Probabilistic Graphical Models

Solution to HWK#1
Assigned Tuesday, Jan. 14, 2020
Due (both Parts A and B) : Thursday, Feb. 6, 2020

1 Fundamentals

Problem 1.1

1. Season $\perp$ Chills

   False: influence can flow along the path Season $\rightarrow$ Flu $\rightarrow$ Chills, since Flu is unobserved

2. Season $\perp$ Chills | Flu

   True: influence cannot flow through Flu, since it is observed; there are no other paths linking Season and Chills

3. Season $\perp$ Headache | Flu

   False: influence can flow along the path Season $\rightarrow$ Dehydration $\rightarrow$ Headache, since Dehydration is unobserved

4. Season $\perp$ Headache | Flu, Dehydration

   True: since both Flu and Dehydration are observed, influence cannot flow along any path that links Season and Headache

5. Season $\perp$ Nausea | Dehydration

   False: influence can flow along the path formed by Season $\rightarrow$ Flu $\rightarrow$ Headache $\rightarrow$ Dizziness $\rightarrow$ Nausea, since Flu, Headache, and Dizziness are unobserved

6. Season $\perp$ Nausea | Dehydration, Headache

   True: influence cannot flow along the path Season $\rightarrow$ Dehydration $\rightarrow$ Nausea, since Dehydration is observed; influence cannot flow along the path Season $\rightarrow$ Flu $\rightarrow$ Headache $\rightarrow$ Dizziness $\rightarrow$ Nausea, since Headache is observed; influence cannot flow along the path Season $\rightarrow$ Flu $\rightarrow$ Headache $\leftarrow$ Dehydration $\rightarrow$ Nausea, even though there is an observed v-structure centered at Headache, because Dehydration is observed

7. Flu $\perp$ Dehydration

   False: influence can flow along the path Flu $\leftarrow$ Season $\rightarrow$ Dehydration, since Season is unobserved

8. Flu $\perp$ Dehydration | Season, Headache

   False: influence can flow along the path Flu $\rightarrow$ Headache $\leftarrow$ Dehydration, since this is a v-structure and Headache is observed

9. Flu $\perp$ Dehydration | Season

   True: influence cannot flow through Season, which is observed, nor through Headache or Nausea, since both form v-structures and both are unobserved

10. Flu $\perp$ Dehydration | Season, Nausea

    False: influence can flow along the path Flu $\rightarrow$ Headache $\rightarrow$ Dizziness $\rightarrow$ Nausea $\leftarrow$ Dehydration, since Headache and Dizziness are unobserved and there is a v-structure at Nausea, which is observed

11. Chills $\perp$ Nausea

    False: influence can flow along the path Chills $\leftarrow$ Flu $\leftarrow$ Season $\rightarrow$ Dehydration $\rightarrow$ Nausea, since Flu, Season, and Dehydration are all unobserved

12. Chills $\perp$ Nausea | Headache

    False: influence can flow along the path Chills $\leftarrow$ Flu $\rightarrow$ Headache $\leftarrow$ Dehydration $\rightarrow$ Nausea, since there is a v-structure at Headache, which is observed

Problem 1.2

1. Using the <u>directed model</u> shown in Figure 1, write down the factorized form of the joint distribution over all of the variables, $P(S, F, D, C, H, N, Z)$.

   $$P(S, F, D, C, H, Z, N) = P(S)\, P(F|S)\, P(D|S)\, P(C|F)\, P(H|F, D)\, P(Z|H)\, P(N|D, Z)$$

2. Using the <u>undirected model</u> shown in Figure 2, write down the factorized form of the joint distribution over all of the variables, assuming the model is parameterized by one factor over each node and one over each edge in the graph.

   $$\frac{1}{Z}\, \phi_1(S)\, \phi_2(F)\, \phi_3(D)\, \phi_4(C)\, \phi_5(H)\, \phi_6(N)\, \phi_7(Z)$$
   $$\cdot\, \phi_8(S, F)\, \phi_9(S, D)\, \phi_{10}(F, C)\, \phi_{11}(F, H)\, \phi_{12}(D, H)\, \phi_{13}(D, N)\, \phi_{14}(H, Z)\, \phi_{15}(N, Z)$$

Problem 1.3

1. What is the probability that you have the flu, when no prior information is known?

   This translates to $P(\text{Flu} = \text{true})$

   $$\begin{aligned} &P(F = \text{true}) \\ &= \textstyle\sum_s P(F = \text{true}, S = s) \\ &= \textstyle\sum_s P(F = \text{true} \mid S = s) P(S = s) \\ &= P(F = \text{true} \mid S = \text{wint}) P(S = \text{wint}) + P(F = \text{true} \mid S = \text{summ}) P(S = \text{summ}) \\ &= 0.4 \cdot 0.5 + 0.1 \cdot 0.5 = 0.25 \end{aligned}$$

2. What is the probability that you have the flu, given that it is winter?

This translates to $P(\text{Flu} = \text{true} \mid \text{Season} = \text{winter})$

$P(F = \text{true} \mid S = \text{wint}) = 0.4$

3. What is the probability that you have the flu, given that it is winter and that you have a headache?

   This translates to $P(\text{Flu} = \text{true} \mid \text{Season} = \text{winter}, \text{Headache} = \text{true})$

$$
\begin{aligned}
&P(F = \text{true} \mid S = \text{wint}, H = \text{true}) \\
&= \frac{P(F = \text{true}, S = \text{wint}, H = \text{true})}{P(S = \text{wint}, H = \text{true})} \\
&= \frac{\sum_d P(F = \text{true}, S = \text{wint}, H = \text{true}, D = d)}{\sum_{f,d} P(F = f, S = \text{wint}, H = \text{true}, D = d)} \\
&= \frac{\sum_d P(H = \text{true} \mid F = \text{true}, D = d) P(F = \text{true} \mid S = \text{wint}) P(D = d \mid S = \text{wint}) P(S = \text{wint})}{\sum_{f,d} P(H = \text{true} \mid F = f, D = d) P(F = f \mid S = \text{wint}) P(D = d \mid S = \text{wint}) P(S = \text{wint})} \\
&= \frac{0.9 \cdot 0.4 \cdot 0.1 \cdot 0.5 + 0.8 \cdot 0.4 \cdot 0.9 \cdot 0.5}{0.9 \cdot 0.4 \cdot 0.1 \cdot 0.5 + 0.8 \cdot 0.4 \cdot 0.9 \cdot 0.5 + 0.8 \cdot 0.6 \cdot 0.1 \cdot 0.5 + 0.3 \cdot 0.6 \cdot 0.9 \cdot 0.5} \\
&= \frac{0.018 + 0.144}{0.018 + 0.144 + 0.024 + 0.081} = 0.61
\end{aligned}
$$

4. What is the probability that you have the flu, given that it is winter, you have a headache, and you know that you are dehydrated?

   This translates to $P(\text{Flu} = \text{true} \mid \text{Season} = \text{winter}, \text{Headache} = \text{true}, \text{Dehydration} = \text{true})$

$$
\begin{aligned}
&P(F = \text{true} \mid S = \text{wint}, H = \text{true}, D = \text{true}) \\
&= \frac{P(F = \text{true}, S = \text{wint}, H = \text{true}, D = \text{true})}{P(S = \text{wint}, H = \text{true}, D = \text{true})} \\
&= \frac{P(F = \text{true}, S = \text{wint}, H = \text{true}, D = \text{true})}{\sum_f P(F = f, S = \text{wint}, H = \text{true}, D = \text{true})} \\
&= \frac{P(H = \text{true} \mid F = \text{true}, D = \text{true}) P(F = \text{true} \mid S = \text{wint}) P(D = \text{true} \mid S = \text{wint}) P(S = \text{wint})}{\sum_f P(H = \text{true} \mid F = f, D = \text{true}) P(F = f \mid S = \text{wint}) P(D = \text{true} \mid S = \text{wint}) P(S = \text{wint})} \\
&= \frac{0.9 \cdot 0.4 \cdot 0.1 \cdot 0.5}{0.9 \cdot 0.4 \cdot 0.1 \cdot 0.5 + 0.8 \cdot 0.6 \cdot 0.1 \cdot 0.5} = \frac{0.018}{0.018 + 0.024} = 0.43
\end{aligned}
$$

5. Does knowing you are dehydrated increase or decrease your likelihood of having the flu? Intuitively, does this make sense?

   Knowing that you are dehydrated decreases the likelihood that you have the flu. This makes sense because the headache symptom is “explained away” by the dehydration.

Problem 1.4

Now consider the undirected model shown in Figure 2.

1. Are there any differences between the set of marginal independencies encoded by the directed and undirected versions of this model? If not, state the full set of marginal independencies encoded by both models. If so, give one example of a difference.

   There are no differences, because neither model encodes any marginal independencies at all.

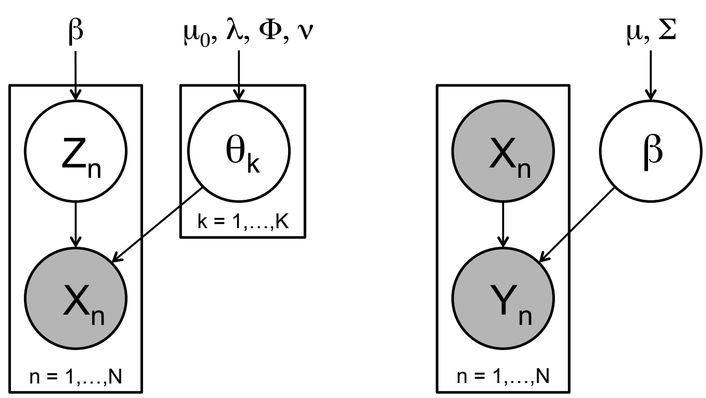
2. Are there any differences between the set of conditional independencies encoded by the directed and undirected versions of this model? If so, give one example of a difference.

   There are several differences. One example is that in the Markov network, we have Flu $\perp$ Dehydration | Season, Headache. However, this is not the case in the Bayesian network because observing Headache creates an active v-structure at Flu $\rightarrow$ Headache $\leftarrow$ Dehydration.

2 Bayesian Networks

Problem 2.1

The correct graphical models are shown below. Note that for Bayesian logistic regression, it's also correct to draw $\{X_n\}$ as a set of fixed "parameters" since they are technically not random variables.



Gaussian Mixture Model

Bayesian Logistic Regression

Problem 2.2

Prove the following theorem.

Theorem 1. *Let $\mathcal{G}$ be a Bayesian Network structure over a set of random variables $\mathcal{X}$ and let P be a joint distribution over $\mathcal{X}$. If P factorizes according to $\mathcal{G}$, then $\mathcal{G}$ is an I-map for P.*

Answer:

Given that P factorizes according to $\mathcal{G}$, we have that:

$$P(X_1, \ldots, X_n) = \prod_{i=1}^{n} P(X_i \mid \mathrm{Pa}(X_i)), \tag{1}$$

where $X_1, \ldots, X_n$ are the random variables corresponding the nodes of graph $\mathcal{G}$ and $\mathrm{Pa}(X_i)$ denotes the parents of node X_i in graph $\mathcal{G}$. Furthermore, let $\mathrm{ND}(X_i)$ denote the non-descendant nodes of node X_i in graph $\mathcal{G}$ and $\mathrm{D}(X_i)$ denote its descendant nodes. Using the above factorization, we can see that:

$$\begin{aligned}
P(X_i \mid \mathrm{ND}(X_i)) &= \frac{P(X_i, \mathrm{ND}(X_i))}{P(\mathrm{ND}(X_i))}, \\
&= \frac{\sum_{\mathrm{D}(X_i)} P(X_1, \ldots, X_n)}{\sum_{X_i, \mathrm{D}(X_i)} P(X_1, \ldots, X_n)}, \\
&= \frac{\sum_{\mathrm{D}(X_i)} \prod_{j=1}^{n} P(X_j \mid \mathrm{Pa}(X_j))}{\sum_{X_i, \mathrm{D}(X_i)} \prod_{j=1}^{n} P(X_j \mid \mathrm{Pa}(X_j))}, \\
&= \frac{\prod_{X_j \in (\mathrm{ND}(X_i) \cup X_i)} P(X_j \mid \mathrm{Pa}(X_j)) \sum_{\mathrm{D}(X_i)} \prod_{X_j \in \mathrm{D}(X_i)} P(X_j \mid \mathrm{Pa}(X_j))}{\prod_{X_j \in \mathrm{ND}(X_i)} P(X_j \mid \mathrm{Pa}(X_j)) \sum_{X_i, \mathrm{D}(X_i)} \prod_{X_j \in (\mathrm{D}(X_i) \cup X_i)} P(X_j \mid \mathrm{Pa}(X_j))}, \\
&= \frac{\prod_{X_j \in (\mathrm{ND}(X_i) \cup X_i)} P(X_j \mid \mathrm{Pa}(X_j)) \cdot 1}{\prod_{X_j \in \mathrm{ND}(X_i)} P(X_j \mid \mathrm{Pa}(X_j)) \cdot 1}, \\
&= P(X_i \mid \mathrm{Pa}(X_i)).
\end{aligned} \tag{2}$$

This means that:

$$\{X_i \perp \mathrm{ND}(X_i) \mid \mathrm{Pa}(X_i); i = 1, \ldots, n\}, \tag{3}$$

which by the local Markov assumptions we know to be the independence assertions included in set $I(\mathcal{G})$ (using the notation from our lecture notes). Therefore, we see that those assertions are included in $I(P)$, i.e. $I(\mathcal{G}) \subseteq I(P)$, and so we have proven that if P factorizes according to $\mathcal{G}$, then $\mathcal{G}$ is an I-map for P.

Problem 2.3

Prove the following theorem:

Let $\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{Z}$ be three disjoint sets of nodes in a Bayesian network $\mathcal{G}$. Let $\boldsymbol{U} = \boldsymbol{X} \cup \boldsymbol{Y} \cup \boldsymbol{Z}$, and let $\mathcal{G}' = \mathcal{G}^+[\boldsymbol{U}]$ be the induced Bayesian network over $\boldsymbol{U} \cup \mathrm{Ancestors}_{\boldsymbol{U}}$. Let $\mathcal{H}$ be the moralized graph $\mathcal{M}[\mathcal{G}']$. Then $\text{d-sep}_{\mathcal{G}}(\boldsymbol{X}; \boldsymbol{Y} \mid \boldsymbol{Z})$ if and only if $\mathrm{sep}_{\mathcal{H}}(\boldsymbol{X}; \boldsymbol{Y} \mid \boldsymbol{Z})$.

Answer:

Suppose not *d-sep*$_{\mathcal{G}}(\boldsymbol{X}; \boldsymbol{Y} \mid \boldsymbol{Z})$. Then there is an active path in $\mathcal{G}$ between some $X \in \boldsymbol{X}$ and some $Y \in \boldsymbol{Y}$. The active path, like any path, is formed of overlapping unidirectional segments $W_1 \to \ldots \to W_n$. Note that W_n in each segment is either X, Y, or the base of some active v-structure. As such, either W_n or one of its descendants is in $\boldsymbol{U}$, and so all nodes and edges in teh segment $W_1 \to \ldots \to W_n$ are in the induced graph over *Ancestors*$(\boldsymbol{U}) \cup \boldsymbol{U}$. So the active path in $\mathcal{G}$ is a path in $\mathcal{H}$, and in $\mathcal{H}$, the path can only contain members of $\boldsymbol{Z}$ at the bases of v-structures in $\mathcal{G}$ (otherwise, the path would have been active in $\mathcal{G}$). Because $\mathcal{H}$ is the moralized graph $\mathcal{M}[\mathcal{G}']$, we know that those members of $\boldsymbol{Z}$ have been bypassed: their parents in the v-structure must have an edge between them in $\mathcal{H}$. So there is an active path between X and Y in $\mathcal{H}$, so not *sep*$_{\mathcal{H}}(\boldsymbol{X}; \boldsymbol{Y} \mid \boldsymbol{Z})$.

Now we prove the converse. Suppose that $\boldsymbol{X}$ and $\boldsymbol{Y}$ are d-separated given $\boldsymbol{Z}$. Consider an arbitrary path in $\mathcal{G}$ between some $X \in \boldsymbol{X}$ and some $Y \in \boldsymbol{Y}$. Any path between X and Y in $\mathcal{G}$ must either be blocked by a member of $\boldsymbol{Z}$ or an inactive v-structure in $\mathcal{G}$. First, suppose the path is blocked by a member of $\boldsymbol{Z}$. Then the path in $\mathcal{H}$ (if it exists–it may not because $\mathcal{H}$ is the induced graph) will also be blocked by that member of $\boldsymbol{Z}$.

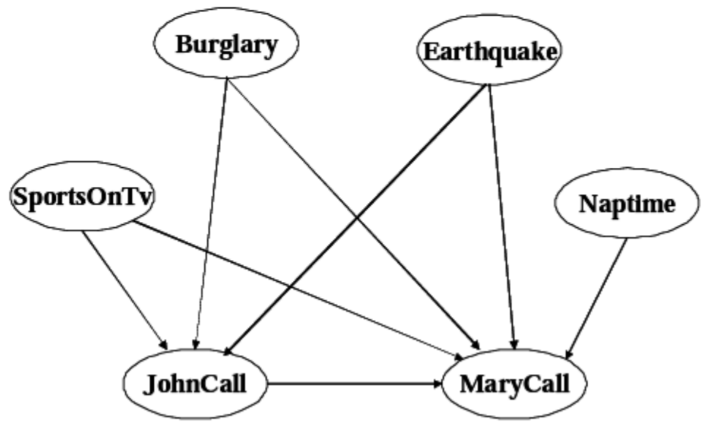
Of course, if the path between X and Y in $\mathcal{G}$ is not blocked by a member of $\boldsymbol{Z}$, it must be blocked by an inactive v-structure. Because the v-structure is inactive, its base (and any of its descendants) must not be in the induced graph $\mathcal{H}$. As such, the path will not exist in $\mathcal{H}$. Neither the base nor its descendants can be in $\boldsymbol{Z}$, and they cannot be in $\boldsymbol{X}$ or $\boldsymbol{Y}$ either, because then we would have an active path from a member of $\boldsymbol{X}$ to a member of $\boldsymbol{Y}$.

Recall that the edges added in moralization are necessary to create an active path in $\mathcal{H}$ only when paths would be blocked by the observed root node of a v-structure in $\mathcal{G}$. In this case, the segment would have been active in $\mathcal{G}$, so moralization edges cannot effect segments in $\mathcal{H}$ corresponding to inactive segments in $\mathcal{G}$; this is what our proof depends on. Because of all the above, there are no active paths in $\mathcal{H}$ between arbitrary $X \in \boldsymbol{X}$ and $Y \in \boldsymbol{Y}$, so we have *sep*$_{\mathcal{H}}(\boldsymbol{x}; \boldsymbol{Y} \mid \boldsymbol{Z})$.

Problem 2.4

Exercise 3.11 (part a only) from the book, at page 98.

Answer:



In order to construct a minimal I-map, we would like to preserve all independencies (assuming Alarm is always unobserved) that were present in the original graph, without adding any unnecessary edges. Let's start with the remaining nodes and add edges only as needed.

We see that with Alarm unobserved, there exist active paths between Alarm's direct ancestors and children. Thus, direct edges between the parents, Burglary and Earthquake, should be added to connect to both children, JohnCall and Mary Call. Similarly, since any two children of Alarm also now have an active path between them, a direct edge between JohnCall and MaryCall should be added. Without loss of generality, we direct this edge to go from JohnCall to MaryCall.

Next, since SportsOnTv and JohnCall as well as Naptime and MaryCall were directly connected in the original graph, removing Alarm doesn't affect their dependencies and the two edges must be preserved.

Now we must consider any independencies that may have changed. In the old graph, because of the v-structure between Alarm and co-parent SportsOnTv, if Alarm was unobserved and JohnCall observed, there existed an active path between SportsOnTv and MaryCall. In the new graph however, because of the added direct edge between the two children JohnCall and MaryCall, if JohnCall is observed, the path between SportsOnTv and MaryCall is actually rendered inactive. Thus, an alternate path that does not introduce any other dependencies needs to be introduced, and a direct edge is added between SportsOnTv and MaryCall.

Problem 2.5

Exercise 4.10 from the book, at page 154.

In this problem we are only considering the independences defined by the separation i.e. $I = I(H) = \{\boldsymbol{X} \perp \boldsymbol{Y} | \boldsymbol{Z}: sep_H(\boldsymbol{X}; \boldsymbol{Y} | \boldsymbol{Z})\}$.

<u>Strong Union: $(\boldsymbol{X} \perp \boldsymbol{Y} | \boldsymbol{Z}) \Rightarrow (\boldsymbol{X} \perp \boldsymbol{Y} | \boldsymbol{Z}, \boldsymbol{W})$</u>

Since, $\boldsymbol{X} \perp \boldsymbol{Y} | \boldsymbol{Z}$ by definition we have $sep_H(\boldsymbol{X}; \boldsymbol{Y} | \boldsymbol{Z})$.

1. If $\boldsymbol{W}$ is a subset of $\boldsymbol{Z}$ then the statement is trivially true because $sep_H(\boldsymbol{X}; \boldsymbol{Y} | \boldsymbol{Z} \cup \boldsymbol{W})\}$ is the same as $sep_H(\boldsymbol{X}; \boldsymbol{Y} | \boldsymbol{Z})$
2. If $\boldsymbol{W}$ is not a subset of $\boldsymbol{Z}$ then we can define $\boldsymbol{W_1} := \boldsymbol{W} - \boldsymbol{Z}$ and $\boldsymbol{W_2} := \boldsymbol{W} \cap \boldsymbol{Z}$. $sep_H(\boldsymbol{X}; \boldsymbol{Y} | \boldsymbol{Z} \cup \boldsymbol{W_2})\}$ follows from Part 1. Since $(\boldsymbol{X} \perp \boldsymbol{Y} | \boldsymbol{Z})$ we know that $\boldsymbol{Z}$ separates $\boldsymbol{X}$ and $\boldsymbol{Y}$. Thus, $\boldsymbol{W_1}$ which is disjoint from $\boldsymbol{Z}$ cannot have a path connecting $\boldsymbol{X}$ and $\boldsymbol{Y}$, because that would be a contradiction to $sep_H(\boldsymbol{X}; \boldsymbol{Y} | \boldsymbol{Z})$. Thus $sep_H(\boldsymbol{X}; \boldsymbol{Y} | \boldsymbol{Z} \cup \boldsymbol{W_1})\}$ is true which is the same as $sep_H(\boldsymbol{X}; \boldsymbol{Y} | \boldsymbol{Z} \cup \boldsymbol{W})\}$.

<u>Transitivity: $\neg(\boldsymbol{X} \perp A | \boldsymbol{Z})$ & $\neg(A \perp \boldsymbol{Y} | \boldsymbol{Z}) \Rightarrow \neg(\boldsymbol{X} \perp \boldsymbol{Y} | \boldsymbol{Z})$</u>

If for some node A, the following relation $\neg(\boldsymbol{X} \perp A | \boldsymbol{Z})$ & $\neg(A \perp \boldsymbol{Y} | \boldsymbol{Z})$ is true then $\neg(\boldsymbol{X} \perp \boldsymbol{Y} | \boldsymbol{Z})$.

1. Since, $\neg(\boldsymbol{X} \perp A | \boldsymbol{Z}) \Rightarrow \neg sep_H(\boldsymbol{X}; A | \boldsymbol{Z})$, there is a path between $\boldsymbol{X}$ and A that does not go through any nodes of $\boldsymbol{Z}$.
2. Since, $\neg(\boldsymbol{Y} \perp A | \boldsymbol{Z}) \Rightarrow \neg sep_H(\boldsymbol{Y}; A | \boldsymbol{Z})$, there is a path between $\boldsymbol{Y}$ and A that does not go through any nodes of $\boldsymbol{Z}$.

Thus, we have a path from $\boldsymbol{X}$ to $\boldsymbol{Y}$ through this node A that does not go through any nodes of $\boldsymbol{Z}$. Thus, the sets $\boldsymbol{X}$ and $\boldsymbol{Y}$ are not separated by $\boldsymbol{Z}$. Therefore, $\neg(\boldsymbol{X} \perp \boldsymbol{Y} | \boldsymbol{Z})$.