CS 475 Machine Learning: Homework 5
Graphical Models

Due: Friday Nov 18, 2016, 11:59pm

100 Points Total Version 1.0

Jin Yong Shin
JHED : Jshin44

1 Analytical (40 points)

1. (12 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.

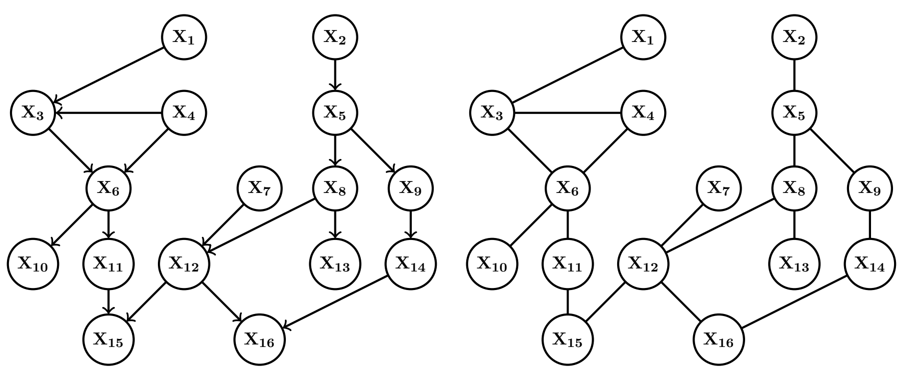
a. $\mathbf{A} = \{x_1\}$, $\mathbf{B} = \{x_9\}$, $\mathbf{C} = \{x_5, x_{14}\}$

b. $\mathbf{A} = \{x_{11}\}$, $\mathbf{B} = \{x_{13}\}$, $\mathbf{C} = \{x_1, x_{15}\}$

c. $\mathbf{A} = \{x_4\}$, $\mathbf{B} = \{x_5\}$, $\mathbf{C} = \{x_{10}, x_{16}\}$

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

Now consider a Markov Random Field in Figure 1(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.

Directed:

(a) : For both direct graph figure (a) and undirected graph figure (b), set A and set B are d-separated for the given set C. The reason is that for directed graph, it is clear that x_9 is sitting between x_5 and x_{14} which are only connected by one node and the only node connecting x_5 and x_{14}. So only one edge is going from x_5 to x_9 and one edge is going from x_9 and x_{14}. Therefore, all the paths that are going from set A which contains x_1 to set B which contains x_9 are blocked by set C which contains x_5 and x_{14}. So if we remove given nodes, there is no possible edge to connect x_1 and x_9.

(b) : For the directed graph, it's not d-separated in this case. Since we are given set C which contains x_{15}, the paths are not blocked. $x_1 5$ is descendent of the nodes in set A and set B, they are connected through x_{15}. Therefore, we can say they are not conditionally independent. Hence set A and set B are not d-separated.

(c) : Because node x_{15} is unobserved and is descendent of x_5 and x_4, we can say that x_4 and x_5 are blocked by x_{15} which is not given. Hence set A and set B are d-separated.

(d) : For the directed graph, it's not d-separated in this case. Since we are given set C which contains x_{15}, the paths are not blocked. $x_1 5$ is descendent of the nodes in set A and set B, they are connected through x_{15}. Since x_{15} is given which means it is observed, thus set A and set B are not conditionally independent. Thus not d-separated.

Undirected

(a) : Similar to directed graph version,this is also same for undirected graph as well. Basically, x_1 and x_9 do not share parent's node and also if we remove given nodes x_5 and x_{14}, then there is no route or path to connect x_1 and x_9. Therefore, set A and set B are independent which shows it's d-separation.

(b) : In the case of undirected graph, all the paths from set A to set B have to pass x_{15} which is in set C. Therefore, it is blocked. Hence set A and set B are conditionally independent (when we erase given nodes). Thus, d-separated

(c) : Here, for undirected graph, if we remove the given nodes x_{10} and x_{16}, the set A and set B are still connected. Therefore path from x_4 to x_5 are still valid. Therefore, set A and set B are not blocked. Which means not d-separated.

(d) : Similar to (b), if we remove given nodes in set C, we will eventually remove node x_{15}. Then there is no long valid way to connect set A and set B. Hence they are blocked, thus d-separated.

2. (15 points) Let $X = (X_1, ..., X_d^T)$ be a random vector, which follows a d-variate Gaussian distribution $N(0, \Sigma)$. Define the precision matrix as $\Omega = \Sigma^{-1}$. Please prove that $\Omega_{ij} = 0$ implies the conditional independence between variables X_i and X_j given the remaining variables, i.e., a sparsity pattern Ω is equivalent to an adjacency matrix of a Gaussian Markov Random Field.

ASNWER:

We have set X which has a random vector with $d - dimension$ and d-variate Gaussian distribution $N(0, \Sigma)$. Since it's d-variate finite Gaussian with given distribution, this will have nonsingular covariance matrix Σ, which will lead into precision matrix $\Omega = \Sigma^{-1}$. The element of covariance matrix Σ will tell us the covariance between the variable X_i and X_j. This element of the covariance can be written with sum of weights corresponding to the path on a graph between the node X_i and X_j.

$$\sigma_{X_i,X_j} = \sum_{P \in P_{X_i,X_j}} (-1)^n w_{p_1,p_2} w_{p_2,p_3} ... w_{p_{n-1},p_n} \frac{det(\Omega_{\setminus P})}{\Omega}$$

To find out precision matrix giving 0 will essentially imply the conditional independence between variables X_i and X_j when $\Sigma_{i,j} = 0$, we need to find $P(X_i, X_j | X_{-i,-j}) = P(X_i, | X_{-i,-j}) P(X_j | X_{-i,-j})$ where $X_{-i,-j}$ is the $X \setminus X_i, X_j$. We can rewrite the covariance elements in a way that

$$\sigma_{X_i,X_j} = \frac{1}{det(\Omega)} (w_{X_1} d(\Omega, X_1, X_i X_j) + ... + w_{X_n} d(\Omega, X_n, X_i X_j))$$

This is because we can expand $det(\Omega_{\setminus X_i, \setminus X_j})$ and also denote $d(\Omega, X_i, X_j)$ as the power of -1 times the determinant of the matrix with X_i and X_j. Since we have eliminated w_{X_i,X_i} from Ω by reproducing Ω, X_i, X_j, the total sum will drop the case where $w_{iX_i} = 0$. This will eventually lead into that when we compute σ_{xy} the determinant of the remaining matrix is zero thus not contributing to covariance. Therefore, this implies conditional independence.

To show this more precisely, let's prove more formally. If $\Omega_{i,j} = \Omega_{j,i} = 0$, then X_i and X_j are conditionally independent given set $X_{-i,-j}$. Let's denote $-i, -j$ which is the set excluding i, j with just r. So set X excluding i, j will be noted as X_r.

Now, let's calculate

$$P(X_{i,j} | X_r) = N(\mu_{i,j} | \Sigma_{ij,ji|r})$$

We know that $\Sigma_{ij,ji|r} = \Omega_{ij,ji}^{-1}$. Since we know how to compute the covariance, we can also get partial correlation coefficient of X_i and X_j.

$$\rho_{i,j|r} = \frac{\Sigma_{ij,ji|r}(i,j)}{\sqrt{\Sigma_{ij,ji|r}(i,i) * \Sigma_{ij,ji|r}(j,j)}} = \frac{-\lambda_{ij}}{\sqrt{\lambda_{ii}\lambda_{jj}}}$$

As we can see here, when $\rho_{i,j|r} = 0$ implies $\lambda_{i,j} = 0$. Therefore uncorrelated implies independence in here. This goes into precision matrix which represent conditional independence with zero values.

ANOTHER VERSION: Finally, my last version of proving conditional probability will be using conditional expectation. If X_i and X_j are conditional independent, then it will have

$$E(X_i | X_j; \theta) = E(X_i | \theta)$$

which is from the conditional independency of

$$P(X_i|X_j;\theta) = P(X_i|\theta)$$

We have Gaussian distribution of $\theta = (0, \Sigma)$. Then we can now rewrite the expectation with

$$\begin{aligned} E(X_i|X_j;\theta) &= E(X_i|\theta) + \Sigma_{ij}\Sigma_{jj}^{-}1(E(X_j|\theta) - X_j) \\ &= E(X_i|\theta) + \Omega_{ii}^{-}1\Sigma_{ij}(E(X_j|\theta) - X_j) \\ &= E(X_i|\theta) if(\Omega_{ij} = 0) \end{aligned} \tag{1}$$

Therefore, we can prove conditional independency by giving $\Omega_{ij} = 0$. Thus we can conclude that $\Omega_{ij} = 0$ correctly implies conditional independency

3. (13 points) The probability density function of most Markov Random Fields cannot be factorized as the product of a few conditional probabilities. This question explores some MRFs which can be factorized in this way.

Consider the graph structure in Figure 2. From this graph, we know that X_2 and X_3

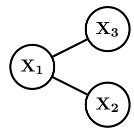


Figure 2: The Original Undirected Graph

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:

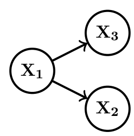


Figure 3: The Converted Directed Graph

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

Now consider the following graphical model in Figure 6.

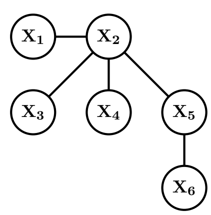


Figure 4: An Undirected Graph

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

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

(b) Is this factorization unique, meaning, could you have written other factorizations that correspond this model?

(c) If the factorization is unique, explain why it is unique. If it is not unique, provide an alternate factorization.

ANSWER:

a. If we change the undirected graph into directed graph such that x_2 node is the center and all the edges are directed out to other nodes.

Then we will have the following joint distribution such that

$$P(X_1, X_2, X_3, X_4, X_5, X_6) = P(X_1|X_2)P(X_3|X_2)P(X_4|X_2)P(X_5|X_2)P(X_6|X_5)P(X_2)$$

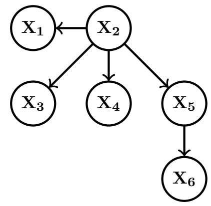


Figure 5: An Directed Graph for CounterExample

b. This factorization is not unique. The reason for the not unique factorization is given in part (c) with the counter example. Therefore, this could have written other factorization that correspond to this model.

c. If we change the direction of the directed graph, then we can have another factorization corresponding to this model. Following is the counter example: Therefore, the

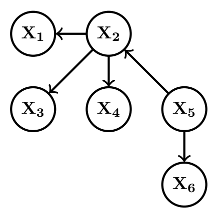


Figure 6: An Directed Graph for CounterExample

factorization for this model will be:

$$P(X_1, X_2, X_3, X_4, X_5, X_6) = P(X_1|X_2)P(X_3|X_2)P(X_4|X_2)P(X_6|X_5)P(X_2|X_5)P(X_5)$$