CS 475 Machine Learning: Homework 5
Graphical Models

Due: Tuesday November 24, 2015, 11:59pm

100 Points Total Version 1.0

Li-Yi Lin / llin34@jhu.edu

1 Analytical (40 points)

1. (10 points) Consider the Bayesian Network given in Figure 1(a). Are the sets **A** and **B** d-separated given set **C** for each of the following definitions of **A**, **B** and **C**? Justify each answer.

a. $\mathbf{A} = \{x_1\}$, $\mathbf{B} = \{x_9\}$, $\mathbf{C} = \{x_5, x_{14}\}$
Ans:
In Figure 3(a), there are two paths from x_9 to x_1. The path that passes through x_5 is blocked because x_5 is observed and the directions to x_5 are "tail to tail". Another path that passes though x_{14} is also blocked because x_14 is observed and the directions to x_14 are "tail to head". Therefore, x_1 and x_9 are conditional independent given $\{x_5, x_{14}\}$.

In Figure 3(b), the only two paths that can go from x_9 to x_1 are blocked by $\{x_5, x_{14}\}$. Thus, x_1 and x_9 are conditional independent given $\{x_5, x_{14}\}$.

b. $\mathbf{A} = \{x_{11}\}$, $\mathbf{B} = \{x_{13}\}$, $\mathbf{C} = \{x_1, x_{15}\}$
Ans:
In Figure 3(a), The path from x_{11} to x_{13} must pass node x_{15}. Since x_{15} is observed and the directions to x_{15} is "head to head", the path is unblocked. So x_{11} and x_{15} are not conditional indepnedent given $\{x_1, x_{15}\}$.

In Figure 3(b), since the path from x_{11} to x_{13} must go through x_{15} and x_{15} is observed, the path is blocked. Thus, x_{11} and x_{13} are conditional independent given x_{15} is observed.

c. $\mathbf{A} = \{x_4\}$, $\mathbf{B} = \{x_5\}$, $\mathbf{C} = \{x_{10}, x_{16}\}$
Ans:
In Figure 3(a), the path from x_4 to x_5 must pass through x_15. Since x_{15} is not observed and the directions to x_{15} are "head to head", the path is blocked. Therefore, x_4 and x_5 are conditional independent given $\{x_{10}, x_{16}\}$.

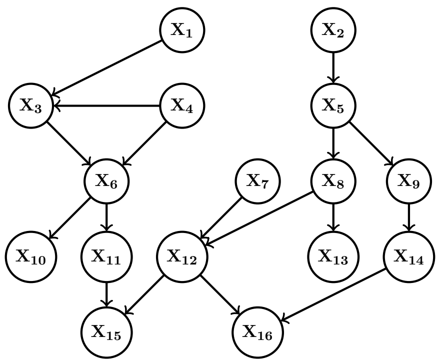
In Figure 3(b), since the paths from x_4 to x_5 are not blocked. Thus, x_4 and x_5 are not conditional independent given $\{x_{10}, x_{16}\}$.

d. $\mathbf{A} = \{x_3, x_4\}$, $\mathbf{B} = \{x_{13}, x_9\}$, $\mathbf{C} = \{x_{10}, x_{15}, x_{16}\}$
Ans:

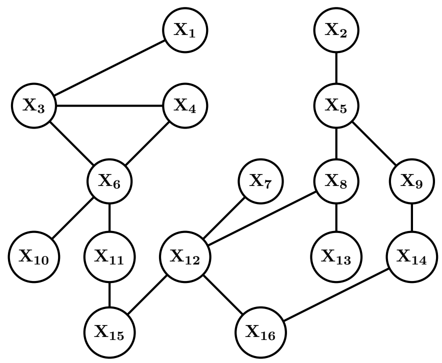
In Figure 3(a), the path from $\{x_3, x_4\}$ to $\{x_{13}, x_9\}$ must go through x_{15}. Since x_{15} is observed and the directions to x_{15} is "head to head", the path is not blocked. Thus, $\{x_3, x_4\}$ and $\{x_{13}, x_9\}$ are not conditional independent given $\{x_{10}, x_{15}, x_{16}\}$.

In Figure 3(b), since the path from $\{x_3, x_4\}$ to $\{x_{13}, x_9\}$ is blocked by x_{15}. Thus, $\{x_3, x_4\}$ and $\{x_{13}, x_9\}$ are conditional independent given $\{x_{10}, x_{15}, x_{16}\}$.

Now consider a Markov Random Field in Figure 3(b), which has the same structure as the previous Bayesian network. Re-answer each of the above questions with justifications for your answers.



(a) Directed Graphical Model



(b) Undirected Graphical Model

Figure 1: Two graphs are the same. However since (a) is directed and (b) is undirected the two graphs have different a conditional independence interpretation.

2. (10 points) Let $X = (X_1, ..., X_{16})^T$ be a random vector with distribution given by the graphical model in Figure 3(a). Consider variables $X_i \in \{X_6, X_8, X_{12}\}$. For each X_i, what is the minimal subset of the variables, $A \subset \mathcal{X} - \{X_i\}$, such that X_i is independent of the rest of the variables $(\mathcal{X} - (A \cup \{X_i\}))$ given A? Justify your answer. Answer the same questions for Figure 3(b).

Ans:

For x_6 in Figure 3(a), if x_3 and x_4 is blocked, x_1 and x_6 are conditional independent. If x_{11} is blocked, then x_6 is conditional independent with all the node in the right part. x_{10} also has to be blocked since it directly connects to x_6. So the subset for x_6 is $\{x_3, x_4, x_{10}, x_{11}\}$. For x_8 in Figure 3(a), if x_5 is blocked, x_2 and x_9 are conditional independent with x_8. x_{16} and x_{12} also need to be blocked. Since the direction to x_{12} is "head to head", x_7 also need to be blocked. x_{13} also needs to be blocked since it connects to x_8 directly. So the subset for x_8 is $\{x_5, x_7, x_{12}, x_{13}\}$. For x_{12} in Figure 3(a), x_7 and x_8 need to be blocked since they connect to x_{12} directly. Since the directions to x_{15} and x_{16} are both "head to head", x_{15} and x_{16} need not to be blocked. So the subset of X_{12} is $\{x_7, x_8, x_{11}, x_{14}, x_{15}, x_{16}\}$.

For x_6 in Figure 3(b), the minimal subset will be all the variables that are direct neighbors of x_6. So the subset of x_6 is $\{x_3, x_4, x_{10}, x_{11}\}$. For x_8, the subset is $\{x_5, x_9, x_{12}, x_{13}\}$. For x_{12}, the subset is $\{x_7, x_8, x_{15}, x_{16}\}$.

4. (10 points) The notation $(A \perp B|C)$ means A and B are conditionally independent given C. Prove or disprove (by providing a counter-example) each of the following properties of independence:

(a) $(X \perp Y, W|Z)$ implies $(X \perp Y|Z)$.

Ans:

We know $(X \perp Y, W|Z)$, so we can write

$$P(X, Y, W|Z) = P(X|Z)P(Y, W|Z)$$

Then we sum over W to eliminate W in the equation. So we get

$$\sum_W P(X, Y, W|Z) = \sum_W P(X|Z)P(Y, W|Z)$$
$$P(X, Y|Z) = P(X|Z)\sum_W P(Y, W|Z)$$
$$P(X, Y|Z) = P(X|Z)P(Y|Z)$$

Then, we get the result that $(X \perp Y|Z)$. Thus, we have proved that $(X \perp Y, W|Z)$ implies $(X \perp Y|Z)$.

(b) $(X \perp Y|Z)$ and $(X \perp W|Y, Z)$ imply $(X \perp Y, W|Z)$.

Ans:

We first apply Bayes rule and get following equation:

$$P(X, Y, W|Z) = P(X|Y, W, Z)P(Y, W|Z)$$

Since $(X \perp W|Y, Z)$, we know $P(X|W, Y, Z) = P(X|Y, Z)$. Then we have

$$P(X, Y, W|Z) = P(X|Y, Z)P(Y, W|Z)$$

Since $(X \perp Y|Z)$, we know $P(X|Y, Z) = P(X|Z)$. Then we have

$$P(X, Y, W|Z) = P(X|Z)P(Y, W|Z)$$

So we get the result that $(X \perp Y, W|Z)$. Hence, we have proved that $(X \perp Y|Z)$ and $(X \perp W|Y, Z)$ imply $(X \perp Y, W|Z)$.

(c) $(X \perp Y|Z)$ and $(X \perp W|Z)$ imply $(X \perp Y, W|Z)$.

Ans:

Given Z is the XOR space. The relationship between X, Y and W is shown below:

X	Y	W
0	0	0
0	1	1
1	0	1
1	1	0

For showing the independence between X and Y, we need to prove that $P(X, Y|Z) = P(X|Z)P(Y|Z)$. We first prove $P(X = 0, Y = 0|Z) = P(X = 0|Z)P(Y = 0|Z)$. We know that $P(X = 0, Y = 0|Z) = 1/4$, $P(X = 0|Z) = 1/2$ and $P(Y = 0|Z) = 1/2$, so $P(X = 0, Y = 0|Z) = P(X = 0|Z)P(Y = 0|Z) = 1/4$. We can use the same way to prove all the combinations of X and Y that $P(X, Y|Z) = P(X|Z)P(Y|Z)$. Thus, X and Y are independent given Z. We can also apply the same proof for X and W and prove that X and W are independent given Z. But $P(X = 0, Y = 0, W = 0|Z) \neq P(X = 0|Z)P(Y = 0, W = 0|Z)$ because $P(X = 0, Y = 0, W = 0|Z) = 1/4$ is not equal to $P(X = 0|Z)P(Y = 0, W = 0|Z) = 1/2 \times 1/4 = 1/8$. Therefore, the given expression is wrong.

5. (10 points) A Markov Random Field usually cannot help us with the factorization of the distribution function. However, some MRFs can be converted to Bayesian Networks. For example, consider the graph structure in Figure 2. From this graph, we know that X_2

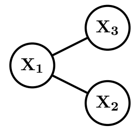


Figure 2: The Original Undirected Graph

and X_3 are conditionally independent given X_1. We can draw the corresponding directed graph as Figure 3. This suggests the following factorization of the joint probability:

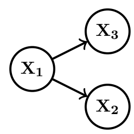


Figure 3: The Converted Directed Graph

$$P(X_1, X_2, X_3) = P(X_3|X_1)P(X_2|X_1)P(X_1)$$

Now consider the following graphical model in Figure 4.

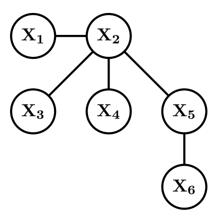


Figure 4: An Undirected Graph

As before, we can read the conditional independence relations from the graph.

(a) Following the example above, show a factorization of the joint distribution.
Ans:
$P(X_1, X_2, X_3, X_4, X_5, X_6) = P(X_1|X_2)P(X_3|X_2)P(X_4|X_2)P(X_6|X_5)P(X_5|X_2)P(X_2)$

(b) Is this factorization unique, meaning, could you have written other factorizations that correspond this model?
Ans:
This factorization is not unique. There still are other factorizations for this graph.

(c) If the factorization is unique, show why it is unique. If it is not unique, provide an alternate factorization.
Ans:
$P(X_1, X_2, X_3, X_4, X_5, X_6) = P(X_1|X_2)P(X_3|X_2)P(X_4|X_2)P(X_6|X_5)P(X_2|X_5)P(X_5)$