10-708 Probabilistic Graphical Models

Homework 1

Due Feb 10, 7:00 PM

Rules:

1. Homework is due on the due date at 7:00 PM. The homework should be submitted via Gradescope. Solution to each problem should start on a *new page* and marked appropriately on Gradescope. For policy on late submission, please see course website.
2. We recommend that you typeset your homework using appropriate software such as LaTeX. If you are writing, please make sure your homework is cleanly written up and legible. The TAs will not invest undue effort to decrypt bad handwriting.
3. **Code submission:** for programming questions, you must submit the complete source code of your implementation also via Gradescope. Remember to include a small README file and a script that would help us execute your code.
4. **Collaboration:** You are allowed to discuss the homework, but you should write up your own solution and code. Please indicate anyone you collaborated with in your submission.

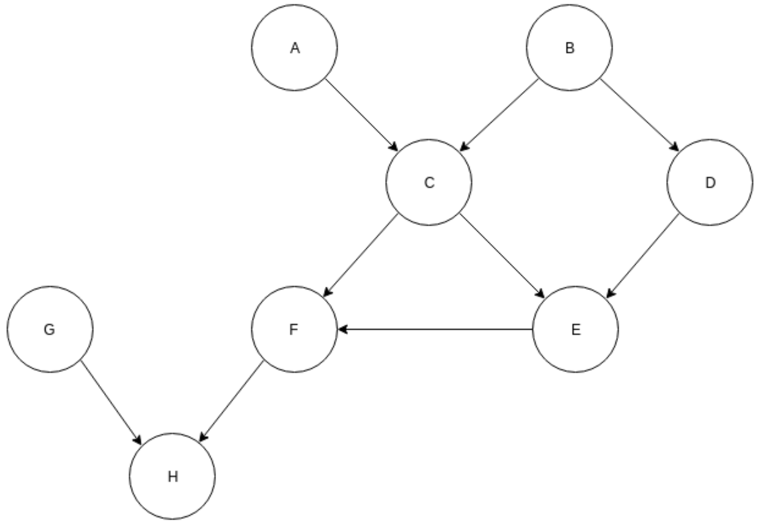

Figure 1: Figure for Problem 1.2

1 Bayesian Network [30 pts] (Haohan)

1.1 I-map (10 pts)

Let $\mathbb{G}$ be a Bayesian Network structure over a set of random variables $\mathbb{X}$, and let $\mathbb{P}$ be a joint distribution over the same space. If $\mathbb{G}$ is an I-map for $\mathbb{P}$, show that $\mathbb{P}$ factorizes according to $\mathbb{G}$.

1.2 D-seperation (10 pts)

According to Figure 1, determine whether the following claims to be True or False and justify your answer.

- $B \perp G \quad | \quad A$
- $C \perp D \quad | \quad F$
- $C \perp D \quad | \quad A$
- $H \perp B \quad | \quad C, F$
- There is at least one node in this BN, that all the other nodes are in its Markov Blanket.

1.3 I-Equivalence (10 pts)

I-equivalence of two graphs means that any distribution P that can be factorized over one of these graphs can be factorized over other. A relevant concept to describe I-equivalence is called **skeleton**. Formally, The **skeleton** of a Bayesian network graph G over X is an undirected graph over X that contains an edge $\{X, Y\}$ for every edge (X, Y) in G. Informally, just remove all the arrows of a Bayesian network, you will get the **skeleton** of it.

For two graphs G_1 and G_2, briefly justify the following arguments:

- G_1 and G_2 has the same skeleton is a necessary, but not sufficient condition for G_1 and G_2 to be I-equivalent.
- G_1 and G_2 has the same skeleton and same v-structures, is a sufficient, but not necessary condition for G_1 and G_2 to be I-equivalent.

amsmath amssymb epsfig graphicx

10-702:Statistical Machine Learning 10-702, Spring 2017

Homework 1

Lecturer: Ryan Tibshirani, Larry Wasserman *Name: Yuan Liu(yuanl4)*

2 Mathematical Statistics warm-up

Bayesian Network Solution

1.1 I-map

Solution: (Koller and Friedman Textbook Page 62)
Let $X_1, ..., X_n$ is a topological ordering of the variables in $\mathbb{X}$ relative to $\mathbb{G}$. We can first use the chain rule to represent $\mathbb{P}$:

$$P(X_1, ..., X_n) = \prod_{i=1}^{n} P(X_i|X_1, ..., X_{i-1})$$

Now, consider one of the factors $P(X_i|X_1, ..., X_{i-1})$. Because $\mathbb{G}$ is a I-map for P, we have

$$(X_i \perp NonDescendants_{x_i}|Pa_{x_i}^{\mathbb{G}}) \in I(\mathbb{P})$$

By the topological ordering assumption, all of X_i's parents are in the set $X_1, ..., X_{i-1}$. Furthermore, none of X_i's descendants can possibly be in the set. Hence,

$$\{X_1, ..., X_{i-1}\} = Pa_{X_i} \cup Z$$

where $Z \subset NonDescendants_{X_i}$. In addition to it, we already know $(X_i \perp NonDescendants_{x_i}|Pa_{x_i}^{\mathbb{G}}) \in I(\mathbb{P})$, so it follows that $(X_i \perp Z|Pa_{x_i})$. Hence we have that

$$P(X_i|X_1, ..., X_{i-1}) = P(X_i|Pa_{X_i})$$

Applying this transformation to all of the factors in the chain rule decomposition, the result follows.

1.2 D-separation

- $B \perp G|A$: True. Since H is unobserved, and $G \to H \leftarrow F$ is a v-structure, influence cannot flow through H.
- $C \perp D|F$: False. Influence can flow along the path $C \leftarrow B \to D$.
- $C \perp D|A$: False. Influence can flow along the path $C \leftarrow B \to D$.
- $H \perp B|C, F$: True. Since F is observed and $H \leftarrow F \leftarrow C$ or $H \leftarrow F \leftarrow E$ are neither v-structure, then H can not flow through F.

1.3 I-Equivalence

- G_1 and G_2 has the same skeleton is a necessary, but not sufficient condition for G_1 and G_2 to be I-equivalent.

  Not sufficient part proof: (http://www.wisdom.weizmann.ac.il/pgm/Ex/ps3_sol.pdf)The v-structure and common parent structure have the same skeleton, but they are not I-equivalent.

  Necessary part proof: Proof by contradiction. Assume G_1 and G_2 are I-equivalent, but they have different skeleton. In this situation, we can find a trail in one network that does not exist in the other. Assume this trail is $X_1 \rightleftharpoons X_2 \rightleftharpoons ... \rightleftharpoons X_i$ in G_1. Given all the v-structures in this trail, like $A_k \rightarrow B_k \leftarrow C_k$, let all the $\{B_k\}_k$ observed, and left others unobserved. Then $X_1 \perp X_2 | \{B_k\}_k$ in G_2, but $X_1 \not\perp X_2 | \{B_k\}_k$ in G_1

- G_1 and G_2 has the same skeleton and same v-structures, is a sufficient, but not necessary condition for G_1 and G_2 to be I-equivalent.

  Sufficient part proof: First we assume that $(\mathbb{X} \perp \mathbb{Y} | \mathbb{Z}) \in I(G_1)$ and we show that $(\mathbb{X} \perp \mathbb{Y} | \mathbb{Z}) \in I(G_2)$. If two graphs have the same skeleton, then they have the same trails. Lets look on the trails that between $X \in \mathbb{X}$ and $Y \in \mathbb{Y}$ that given $\mathbb{Z}$ is inactive in G_1, and then show this trail in G_2 is in active too. Consider two cases:

  i The trail in G_1 is inactive because some of the nodes on the trail that are not in v-structure are observed in $\mathbb{Z}$. Then clearly these nodes also blocks the trail in G_2

  ii Otherwise, all nodes on the trail that are not in a v-structure are not observed (not in $\mathbb{Z}$), but then for some v-structure V_{i-1}, V_i, V_{i+1} on the trail, none of the of V_i are observed. That is for every node V such that there is a directed path in G_1 from V_i to V, all the nodes on the path are not observed. Consider such a directed path from such a V_i to some V in G_1, then in G_2 this trail must also be directed the same, from V_i to V, because other wise it introduces a v-structure that is not in G_1(either on this path itself, or with respect to $V_{i-1} \rightarrow V_i$), and clearly all it's nodes are not observed too. There fore this path also inactivates the trail between $\mathbb{X}$ and $\mathbb{Y}$(given $\mathbb{Z}$) in G_2

  Thus, $(\mathbb{X} \perp \mathbb{Y} | \mathbb{Z}) \in I(G_1)$ implies that every trail between any pair $X \in \mathbb{X}$ and $Y \in \mathbb{Y}$ that given $\mathbb{Z}$ is inactive in G_1, and showed for each such a trail that it is inactive in G_2. Since the two graphs have the same trails, then $(\mathbb{X} \perp \mathbb{Y} | \mathbb{Z}) \in I(G_2)$. Hence $I(G_1) \subset I(G_2)$, and by symmetry we also get that $I(G_2) \subset I(G_1)$, therefore $I(G_1) = I(G_2)$

  Not necessary part proof: Consider complete graphs over a set of variables. Recall that a complete graph is one to which we cannot add additional arcs without causing cycles. Such graphs encode the empty set of conditional independence assertions. Thus, any two complete graphs are I-equivalent. Although they have the same skeleton, they invariably have different v-structures.