!\perp x_1$ (hence an edge drawn)

(ii) $x_4 \perp\!\!\!\perp \phi \mid \{x_1, x_3\}$ (All others false)

(iii) $x_6 \perp\!\!\!\perp x_1 \{x_3, x_4\}$

(iv) $x_2 \perp\!\!\!\perp \phi \mid \{x_1, x_3, x_4, x_6\}$

(v) $x_5 \perp\!\!\!\perp \{x_1, x_3\} \mid \{x_2, x_4, x_6\}$

..2

(c) List two CIs that holds in H but do not hold in G.

(i) $x_1 \perp\!\!\!\perp x_5 \mid \{x_2, x_4\}$ because $\{x_2, x_4\}$ separate x_1 from x_5 but we cannot deduce this from G as x_1 is the parent of x_5

(ii) $x_3 \perp\!\!\!\perp x_5 \mid \{x_2, x_6\}$ because $\{x_2, x_6\}$ separate x_3 and x_5 In G, x_5 is the parent of x_3

..2

$\boxed{\textbf{Total: 40}}$