AML Assignment-1

Aditya Neeraje

January 20, 2025

Contents

1 Question 1

Without loss of generality, assume the x_i are topologically ordered (from ancestors to descendants w.r.t to the DAG). This assumption can be done because the only difference in the formula made by permuting the indices by a permutation ψ is that π_i is replaced by some π_i' obtained by applying the permutation ψ to the set of parents of x_i.
Now, by factorizing the joint distribution as per the DAG, we get:

$$\begin{aligned} P(x_1, \cdots x_{n-1}) &= \sum_{x_1, \cdots x_{n-1}} P(x_1, \cdots x_{n-1}) \\ &= \prod_{i=1}^{n-1} f_i(x_i, x_{\pi_i}) \cdot \left(\sum_{x_n} f_n(x_n, x_{\pi_n}) \right) \\ &= \prod_{i=1}^{n-1} f_i(x_i, x_{\pi_i}) \end{aligned}$$

Proceeding inductively, we can observe:

$$\begin{aligned} P(x_1, \cdots x_j) &= \prod_{i=1}^{j} f_i(x_i, x_{\pi_i}) \cdot \prod_{i=j+1}^{n} \left(\sum_{x_i} f_i(x_i, x_{\pi_i}) \right) \\ &= \prod_{i=1}^{j} f_i(x_i, x_{\pi_i}) \cdot \prod_{i=j+1}^{n} 1 \\ &= \prod_{i=1}^{j} f_i(x_i, x_{\pi_i}) \end{aligned}$$

Thus, $P(x_i|\{x_1, \ldots, x_{i-1}\}) = \frac{P(x_1, \ldots, x_i)}{P(x_1, \ldots, x_{i-1})} = f_i(x_i, x_{\pi_i})$
However, by the fact that P factorizes over the given DAG, $P(x_i|\{x_1, \ldots, x_{i-1}\}) = P(x_i|x_{\pi_i})$
Thus, $f_i(x_i, x_{\pi_i}) = P(x_i|x_{\pi_i})$

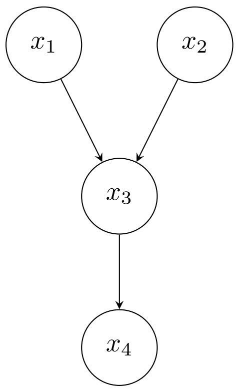
2 Question 2

2.1 Part a

$ZP(1, 0, 1, 1) = e^0 \cdot e^0 = 1$
$ZP(1, 1, 0, 0) = e^0 \cdot e^0 = 1$
$ZP(1, 1, 0, 1) = e^0 \cdot e^1 = e$
$ZP(1, 1, 1, 0) = e^1 \cdot e^1 = e^2$
$ZP(1, 1, 1, 1) = e^1 \cdot e^0 = e$

2.2 Part b

First, x_1 is added to the BN. It has no parents. When adding x_2, observe that $P(x_1, x_2) = \frac{1}{4}$ for all combinations of (x_1, x_2). This equals $P(x_1) \cdot P(x_2)$. Thus, $x_2 \perp\!\!\!\perp x_1$.
When x_3 is added, given both x_1 and x_2, the probability that $x_3 = x_1 \oplus x_2$ is $\frac{1}{e}$ times the probability of $x_3 = \neg(x_1 \oplus x_2)$. Similarly, x_3 is dependent on x_2 given x_1. Thus, we add edges from x_1 and x_2 to x_3.
When x_4 is added, given x_3, the probability that $x_4 = x_3$ is e times the probability that $x_4 = \neg x_3$ and is independent of x_1 and x_2. Thus, we add an edge from x_3 to x_4.



2.3 Part c

Node	**0**	**1**
x_1	$\frac{1}{2}$	$\frac{1}{2}$

Node	**0**	**1**
x_2	$\frac{1}{2}$	$\frac{1}{2}$

x_1, x_2	**(0, 0)**	**(0, 1)**	**(1, 0)**	**(1, 1)**
$x_3 = 0$	$\frac{1}{e+1}$	$\frac{e}{e+1}$	$\frac{e}{e+1}$	$\frac{1}{e+1}$
$x_3 = 1$	$\frac{e}{e+1}$	$\frac{1}{e+1}$	$\frac{1}{e+1}$	$\frac{e}{e+1}$

x_3	**0**	**1**
$x_4 = 0$	$\frac{1}{e+1}$	$\frac{e}{e+1}$
$x_4 = 1$	$\frac{e}{e+1}$	$\frac{1}{e+1}$

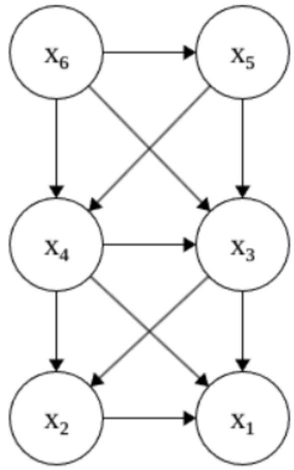
3 Question 3

3.1 Part a

x_1, x_2, x_3, x_4 are all unconditionally independent of x_6. This is because the only possible path from x_6 to any of them uses x_5 as the center of a v-structure, and x_5 is not observed and has no observed descendants when no variables are observed, and hence cannot be the center of a v-structure. x_5 is dependent on x_6 because they are directly connected.

3.2 Part b

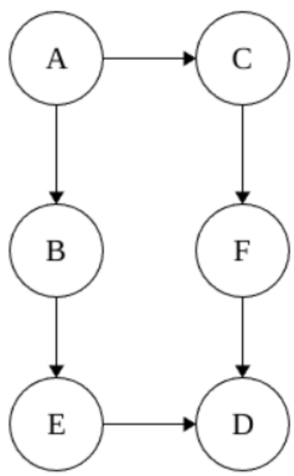
x_5 is dependent on x_6. $x_4 \not\perp\!\!\!\perp x_5|S$ for any S by virtue of a direct edge between them. $x_4 \not\perp\!\!\!\perp x_6|x_5$ by virtue of x_5 being a valid center of a v-structure.
$x_3 \not\perp\!\!\!\perp x_5|S$ and $x_3 \not\perp\!\!\!\perp x_4|S$ for any S. $x_3 \not\perp\!\!\!\perp x_6|\{x_4, x_5\}$ due to x_5 being a valid center of a v-structure.
$x_2 \not\perp\!\!\!\perp x_4|S$ for any S. $x_2 \not\perp\!\!\!\perp x_3|x_4$. Thus, $x_2 \perp\!\!\!\perp x_5, x_6|\{x_4, x_3\}$ is the best we can do.
$x_1 \not\perp\!\!\!\perp x_4|S$ for any S. $x_1 \not\perp\!\!\!\perp x_3|x_4$ and $x_1 \not\perp\!\!\!\perp x_2|x_4$. Thus, $x_1 \perp\!\!\!\perp x_5, x_6|\{x_4, x_3, x_2\}$ is the best we can do.



4 Question 4

4.1 Part 1

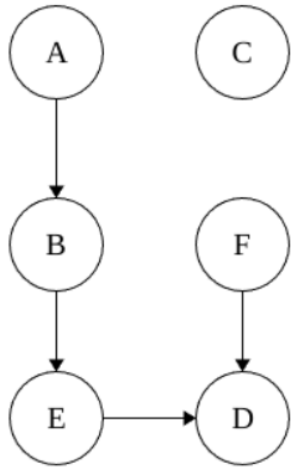
Consider the following BN for the given model:



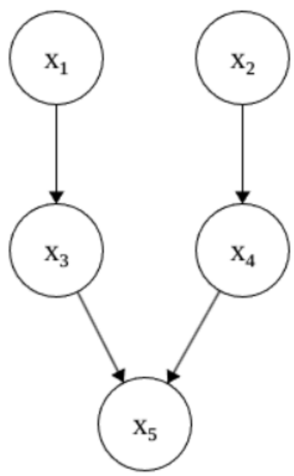
Since P factorizes over G, G is an I-map for P. Since $I(G) \subseteq I(P)$, the conditional independence $F \perp\!\!\!\perp B|A$, which holds in this graph, holds in P as well.

4.2 Part 2

We cannot guarantee that C is independent of E. In this very graph, C-A-B-E is an active path from C to E given no observed variables. We at the same time cannot guarantee that C is dependent on E. Consider the graph below, which satisfies all the conditional independencies satisfied in P and some more.



5 Question 5



6 Question 6

$$x_4 \perp\!\!\!\perp x_3, x_1 | x_2 (\text{given}) \tag{1}$$
$$x_4 \perp\!\!\!\perp x_3 | x_1, x_2 (\text{Weak Union}) \tag{2}$$
$$x_3 \perp\!\!\!\perp x_2 | x_1 (\text{given}) \tag{3}$$
$$x_3 \perp\!\!\!\perp x_4, x_2 | x_1 (\text{Contraction on 2 and 3}) \tag{4}$$
$$x_3 \perp\!\!\!\perp x_4 | x_1 (\text{Decomposition}) \tag{5}$$
$$x_4 \perp\!\!\!\perp x_1 | x_2 (\text{Decomposition on 1}) \tag{6}$$
$$x_1 \perp\!\!\!\perp x_2 (\text{given}) \tag{7}$$
$$x_1 \perp\!\!\!\perp x_4 (\text{Contraction on 6 and 7}) \tag{8}$$
$$x_4 \perp\!\!\!\perp x_3 (\text{Contraction on 5 and 8}) \tag{9}$$

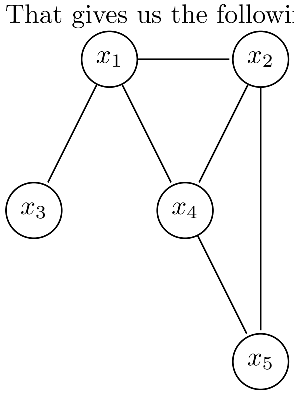
7 Question 7

Given that $Pa(x_i)$ and $Pa(x_j)$ are observed, notice that no path emerging out of x_i in the direction of one of its parents can be active (since the edge is directed towards x_i, the parent is not the center of a v-structure, and is observed, hence the parent is blocked).
Thus, any active path must come out of x_i towards one of its children. If we consider this path to keep going downwards in the graph until some node x_k which it enters from a parent of x_k and exits via another parent of x_k. Since x_k is the center of a v-structure, it has to be an ancestor of $Pa(x_j)$ (since if it was an ancestor of $Pa(x_i)$ and also a descendant of x_i, the graph would no longer be a DAG.)
Thus, x_j is a descendant of x_k and x_k is a descendant of x_i. Thus, x_j is a descendant of x_i.
By the same argument with x_i and x_j swapped, x_i is a descendant of x_j. Thus, x_i and x_j are descendants of each other.
This is not possible in a DAG.

8 Question 8

8.1 Part a

We use Proposition 4.8 from Koller-Friedman: Let G be any Bayesian network graph. The moralized graph M[G] is a minimal I-map for G.
That gives us the following diagram equivalent Markov network.



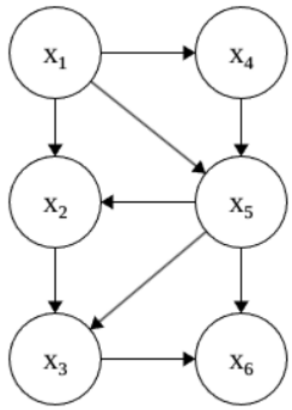
We now prove Proposition 4.8 from more basic principles for completeness. We need to show that our assumed Markov blanket of X, consisting of X's parents, children and parents of X's children, is indeed d-separating X from all other nodes and is the minimal subset with this property. Firstly, note that all children and parents of X are trivially in the Markov blanket of X, since the direct edge they share with X ensures that they are dependent on X given any subset of the other nodes. Also, since when $Ch(X)$ is observed and therefore can form the center of a v-structure, all parents of these children are dependent on X given any superset of $Ch(X)$ (unless the parent itself is in the superset, which we observe is in conformity with our claim (Proposition 4.8) anyway). Thus, the Markov blanket of X at least contains all of these points. To showcase that this subset of nodes d-separates X from the rest of the graph, observe that any path cannot exit X upwards (as its parents are observed). Any path exiting X towards its children must go towards a parent of its children since $Ch(X)$ is observed and hence these nodes can only be centers of v-structures. But these parents are observed and hence these paths are blocked.

8.2 Part b

$x_1 \perp\!\!\!\perp x_2$ and $x_1 \perp\!\!\!\perp x_2 | x_3$ are two CIs that hold in G but not in H.

9 Question 9

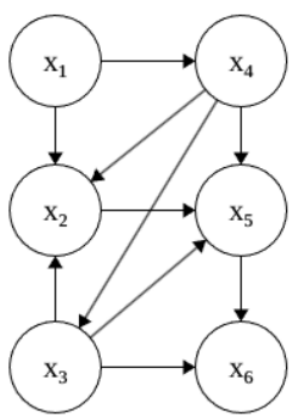
9.1 Part a



Justification: We use the algorithm for constructing an I-map of a given distribution given in Algorithm 3.2 of Friedman-Koller.
When x_5 is inserted, it has of course to be connected with x_4, but x_1 is also not independent of x_5 given x_4 in H. For x_2, x_5 and x_1 together separate it from x_4, and this blanket is minimal since it has to be connected to both x_1 and x_5. For x_3, x_2 is necessarily its parent. x_3 cannot be rendered independent of x_5 by any subset of $\{x_1, x_2, x_4\}$, so x_5 is also a parent. At this point, x_3 is separated from x_1 and x_4. For x_6, it is necessary to add its neighbours in H as its parents in G, and these d-separate it in H from all other vertices, so we are done. In total, this approach created 9 directed edges.

9.2 Part b

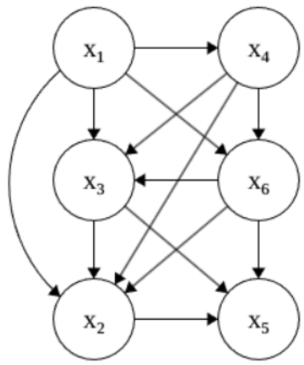
A simple change that generates more variables is to consider the order $x_1, x_4, x_3, x_2, x_5, x_6$.



When x_3 is inserted, no subset of $\{x_4, x_1\}$ d-separates it from the other vertex. When x_2 is now added, in the absence of x_5, it cannot be d-separated from x_4 by x_1, x_3 (note that due to monotonicity of d-separation in Markov networks, we do not need to separately consider subsets of $\{x_1, x_3\}$). x_5 needs to be connected with both its neighbours, and also with x_3 since $\{x_2, x_4\}$ does not d-separate it from x_3. x_6 has to be connected to all its neighbours, who form a Markov blanket.

In total, this approach created 10 directed edges.

In fact, note that one particular ordering gave me a directed graph with 13 nodes as the minimal BN which is an I-map for this Markov Network. That ordering was $x_1, x_4, x_6, x_3, x_2, x_5$.



This is provably the largest number of edges that a minimal and correct BN for this Markov network can have, since the last vertex we add need only be connected to its neighbours, hence the largest number of possible edges is $\frac{6 \cdot 5}{2} - 2 = 13$.

9.3 Part c

$x_3 \perp\!\!\!\perp x_5 | x_2, x_6$ and $x_5 \perp\!\!\!\perp x_1 | x_2, x_4$ are two CIs that hold in H but not in G (since x_3 and x_5, x_5 and x_1 now have edges between them in G, but were d-separated from each other by an appropriate set of vertices in H).