

CS 6190: Probabilistic Machine Learning Spring 2022

Homework 3

Handed out: 15 Mar, 2022
Due: 11:59pm, 29 Mar, 2022

- You are welcome to talk to other members of the class about the homework. I am more concerned that you understand the underlying concepts. However, you should write down your own solution. Please keep the class collaboration policy in mind.
- Feel free discuss the homework with the instructor or the TAs.
- Your written solutions should be brief and clear. You need to show your work, not just the final answer, but you do not need to write it in gory detail. Your assignment should be **no more than 10 pages**. Every extra page will cost a point.
- Handwritten solutions will not be accepted.
- The homework is due by **midnight of the due date**. Please submit the homework on Canvas.

Analytical problems [100 points + 40 bonus]

1. [13 points] The joint distribution over three binary variables are given in Table 1. Show by direct evaluation that this distribution has the property that a and b are marginally dependent, so that $p(a, b) \neq p(a)p(b)$, but that they become independent conditioned on c , i.e., $p(a, b|c) = p(a|c)p(b|c)$.
Answer: Let us first compute $p(a), p(b), p(a|c), p(b|c)$. In view of Table 2, we can see that

$$p(a = 0)p(b = 0) \neq p(a = 0, b = 0)$$

which indicates that a and b are not independent. Also, Table 3 shows that for each value of a, b, c , we have the equality $p(a|c)p(b|c) = p(a, b|c)$ which means a and b are conditioned on c .

a	b	c	p(a,b,c)
0	0	0	0.192
0	0	1	0.144
0	1	0	0.048
0	1	1	0.216
1	0	0	0.192
1	0	1	0.064
1	1	0	0.048
1	1	1	0.096

Table 1: Joint distribution of a, b, c .

a	p(a)	b	p(b)	c	p(c)	and	a	b	p(a, b)
0	0.6	0	0.592	0	0.48		0	0	0.336
1	0.4	1	0.408	1	0.52		0	1	0.264
							1	0	0.256
							1	1	0.144

Table 2: Distributions of $p(a), p(b), p(c)$ and $p(a, b)$.

a	c	p(a c)	and	b	c	p(b c)	and	a	b	c	p(a, b c)
0	0	0.5		0	0	0.8		0	0	0	0.4
1	0	0.5		1	0	0.2		0	1	0	0.1
								1	0	0	0.4
								1	1	0	0.1
0	1	0.6923		0	1	0.4		0	0	1	0.2769
1	1	0.3076		1	1	0.6		0	1	1	0.4154
								1	0	1	0.1232
								1	1	1	0.18462

Table 3: Conditional distributions of $a|c$ and $b|c$.

2. [12 points] Using the d -separation algorithm/criterion (Bayes ball algorithm) to show that the conditional distribution for a node x in a directed graph, conditioned on all of the nodes in its Markov blanket, is independent of the remaining variables in the graph.

Answer: Let x be an arbitrary vertex in an acyclic graph T and denotes its Markov blanket $MB(x)$. Assume that $P = xz_1z_2 \cdots z_m = y$ is a path from x to a vertex y outside of $MB(x)$. In view of the d -separation d -separation criterion, to prove the desired assertion, we need to prove that the path P is blocked. Distinguish the following cases:

- (1) z_1 is a child of x . In this case, if z_2 is a child of z_1 , then the arrows on the path meet the vertex z_1 in head-to-tail. Since z_1 is in $MB(x)$, the path is blocked. If z_2 is a parent of z_1 , then $z_2 \in MB(x)$ (thus $z_2 \neq y$) and the arrows on the path meet the vertex z_2 either in head-to-tail or tail-to-tail (depending on whether z_3 is a parent of a child of z_2) which implies that the path is blocked.
 - (2) z_1 is a parent of x . In this case, if z_2 is a child of z_1 , then the arrows on the path meet the vertex z_1 in tail-to-tail and if z_2 is a parent of z_1 , then the arrows on the path meet the vertex z_1 in head-to-tail. Therefore, since z_1 is in $MB(x)$, the path is blocked.
3. [15 points] See the graphical model in Figure 1. Recall what we have discussed in the class. Show that $a \perp b | \emptyset$. Suppose we have observed the variable d . Show that in general $a \not\perp b | d$.

Answer: According to the graphical model,

$$p(a, b, c, d) = p(a)p(b)p(c|a, b)p(d|c).$$

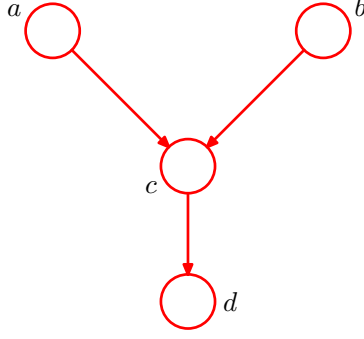


Figure 1: Graphical model.

Therefore,

$$\begin{aligned}
 p(a, b) &= \int_{c,d} p(a)p(b)p(c|a, b)p(d|c)dcdd \\
 &= p(a)p(b) \int_c cp(c|a, b) \underbrace{\int_d p(d|c)dd}_{=1} dc \\
 &= p(a)p(b) \underbrace{\sum_c p(c|a, b)}_{=1} dc = p(a)p(b)
 \end{aligned}$$

which implies that $a \perp b | \emptyset$. To prove $a \not\perp b | d$,

$$\begin{aligned}
 p(a, b|d) &= \frac{p(a, b, d)}{p(d)} \\
 &= \frac{\int_c p(a, b, c, d)dc}{p(d)} \\
 &= \frac{\int_c p(a)p(b)p(c|a, b)p(d|c)dc}{p(d)} \\
 &= \frac{p(a)p(b) \int_c p(c|a, b)p(d|c)dc}{p(d)}.
 \end{aligned}$$

However, $\int_c p(c|a, b)p(d|c)dc$ generally cannot be decompose into $m(a, d) \times n(b, d)$ and consequently,

$$p(a, b|d) \neq p(a|d)p(b|d).$$

4. [10 points] Convert the directed graphical model in Figure 1 into an undirected graphical model. Draw the structure and write down the definition of the potential functions.

Answer: The converted undirected graphical model is shown in Figure 2. In view of the directed graphical model in Figure 1, we obtain $p(a, b, c, d) = p(a)p(b)p(c|a, b)p(d|c)$. We can rewrite it as

$$p(a, b, c, d) = \underbrace{p(a)p(b)p(c|a, b)}_{\psi_1(a, b, c)} \underbrace{p(d|c)}_{\psi_2(c, d)}$$

and set $Z = 1$. Therefore,

$$p(a, b, c, d) = \psi_1(a, b, c)\psi_2(c, d),$$

where $\psi_1(a, b, c) = p(a)p(b)p(c|a, b)$ and $\psi_2(c, d) = p(d|c)$.

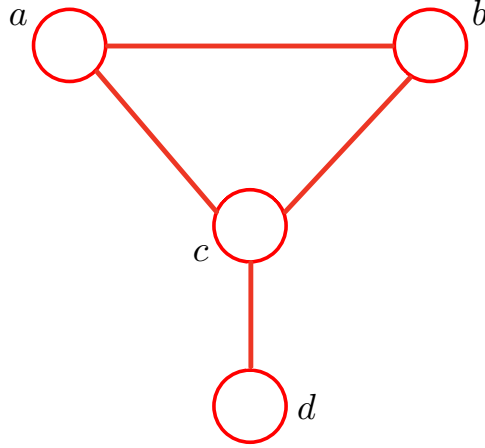


Figure 2: Undirected graphical model of Figure 1

5. [15 points] Write down every step of the sum-product algorithm for the graphical model shown in Figure 3. Note that you need to first choose a root node, and write down how to compute each message. Once all your messages are ready, please explain how to compute the marginal distribution $p(x_4, x_5)$.

Answer: We set x_4 to be the root. We have three leaf nodes x_1 , x_3 , and f_d . Also, note that unnormalized joint distribution is given by

$$\tilde{p}(x_1, x_2, x_3, x_4, x_5) = f_a(x_1, x_2, x_3) f_b(x_2, x_4) f_c(x_4, x_5) f_d(x_5).$$

• **Propagation messages from the leaf nodes out to the root node.**

We have the following eight messages.

$$\mu_{f_d \rightarrow x_5}(x_5) = f_d(x_5) \tag{1}$$

$$\mu_{x_5 \rightarrow f_c}(x_5) = \prod_{l \in \text{ne}(x_5) \setminus f_c} \mu_{f_l \rightarrow x_5}(x_5) = f_d(x_5) \tag{2}$$

$$\begin{aligned} \mu_{f_c \rightarrow x_4}(x_4) &= \sum_{x_5} f_c(x_4, x_5) \mu_{x_5 \rightarrow f_c}(x_5) \\ &= \sum_{x_5} f_c(x_4, x_5) f_d(x_5) \end{aligned} \tag{3}$$

$$\mu_{x_1 \rightarrow f_a}(x_1) = 1 \tag{4}$$

$$\mu_{x_3 \rightarrow f_a}(x_3) = 1 \tag{5}$$

$$\begin{aligned} \mu_{f_a \rightarrow x_2}(x_2) &= \sum_{x_1, x_3} f_a(x_1, x_2, x_3) \mu_{x_1 \rightarrow f_a}(x_1) \mu_{x_3 \rightarrow f_a}(x_3) \\ &= \sum_{x_1, x_3} f_a(x_1, x_2, x_3) \end{aligned} \tag{6}$$

$$\begin{aligned}
\mu_{x_2 \rightarrow f_b}(x_2) &= \prod_{l \in ne(x_2) \setminus f_b} \mu_{f_l \rightarrow x_2}(x_2) \\
&= \mu_{f_a \rightarrow x_2}(x_2) \\
&= \sum_{x_1, x_3} f_a(x_1, x_2, x_3)
\end{aligned} \tag{7}$$

$$\begin{aligned}
\mu_{f_b \rightarrow x_4}(x_4) &= \sum_{x_2} f_b(x_2, x_4) \mu_{x_2 \rightarrow f_b}(x_2) \\
&= \sum_{x_2} f_b(x_2, x_4) \sum_{x_1, x_3} f_a(x_1, x_2, x_3)
\end{aligned} \tag{8}$$

• Propagation messages from the root node out to the leaf nodes.

$$\begin{aligned}
\mu_{x_4 \rightarrow f_c}(x_4) &= \prod_{l \in ne(x_4) \setminus f_c} \mu_{f_l \rightarrow x_4}(x_4) \\
&= \mu_{f_b \rightarrow x_4}(x_4)
\end{aligned} \tag{9}$$

$$= \sum_{x_2} f_b(x_2, x_4) \sum_{x_1, x_3} f_a(x_1, x_2, x_3) \tag{10}$$

$$\begin{aligned}
\mu_{f_c \rightarrow x_5}(x_5) &= \sum_{x_4} f_c(x_4, x_5) \mu_{x_4 \rightarrow f_c}(x_4) \\
&= \sum_{x_4} f_c(x_4, x_5) \sum_{x_2} f_b(x_2, x_4) \sum_{x_1, x_3} f_a(x_1, x_2, x_3)
\end{aligned} \tag{11}$$

$$\begin{aligned}
\mu_{x_5 \rightarrow f_d}(x_5) &= \prod_{l \in ne(x_5) \setminus f_d} \mu_{f_l \rightarrow x_5}(x_5) \\
&= \mu_{f_c \rightarrow x_5}(x_5) \\
&= \sum_{x_4} f_c(x_4, x_5) \sum_{x_2} f_b(x_2, x_4) \sum_{x_1, x_3} f_a(x_1, x_2, x_3)
\end{aligned} \tag{12}$$

$$\begin{aligned}
\mu_{x_4 \rightarrow f_b}(x_4) &= \prod_{l \in ne(x_4) \setminus f_b} \mu_{f_l \rightarrow x_4}(x_4) \\
&= \mu_{f_c \rightarrow x_4}(x_4) \\
&= \sum_{x_5} f_c(x_4, x_5) f_d(x_5)
\end{aligned} \tag{13}$$

$$\begin{aligned}
\mu_{f_b \rightarrow x_2}(x_2) &= \sum_{x_4} f_b(x_2, x_4) \mu_{x_4 \rightarrow f_b}(x_4) \\
&= \sum_{x_4} f_b(x_2, x_4) \sum_{x_5} f_c(x_4, x_5) f_d(x_5)
\end{aligned} \tag{14}$$

$$\begin{aligned}
\mu_{x_2 \rightarrow f_a}(x_2) &= \prod_{l \in ne(x_2) \setminus f_a} \mu_{f_l \rightarrow x_2}(x_2) \\
&= \mu_{f_b \rightarrow x_2}(x_2) \\
&= \sum_{x_4} f_b(x_2, x_4) \sum_{x_5} f_c(x_4, x_5) f_d(x_5)
\end{aligned} \tag{15}$$

$$\begin{aligned}
\mu_{f_a \rightarrow x_1}(x_1) &= \sum_{x_2, x_3} f_a(x_1, x_2, x_3) \mu_{x_2 \rightarrow f_a}(x_2) \mu_{x_3 \rightarrow f_a}(x_3) \\
&= \sum_{x_2, x_3} f_a(x_1, x_2, x_3) \sum_{x_4} f_b(x_2, x_4) \sum_{x_5} f_c(x_4, x_5) f_d(x_5)
\end{aligned} \tag{16}$$

$$\begin{aligned}
\mu_{f_a \rightarrow x_3}(x_3) &= \sum_{x_1, x_2} f_a(x_1, x_2, x_3) \mu_{x_1 \rightarrow f_a}(x_1) \mu_{x_2 \rightarrow f_a}(x_2) \\
&= \sum_{x_1, x_2} f_a(x_1, x_2, x_3) \sum_{x_4} f_b(x_2, x_4) \sum_{x_5} f_c(x_4, x_5) f_d(x_5)
\end{aligned} \tag{17}$$

Unnormalized Marginal $\tilde{p}(x_4, x_5)$:

$$\begin{aligned}
\tilde{p}(x_4, x_5) &= f_c(x_4, x_5) \mu_{f_d \rightarrow x_5}(x_5) \mu_{f_b \rightarrow x_4}(x_4) \\
&= f_c(x_4, x_5) f_d(x_5) \sum_{x_2} f_b(x_2, x_4) \sum_{x_1, x_3} f_a(x_1, x_2, x_3)
\end{aligned}$$

Therefore,

$$p(x_4, x_5) = \frac{\tilde{p}(x_4, x_5)}{\sum_{x_4, x_5} \tilde{p}(x_4, x_5)}.$$

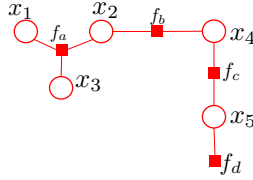


Figure 3: Factor graph.

6. [10 points] Now if x_2 in Figure 3 is observed, explain how to conduct the sum-product algorithm, and compute the posterior distribution $p(x_4, x_5 | x_2)$.

Answer: When $x_2 = \alpha$ is observed, we can run the sum-product algorithm exactly similar to the previous question while we are keeping in mind that the range of x_2 is only $\{\alpha\}$. In other words, in any summation over x_2 , the only possible value for x_2 to take is α . Therefore, we can repeat exactly what we did in the previous question, and finally, set the range of x_2 to be only $\{\alpha\}$.

Unnormalized Marginal $\tilde{p}(x_4, x_5 | x_2)$:

$$\begin{aligned}
\tilde{p}(x_4, x_5 | x_2) &= f_c(x_4, x_5) \mu_{f_d \rightarrow x_5}(x_5) \mu_{f_b \rightarrow x_4}(x_4) \\
&= f_c(x_4, x_5) f_d(x_5) \sum_{x_2 \in \{\alpha\}} f_b(x_2, x_4) \sum_{x_1, x_3} f_a(x_1, x_2, x_3) \\
&= f_c(x_4, x_5) f_d(x_5) f_b(\alpha, x_4) \sum_{x_1, x_3} f_a(x_1, \alpha, x_3)
\end{aligned}$$

Therefore,

$$p(x_4, x_5 | x_2) = \frac{\tilde{p}(x_4, x_5 | x_2)}{\sum_{x_4, x_5} \tilde{p}(x_4, x_5 | x_2)}.$$

7. [10 points] Suppose all the random variables in Figure 3 are discrete, and no one has been observed. Now we want to find the configuration of the x_1, \dots, x_5 to maximize the joint probability. Write down every step of the max-sum algorithm to calculate the maximum joint probability and to find the corresponding configurations of each random variable.

Answer: We set x_4 to be the root. We have three leaf nodes x_1 , x_3 , and f_d .

$$\mu_{f_d \rightarrow x_5}(x_5) = \ln f_d(x_5) \quad (18)$$

$$\mu_{x_5 \rightarrow f_c}(x_5) = \sum_{l \in ne(x_5) \setminus f_c} \mu_{f_l \rightarrow x_5}(x_5) = \ln f_d(x_5) \quad (19)$$

$$\begin{aligned} \mu_{f_c \rightarrow x_4}(x_4) &= \max_{x_5} [\ln f_c(x_4, x_5) + \mu_{x_5 \rightarrow f_c}(x_5)] \\ &= \max_{x_5} [\ln f_c(x_4, x_5) + \ln f_d(x_5)] \end{aligned} \quad (20)$$

$$\phi_{f_c \rightarrow x_4}(x_4) = \arg \max_{x_5} [\ln f_c(x_4, x_5) + \ln f_d(x_5)] \quad (21)$$

$$\mu_{x_1 \rightarrow f_a}(x_1) = 0 \quad (22)$$

$$\mu_{x_3 \rightarrow f_a}(x_3) = 0 \quad (23)$$

$$\begin{aligned} \mu_{f_a \rightarrow x_2}(x_2) &= \max_{x_1, x_3} [\ln f_a(x_1, x_2, x_3) + \mu_{x_1 \rightarrow f_a}(x_1) + \mu_{x_3 \rightarrow f_a}(x_3)] \\ &= \max_{x_1, x_3} \ln f_a(x_1, x_2, x_3) \end{aligned} \quad (24)$$

$$\phi_{f_a \rightarrow x_2}(x_2) = \arg \max_{x_1, x_3} [\ln f_a(x_1, x_2, x_3)] \quad (25)$$

$$\begin{aligned} \mu_{x_2 \rightarrow f_b}(x_2) &= \sum_{l \in ne(x_2) \setminus f_b} \mu_{f_l \rightarrow x_2}(x_2) \\ &= \mu_{f_a \rightarrow x_2}(x_2) \\ &= \max_{x_1, x_3} \ln f_a(x_1, x_2, x_3) \end{aligned} \quad (26)$$

$$\begin{aligned} \mu_{f_b \rightarrow x_4}(x_4) &= \max_{x_2} [\ln f_b(x_2, x_4) + \mu_{x_2 \rightarrow f_b}(x_2)] \\ &= \max_{x_2} [\ln f_b(x_2, x_4) + \max_{x_1, x_3} \ln f_a(x_1, x_2, x_3)] \end{aligned} \quad (27)$$

$$\phi_{f_b \rightarrow x_4}(x_4) = \arg \max_{x_2} [\ln f_b(x_2, x_4) + \max_{x_1, x_3} \ln f_a(x_1, x_2, x_3)] \quad (28)$$

Using Equations (20) and (27), we obtain the maximum of log-joint probability as follows

$$\begin{aligned} \max \ln \tilde{p}(x_1, \dots, x_5) &= \max_{x_4} \sum_{l \in ne(x_4)} \mu_{f_l \rightarrow x_4}(x_4) \\ &= \max_{x_4} [\mu_{f_b \rightarrow x_4}(x_4) + \mu_{f_c \rightarrow x_4}(x_4)] \\ &= \max_{x_4} \left[\max_{x_5} [\ln f_c(x_4, x_5) + \ln f_d(x_5)] + \max_{x_2} [\ln f_b(x_2, x_4) + \max_{x_1, x_3} \ln f_a(x_1, x_2, x_3)] \right] \end{aligned} \quad (29)$$

To find the corresponding configurations of each random variable, using 29, we first get

$$x_4^* = \arg \max_{x_4} \sum_{l \in ne(x_4)} \mu_{f_l \rightarrow x_4}(x_4).$$

Then, recursively, we obtain

$$\begin{aligned} x_5^* &= \phi_{f_c \rightarrow x_4}(x_4^*) && \text{using 21} \\ x_2^* &= \phi_{f_b \rightarrow x_4}(x_4^*) && \text{using 28} \\ x_1^*, x_3^* &= \phi_{f_a \rightarrow x_2}(x_2^*) && \text{using 25} \end{aligned}$$

8. **[Bonus]**[20 points] Show the message passing protocol we discussed in the class is always valid on the tree-structured graphical models— whenever we compute a message (from a factor to a variable or a variable to a factor), the dependent messages are always available. Hint: use induction.

Answer: We prove the following statement by induction on the number of vertices of the tree.

For any tree T with a given root, if the messages at the leaf nodes are available, then the message passing protocol from leaf nodes to root and, after that, the message passing protocol from the root node to leaf nodes are valid at each step on T .

Proof. If the tree has only one vertex, then clearly the statement is true, indeed, there is no message to pass. If the tree has only one edge, then, the outgoing message from the leaf node to the root is available by the assumption, and then the outward message from the root back to the leaf node is available as well.

Now, assume that the statement is true for all the trees with at most N vertices. Let T be a given tree with $N + 1$ vertices for which we want to prove the statement with the root vertex x (variable node). Consider the path P from the root x to a leaf vertex v whose length is maximum. Note that all the siblings of v are also leaf nodes (Since the length of P is maximum). We distinguish the two following cases.

- Assume that $v = g$ is a factor node and z (variable node) is its unique parent. In this case, the message that z should send to its unique parent f_z is

$$\mu_{z \rightarrow f_z}(z) = \prod_{f \in \text{Children}(z)} f(z).$$

Let T' be a tree obtained from T by removing all the children of z (which are leaves). By the induction hypothesis, since T' has fewer vertices than T , we know that the message-passing protocol is always valid for T with any initialization on its leaf nodes. Note that z is a leaf node for T' and we initiate its message to its unique parent by $\mu_{z \rightarrow f_z}(z)$ and the message of other leaf nodes are initiating as normal. However, running the message-passing protocol on T' in this situation completes the message propagation from leaf nodes to root node in T .

- Assume that $v = y_1$ is a variable node and g (factor node) is its unique parent. Let $\text{Children}(g) = \{y_1, \dots, y_k\}$. In this case, the message that g sends to its unique parent z is

$$\mu_{g \rightarrow z}(z) = g(y_1, \dots, y_k).$$

Let T' be a tree obtained from T by removing all the vertices y_1, \dots, y_k (which are leaves). By the induction hypothesis, since T' has fewer vertices than T , we know that the message-passing protocol is always valid for T with any initialization on its leaf nodes. Note that g is a leaf node for T' and we initiate its message to its unique parent by $\mu_{g \rightarrow z}(z)$ and the message of other leaf nodes are initiating as normal. However, running the message-passing protocol on T' in this situation completes the message propagation from the leaf nodes to the root node in T .

Similar to the above discussion, for the propagation of the outward messages from the root back to the leaves, we first follow the propagation schedule for the tree T' which is possible due to the induction hypothesis, and when we want to compute the message from v to its children u , we have access to all the message from its other children to v by initiation (we know how they are initiated). Therefore, we can compute the propagation of the outward messages as well.

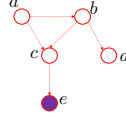


Figure 4: Model 1.

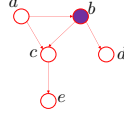


Figure 5: Model 2.

9. [15 points] Use d-separation algorithm (Bayes ball) to determine if $a \perp\!\!\!\perp d|e$ in the graphical model shown in Figure 4, and if $a \perp\!\!\!\perp d|b$ in the graphical model shown in Figure 5.

Answer:

- $a \not\perp\!\!\!\perp d|e$ in the graphical model shown in Figure 4.
To conclude that $a \not\perp\!\!\!\perp d|e$, it suffices to find a path from a to d which is not blocked. Consider the path $P = acbd$. Note that in this path we have two internal vertices c and b . Note that b is a head-to-tail vertex which is not observed and c is a head-to-head vertex with an observed descendant e . Therefore, this path is not blocked and, in view of d-separation algorithm, $a \not\perp\!\!\!\perp d|e$.
- $a \perp\!\!\!\perp d|b$ in the graphical model shown in Figure 5.
There are exactly the two paths $P_1 = abd$ and $P_2 = acbd$ from a to d in graph shown in Figure 5. In the path P_1 , the vertex b is in $C = \{b\}$ and also is head-to-tail. Thus, this path is blocked. In the path P_2 , the vertex b is in $C = \{b\}$ and also is tail-to-tail. Therefore, this path is also blocked as well. Since, every path from a to d is blocked, $\{a\}$ is d -separated from $\{d\}$ by $\{b\}$, we have $a \perp\!\!\!\perp d|b$.

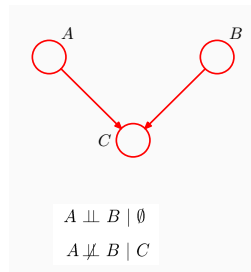


Figure 6: Directed.

10. **[Bonus]**[20 points] We have listed two examples in the class to show that in terms of the expressiveness (i.e., conditional independence) of the directed and undirected graphical models, there is not a guarantee that who is better than who.

- (a) [10 points] Now show that for the directed graphical model in Figure 6, we cannot find an equivalent undirected graphical model to express the same set of conditional independence.

Answer: By Figure 6, $p(A, B, C) = P(A)P(B)P(C|A, B)$. Also, we assume that the model in Figure 6 perfectly presents p . For a contradiction, consider an undirected graphical model with the same vertex presenting p perfectly. Since, as we have the term $P(C|A, B)$ in $P(A, B, C)$ and $A \not\perp\!\!\!\perp B|C$, any undirected graphical model expressing the same conditional independence should have vertices A, B and C as a clique (fully connected). Therefore, the undirected graphical model

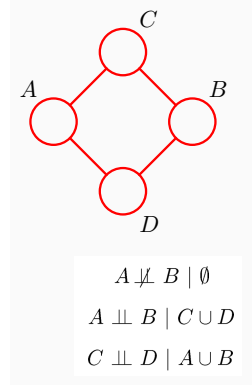


Figure 7: Undirected.

should be a triangle with vertices A, B , and C . In conclusion, there is an edge linking A to B in the undirected graphical model. Note that the undirected graphical model perfectly describes the distribution p which implies $A \not\perp\!\!\!\perp B \mid \emptyset$, a contradiction.

- (b) [10 points] Show that for the undirected graphical model in Figure 7, we cannot find an equivalent directed graphical model to express the same set of conditional independence.

Answer: For a contradiction, assume that there is a directed model that perfectly presents $p(A, B, C, D) = \psi_1(A, C)\psi_2(C, B)\psi_3(B, D)\psi_4(D, A)$. Since $A \not\perp\!\!\!\perp B \mid \emptyset$, by d -separation theorem, we should have a path from A to B which is not blocked by \emptyset . So, there is a path from A to B going through C or D . Therefore, we have one of the cases in Figure 8). Each cases is denied by the following reason (clockwise).

- $A \cup B \not\perp\!\!\!\perp C \cup D$
- The graph has a cycle.
- $B \cup D \not\perp\!\!\!\perp A \cup B$
- $B \cup D \not\perp\!\!\!\perp A \cup B$

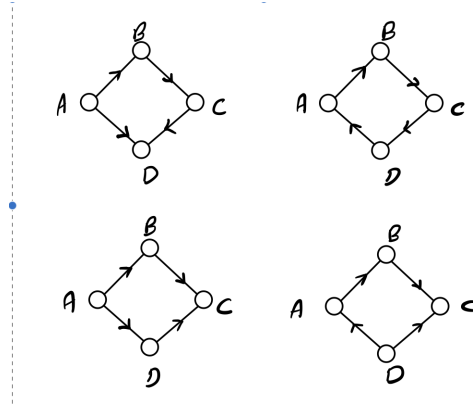


Figure 8: Directed.