CS 6190: Probabilistic Modelling Spring 2019

Homework 3

Handed out: 14 Oct, 2019
Due: 11:59pm, 28 Oct, 2019

- 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 when c, so that $p(a, b|c) = p(a|c)p(b|c)$.

   Please see the figure. 1

2. [12 points] Using the d-separation algorithm/criterion, show that the conditional distribution for a node x in a directed graph, conditioned on all of the nodes in the Markov blanket, is independent of the remaining variables in the graph. As shown in figure 2, let's consider the following cases:

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	b	$p(a,b)$
0	0	0.336
0	1	0.264
1	0	0.256
1	1	0.144

(a) $p(a,b))$

a	c	$p(a,c)$	$p(a\|c)$
0	0	0.240	0.5
0	1	0.360	$\frac{9}{13}$
1	0	0.240	0.5
1	1	0.160	$\frac{9}{13}$

(b) $p(a|c)$

b	c	$p(b,c)$	$p(b\|c)$
0	0	0.384	0.8
0	1	0.208	0.4
1	0	0.096	0.2
1	1	0.312	0.6

(c) $p(b|c)$

a	$p(a)$
0	0.6
1	0.4

(d) $p(a)$

b	$p(b)$
0	0.592
1	0.408

(e) $p(b)$

c	$p(c)$
0	0.480
1	0.520

(f) $p(c)$

a	b	$p(a,b)$	$p(a)p(b)$
0	0	0.336	0.3552
0	1	0.264	0.2448
1	0	0.256	0.2368
1	1	0.144	0.1632

(g) $p(a,b) \neq p(a)p(b)$

a	b	c	$p(a,b\|c)$	$p(a\|c)p(b\|c)$
0	0	0	0.4	0.4
0	0	1	18/65	18/65
0	1	0	0.1	0.1
0	1	1	27/65	27/65
1	0	0	0.4	0.4
1	0	1	8/65	8/65
1	1	0	0.1	0.1
1	1	1	12/65	12/65

(h) $p(a,b|c)p(a|c)p(b|c)$

Figure 1: $p(a,b|c) = p(a|c)p(b|c)$

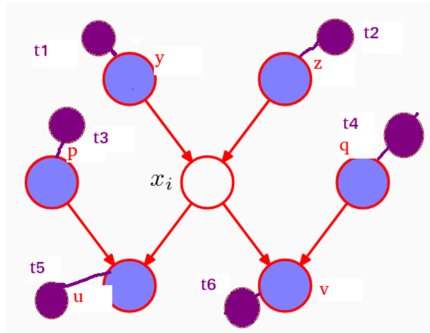


Figure 2: Markov blanket

- Sub-tree connected to x_i via y or z :Given y and z, any variable node present in sub-trees t_1 and t_2 will be linked by head-to-tail path to x_i. Thus, x_i will be independent to them given y and z.
- Sub-tree connected to x_i via p or q : Given p and q, all paths from nodes lying in subt-ree t_3 and t_4 to x_i will be head-to-tail and are blocked. Thus, x_i is independent to them given p and q.
- Sub-tree connected to x_i via u or v : Given u and v, we have head-to-tail path from x_i to variable nodes lying in subtree t_5 and t_6 and thus x_i will be independent to them given u and v.

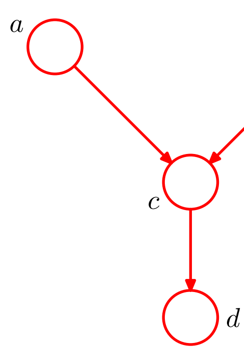


Figure 3: Graphical model.

3. [15 points] See the graphical model in Figure 3. 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 $a \not\perp b|d$.

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

$$\begin{aligned} p(a,b) &= p(a)p(b)\int p(c|a,b)\left(\int p(d|c)dd\right)dc \\ &= p(a)p(b)\int p(c|a,b)dc \\ &= p(a)p(b) \end{aligned}$$

Thus a and b are independent if conditioned on nothing

$$p(a,b|d) = \frac{p(a,b,d)}{p(d)}$$
$$p(a,b,d) = p(a)p(b)\int p(c|a,b)p(d|c)dc$$
$$= p(a)p(b)g(a,b,d) \neq m(a,d)n(b,d) \tag{1}$$

Thus $p(a,b|d) \neq p(a|d)p(b|d)$. Hence a and b are not independent if d is observed.

4. [10 points] Convert the directed graphical model in Figure 3 into an undirected graphical model. Draw the structure and write down the definition of the potential functions.

$$p(a,b,c,d) = p(a)p(b)p(c|a,b)p(d|c)$$
$$\psi_{a,b,c} = p(a)p(b)p(c|a,b)$$
$$\psi_{c,d} = p(d|c)$$

Please see figure 4 for the corresponding undirected graph

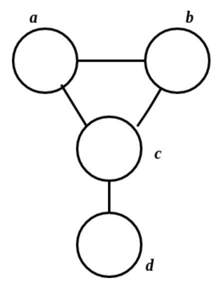


Figure 4: Undirected graph for graph shown in figure 2

5. [15 points] Write down every step of the sum-product algorithm for the graphical model shown in Figure 5. 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 $p(x_4, x_5)$.

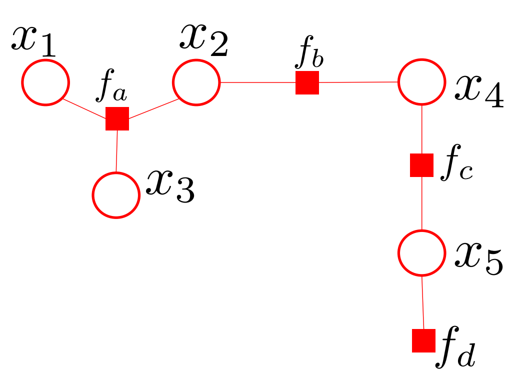


Figure 5: Factor graph.

Let x_1 be the root node. Messages from leaf node to the root node :

$$\begin{aligned}
\mu_{f_d \to x_5}(x_5) &= f_d(x_5) \\
\mu_{x_5 \to f_c}(x_5) &= f_d(x_5) \\
\mu_{f_c \to x_4}(x_4) &= \sum_{x_5} f_c(x_4, x_5) f_d(x_5) \\
\mu_{x_4 \to f_b}(x_4) &= \mu_{f_c \to x_4}(x_4) \\
\mu_{f_b \to x_2}(x_2) &= \sum_{x_4} f_b(x_2, x_4) \mu_{x_4 \to f_b}(x_4) \\
\mu_{x_2 \to f_a}(x_2) &= \mu_{f_b \to x_2}(x_2) \\
\mu_{x_3 \to f_a}(x_3) &= 1 \\
\mu_{f_a \to x_1}(x_1) &= \sum_{x_2, x_3} f_a(x_1, x_2, x_3) \mu_{x_2 \to f_a}(x_2) \mu_{x_3 \to f_a}(x_3)
\end{aligned}$$

Messages from the root node to the leaves are as follows :

$$\begin{aligned}
\mu_{x_1 \to f_a}(x_1) &= 1 \\
\mu_{f_a \to x_3}(x_3) &= \sum_{x_1, x_2} f_a(x_1, x_2, x_3) \mu_{x_1 \to f_a}(x_1) \mu_{x_2 \to f_a}(x_2) \\
\mu_{f_a \to x_2}(x_2) &= \sum_{x_1, x_3} f_a(x_1, x_2, x_3) \mu_{x_1 \to f_a}(x_1) \mu_{x_3 \to f_a}(x_3) = \sum_{x_1, x_3} f_a(x_1, x_2, x_3) \\
\mu_{x_2 \to f_b}(x_2) &= \mu_{f_a \to x_2}(x_2) \\
\mu_{f_b \to x_4}(x_4) &= \sum_{x_2} f_b(x_2, x_4) \mu_{x_2 \to f_b}(x_2) \\
\mu_{x_4 \to f_c}(x_4) &= \mu_{f_b \to x_4}(x_4) \\
\mu_{f_c \to x_5}(x_5) &= \sum_{x_4} f_c(x_4, x_5) \mu_{x_4 \to f_c}(x_4) \\
\mu_{x_5 \to f_d}(x_5) &= \mu_{f_c \to x_5}(x_5)
\end{aligned}$$

To compute the joint-probability of x_4 and x_5, the messages from f_b and f_d will be multiplied.

$$p(x_4, x_5) = \frac{\mu_{f_b \to x_4}(x_4) \mu_{f_d \to x_5}(x_5)}{\sum_{x_4, x_5} \mu_{f_b \to x_4}(x_4) \mu_{f_d \to x_5}(x_5)}$$

6. [10 points] Now if x_2 in Figure 5 is observed, explain how to conduct the sum-product algorithm, and compute the posterior distribution $p(x_4, x_5|x_2)$.

If x_2 is observed and let $x_2 = c$ be the observed value for x_2, then the forward and backward messages assuming x_1 to be the root node will change. Instead of summing up for all possible values of x_2, the messsages will be computed for only $x_2 = c$.Forward messages are :

$$\begin{aligned}
&\mu_{f_d \to x_5}(x_5) = f_d(x_5) \\
&\mu_{x_5 \to f_c}(x_5) = f_d(x_5) \\
&\mu_{f_c \to x_4}(x_4) = \sum_{x_5} f_c(x_4, x_5) f_d(x_5) \\
&\mu_{x_4 \to f_b}(x_4) = \mu_{f_c \to x_4}(x_4) \\
&\mu_{f_b \to x_2}(x_2) = \sum_{x_4} f_b(x_2, x_4) \mu_{x_4 \to f_b}(x_4) \\
&\mu_{x_2 \to f_a}(x_2) = \mu_{f_b \to x_2}(x_2) \\
&\mu_{x_3 \to f_a}(x_3) = 1 \\
&\mu_{f_a \to x_1}(x_1) = \mu_{x_2 \to f_a}(x_2 = c) \sum_{x_3} f_a(x_1, x_2 = c, x_3) \mu_{x_3 \to f_a}(x_3)
\end{aligned}$$

Backward messages are as follows :

$$\begin{aligned}
&\mu_{x_1 \to f_a}(x_1) = 1 \\
&\mu_{f_a \to x_3}(x_3) = \mu_{x_2 \to f_a}(x_2 = c) \sum_{x_1} f_a(x_1, x_2 = c, x_3) \mu_{x_1 \to f_a}(x_1) \\
&\mu_{f_a \to x_2}(x_2) = \sum_{x_1, x_3} f_a(x_1, x_2, x_3) \mu_{x_1 \to f_a}(x_1) \mu_{x_3 \to f_a}(x_3) = \sum_{x_1, x_3} f_a(x_1, x_2, x_3) \\
&\mu_{x_2 \to f_b}(x_2) = \mu_{f_a \to x_2}(x_2 = c) \\
&\mu_{f_b \to x_4}(x_4) = f_b(x_2 = c, x_4) \mu_{x_2 \to f_b}(x_2 = c) \\
&\mu_{x_4 \to f_c}(x_4) = \mu_{f_b \to x_4}(x_4) \\
&\mu_{f_c \to x_5}(x_5) = \sum_{x_4} f_c(x_4, x_5) \mu_{x_4 \to f_c}(x_4) \\
&\mu_{x_5 \to f_d}(x_5) = \mu_{f_c \to x_5}(x_5)
\end{aligned}$$

Once we have the updated messages, the $p(x_4, x_5|x_2)$ can be computed similarly to $p(x_4, x_5)$

$$\begin{aligned}
p(x_2) &= \frac{1}{Z_2} \mu_{f_a \to x_2}(x_2) \mu_{f_b \to x_2}(x_2) \\
p(x_4, x_5 | x_2 = c) &= \frac{1}{Z} \frac{\mu_{f_b \to x_4}(x_4) \mu_{f_d \to x_5}(x_5)}{p(x_2 = c)}
\end{aligned}$$

7. [10 points] Suppose all the random variables in Figure 5 are discrete, and no one has been observed. Now we want to find the configuration of the $x_1, \ldots, x_5$ to maximize the joint probability. Write done every step of the max-sum algorithm to calculate the maximum joint probability and to find the corresponding configurations of each random variable.

$$\begin{aligned}
p(x_1, x_2, .., x_5) &= \prod f_a(x_1, x_2, x_3) f_b(x_2, x_4) f_c(x_4, x_5) f_d(x_5) \\
log(p(x_1, x_2, .., x_5)) &= \sum_i log(f_i) \\
\max_x (log(p(x_1, x_2, .., x_5))) &= \max_x \sum_i log(f_i) \\
&= \max_{x_1, x_2, x_3} (f_a(x_1, x_2, x_3) + \max_{x_4} (log(f_b(x_2, x_4)) \\
&+ \max_{x_5} (log(f_c(x_4, x_5)) + log(f_d(x_5))))) \qquad (2)
\end{aligned}$$

Let x_1 be the root node. Messages from leaf node to the root node :

$$\mu_{f_d \to x_5}(x_5) = \max \log f_d(x_5)$$
$$\mu_{x_5 \to f_c}(x_5) = \max \log f_d(x_5)$$
$$\mu_{f_c \to x_4}(x_4) = \max_{x_5}(\log f_c(x_4, x_5) + \log f_d(x_5))$$
$$\mu_{x_4 \to f_b}(x_4) = \mu_{f_c \to x_4}(x_4)$$
$$\mu_{f_b \to x_2}(x_2) = \max_{x_4}(\log f_b(x_2, x_4) + \log \mu_{x_4 \to f_b}(x_4))$$
$$\mu_{x_2 \to f_a}(x_2) = \mu_{f_b \to x_2}(x_2)$$
$$\mu_{x_3 \to f_a}(x_3) = 1$$
$$\mu_{f_a \to x_1}(x_1) = \max_{x_2, x_3}(\log f_a(x_1, x_2, x_3) + \log \mu_{x_2 \to f_a}(x_2) + \log \mu_{x_3 \to f_a}(x_3))$$

Messages from the root node to the leaves are as follows :

$$\mu_{x_1 \to f_a}(x_1) = 1$$
$$\mu_{f_a \to x_3}(x_3) = \max_{x_1, x_2}(\log f_a(x_1, x_2, x_3) + \log \mu_{x_1 \to f_a}(x_1) + \log \mu_{x_2 \to f_a}(x_2))$$
$$\mu_{f_a \to x_2}(x_2) = \max_{x_1, x_3}(\log f_a(x_1, x_2, x_3) + \log \mu_{x_1 \to f_a}(x_1) + \log \mu_{x_3 \to f_a}(x_3)) = \max_{x_1, x_3} \log f_a(x_1, x_2, x_3)$$
$$\mu_{x_2 \to f_b}(x_2) = \mu_{f_a \to x_2}(x_2)$$
$$\mu_{f_b \to x_4}(x_4) = \max_{x_2}(\log f_b(x_2, x_4) + \log \mu_{x_2 \to f_b}(x_2))$$
$$\mu_{x_4 \to f_c}(x_4) = \mu_{f_b \to x_4}(x_4)$$
$$\mu_{f_c \to x_5}(x_5) = \max_{x_4}(\log f_c(x_4, x_5) + \log \mu_{x_4 \to f_c}(x_4))$$
$$\mu_{x_5 \to f_d}(x_5) = \mu_{f_c \to x_5}(x_5)$$

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.

The tree-structured graphical models have the following properties :

- The graph is connected i.e there exists a path between each and every pair of the graph vertices(node)
- The graph is acyclic i.e there exists a unique path between each and every pair of the graph vertices(node)

Let a tree-structured graphical model has n variables $x_1, x_2, x_3 ... x_n$ and m factor nodes $f_1, f_2, ..., f_n$. Let x_r be the root nodes. Let x_l, f_l be the set of leaf-variable nodes and leaf-factor nodes. Also let the parent of any node x During forward pass, as we start from the leaves, the message is passed using the following algorithm: During the message flow from the root node to the leaves, all the messages

for $x_i \in x_l$ do :
$\mu_{x_i \to p(X_i)}(x_i) = 1$
for $f_i \in f_l$ do :
$\mu_{f_i \to p(f_i)} = f_i(p(f_i))$
level = height-1
while level 0 :
compute message from each x_l at level l to its parent

if root = variable:
$\mu_{root \to child_j} = 1$
else:
$\mu_{root \to child_j} = \sum_{child_i \neq child_j} f_r$
$level = 0$
$while \quad level < height - 1:$
compute message from each x_l at level l to its child node using the forward messages

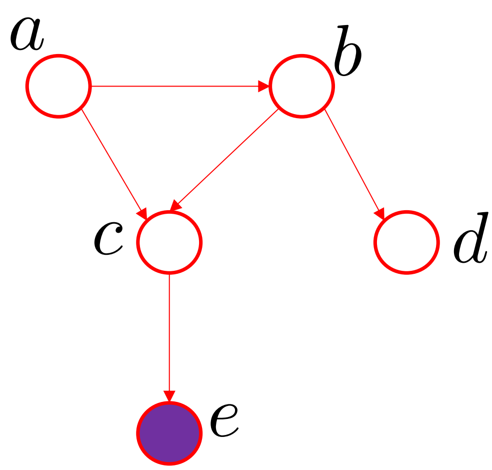


Figure 6: Model 1.

from the lower level nodes have already been computed. Thus, for all the paths from the root node to the leaves, backward messages can be computed using the above algorithm.

9. [15 points] Use d-separation algorithm to determine if $a \perp\!\!\!\perp d|e$ in the graphical model shown in Figure 6, and if $a \perp\!\!\!\perp d|b$ in the graphical model shown in Figure 7.

   - Model 1 : As there exists a path from a to d after deleting the givens, a and d are not required to be conditionally independent given e. See figure 8
   - Model 2 : As there exists no path from a to d after deleting the givens, a and d are conditionally independent given b. See figure 9

   Help taken from http://web.mit.edu/jmn/www/6.034/d-separation.pdf

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 10, we cannot find an equivalent undirected graphical model to express the same set of conditional independence.
    The corresponding undirected graph is shown in figure 12(a). As shown in figure, if nothing is observed, there still exist a path from a to b and thus a and b are not independent given nothing. However, they are also not independent given c.

    (b) [10 points] Show that for the undirected graphical model in Figure 11, we cannot find an equivalent directed graphical model to express the same set of conditional independence.
    As a is not independent of b, there exist a path from a to b via c or d. Now, let's assume graph to be figure 12(b), for c to be independent of d given a and b, they must be connected

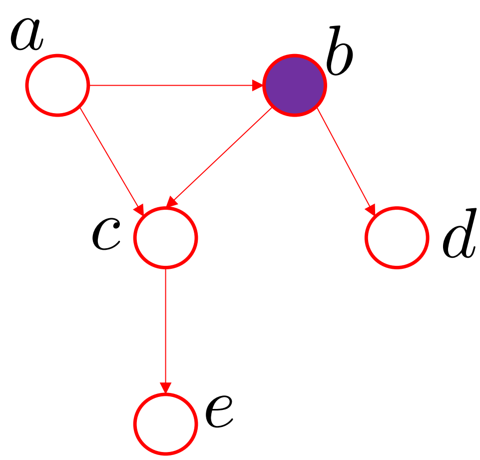
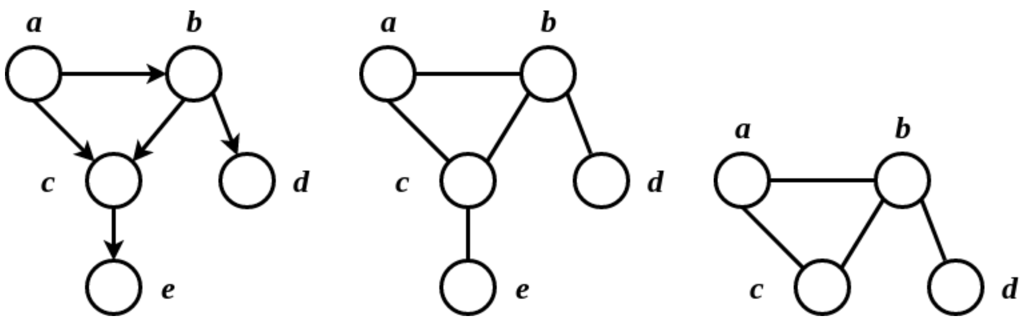


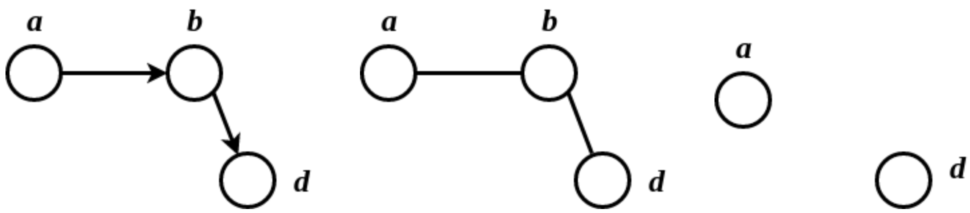
Figure 7: Model 2.



(a) Ancestral Graph (b) Moralise and disorient (c) Delete givens

Figure 8: Model 1

by tail-to-tail or head-to tail at b and c. This configuration however does not make $a \perp\!\!\!\perp b$ given $c \cup d$ as they have head-to-head path at c. Another possible graph can be figure 12(c) but in this case $c \perp\!\!\!\perp d$ given $a \cup b$ is violated.



(a) Ancestral Graph (b) Moralise and disorient (c) Delete givens

Figure 9: Model 2

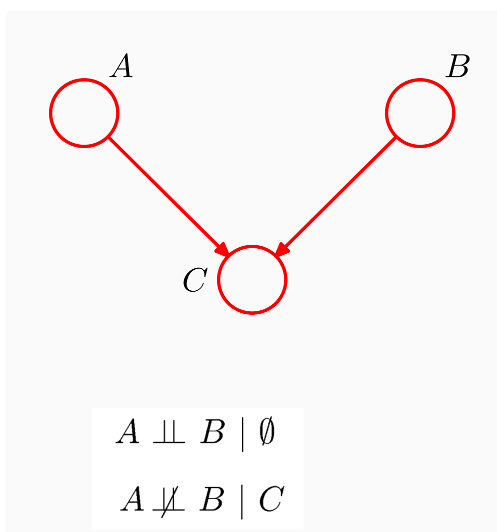


Figure 10: Directed.

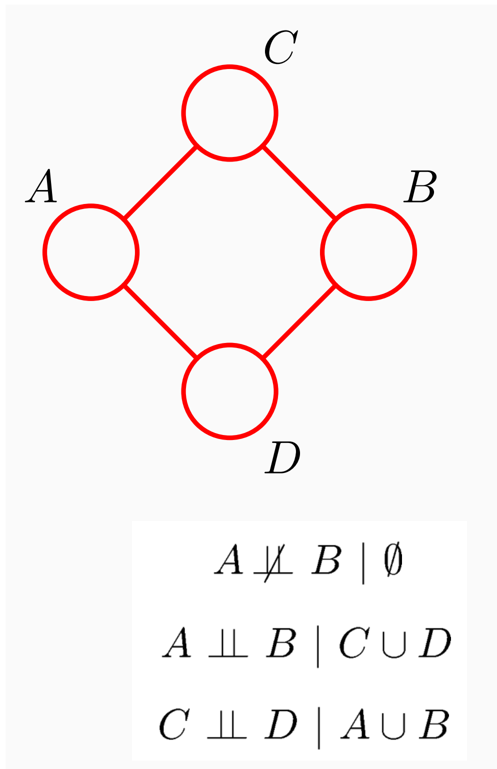
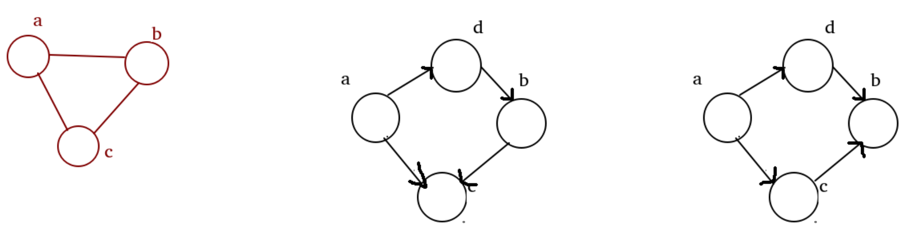


Figure 11: Undirected.



(a) (b) (c)

Figure 12: Ques 10