CS 475 Machine Learning: Homework 5
Graphical Models

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

100 Points Total Version 1.0

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.

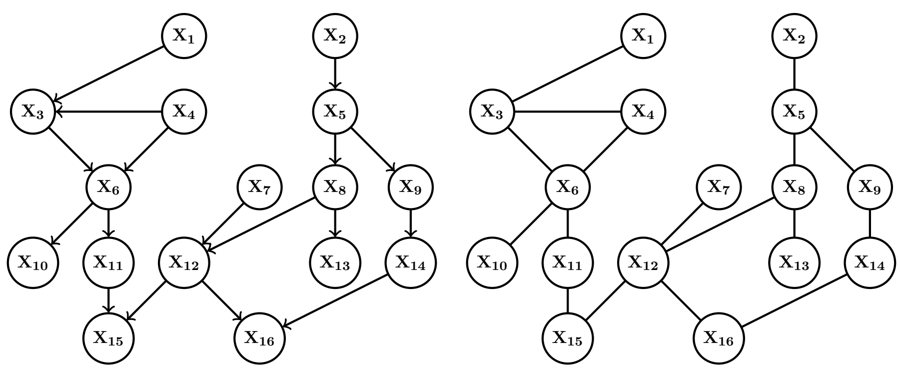
a. $\mathbf{A} = \{x_1\}$, $\mathbf{B} = \{x_9\}$, $\mathbf{C} = \{x_5, x_{14}\}$

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) Sets A and B are d-separated because x_5 is observed which blocks off x_9 from the top because it is a tail to tail path, and x_{14} is observed which blocks off x_9 from the bottom because it is a head to tail connection.

(b) Sets A and B are not d-separated because x_{15} is observed and it is a head to head connection, so it remains unblocked, so the path between x_{11} and x_{13} is unblocked.

(c) Sets A and B are d-separated because x_{15} is unobserved, so it blocks the path between x_4 and x_5.

(d) Sets A and be are not d-separated because x_{15} is observed, and x_{16} is observed so the head to head connection at x_{16} is unblocked and the head to head connection at x_{15} is also unblocked.

Undirected

(a) It is d-separated because all paths from x_1 to x_9 must pass through set C.

(b) It is d-separated because all paths from set A to set B pass through x_{15}

(c) Not d-separated because not all paths from set A to set B pass through set C

(d) It is d-separated because all paths from set A to set B have to pass through x_{15}

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.

My Answer:

Using the hint from piazza, we have to prove that:

$P(X_i, X_j|X_{-i,-j}) = P(X_i|X_{-i,-j})P(X_j|X_{-i,-j})$

And using the rules of conditional probability we can rearrange these probabilities such that the equation becomes:

$$\frac{P(X_i, X_j, X_{-i,-j})}{P(X_{-i,-j})} = \frac{P(X_i, X_{-i,-j})}{P(X_{-i,-j})}\frac{P(X_j, X_{-i,-j})}{P(X_{-i,-j})}$$

$$P(X_i, X_j, X_{-i,-j}) = P(X_i, X_{-i,-j})\frac{P(X_j, X_{-i,-j})}{P(X_{-i,-j})}$$

Now looking at the the multivariate distribution, we can see that the above equation is only true if:

$$exp(X^T\Omega X) = \frac{exp(X_i^T\Omega X_i)exp(X_j^T\Omega X_j)}{exp(X_{-i,-j}^T\Omega X_{-i,-j})}$$

which in terms of sums is:

$$X^T\Omega X = X_i^T\Omega X_i + X_j^T\Omega X_j - X_{-i,-j}^T\Omega X_{-i,-j}$$

Looking at the left hand side, we have:

$$X^T\Omega X = \sum_{p,q|p,q\neq j} X_p^T\Omega_{pq}X_q + \sum_{p,q|p,q\neq i} X_p^T\Omega_{pq}X_q + \sum_{p,q|p,q\in\{i,j\}} X_p^T\Omega_{pq}X_q - \sum_{p,q|p,q\notin\{i,j\}} X_p^T\Omega_{pq}X_q$$

The third summation becomes 0 because Ω_{ij} we know to be 0. So the LHS becomes:

$$\sum_{p,q|p,q\neq j} X_p^T\Omega_{pq}X_q + \sum_{p,q|p,q\neq i} X_p^T\Omega_{pq}X_q - \sum_{p,q|p,q\notin\{i,j\}} X_p^T\Omega_{pq}X_q$$

Now if we exponentiate this, it becomes:

$$\frac{exp(X_i^T\Omega X_i)exp(X_j^T\Omega X_j)}{exp(X_{-i,-j}^T\Omega X_{-i,-j})}$$

which means the LHS equals RHS and our proof is complete.

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

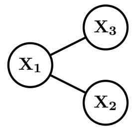


Figure 2: The Original Undirected Graph

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:

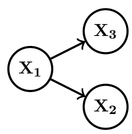


Figure 3: The Converted Directed Graph

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

Now consider the following graphical model in Figure 4.

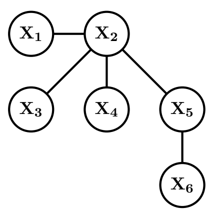


Figure 4: An Undirected Graph

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.

My Answer:

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

(b) No its not unique.

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