Chapter 3, The Bayesian Network Representation

Seminar on Probabilistic Graphical Models
October 27th 2015
Lecture notes and a bit aftermath, by Omer Tabach.

Chapter 3, The Bayesian Network Representation

Motivation

Goals for today:

Student Example

Introduction

Explaining the things on the graph:
"Chain rule for Bayesian networks"

Bayesian Network

Definition 3.5

Intuition:

3.2.2 Basic Independencies in Bayesian Networks

Intuition

Conditional Independencies, the Student example.

Conditional Independencies: the case for roots, and Intercausal reasoning

The window issue

Back to George

3.2.3 Graphs and Distributions

I-Maps introduction Independence Set Definition

I-Map Definition

I-Maps Example

I-Map example 1

I-Map example 2

Factorization: Bayesian Networks Chain Rule strikes again

I-Maps and Factorization

3.3 Independencies in Graphs

3.3.1 D-separation

X -> Y

X <-> Z <-> Y ?

Causal Trail: X -> Z -> Y

Evidential Trail: X <-> Z <-> Y

Common Cause X <-> Z -> Y

Common Effect X -> Z <-> Y

General Case

D-separation to I-Maps

3.2.2 Soundness and Completeness
Soundness
[Lack-of] Completeness
Completeness
D-Separation Conclusion
3.3.3 Linear algorithm for d-separation.
Run example over Figure 3.6

I-Equivalence
Definition
Motivation
Skeleton
V-Structure
Theorem 3.7, If G and K have the same skeleton and same set of v structures then
they are I-Equivalent
Definition 3.11: Immorality
The. 3.8: I-Equivalence

3.4 From Distribution to Graph,
Motivation
Naive Approach
3.4.1 Minimal I-Maps
Minimal I-Map is not unique
Minimal I-Map Fails
Minimal I-Map Algorithm
I-Map Algorithm over the Student example.
Minimal I-Maps Conclusion

3.4.2 Perfect Maps
Definition 3.14: Perfect Map
Perfect Map Fails: Regularity
Perfect Map Fails, Cases where BN is simply not appropriate
Perfect Map: Conclusion

Box 3.C: Skill: Knowledge Engineering
Picking Variables
Picking Structure
Picking Probabilities

Conclusion [terms]
Home work
Bravely Skipped
Motivation

Bayesian Networks, as Haim presented last week, serve as a compact representation for joint distribution over many variables.

Goals for today:

Key terms Covered

CPD,
Chain Rule,
Intercausal Reasoning.
I-Map
D-Separation
I-Equivalence
Minimal I-Map
Perfect Map

Student Example

Introduction
Explaining the things on the graph:

Directed Dependencies and CPDs. [Conditional Probability Distribution.]
Each graph node depends only on its parents.

"Chain rule for Bayesian networks"

Example of use of such network.
Query: 'What are the odds that an intelligent student took an easy course, got a bad grade, a bad letter but still had a high score on his SAT?'
Answer, via evaluation of the graph’s nodes in some topological ordering:

\[ P(I = 1, D = 0, G = 2, L = 0, S = 2) = P(I = 1) \cdot P(D = 0) \cdot P(G = 2 | I = 1, D = 0) \cdot P(L = 0 | G = 2) \cdot P(S = 2 | I = 1). \]
Bayesian Network

Definition 3.5

Bayesian network is a pair, BN=(G, P) where G is graph and P factorizes over G. P is specified as a set of CPDs associated with G's nodes. The distribution P is often annotated Pb.

Intuition:

The graph itself can be viewed in 2 different ways:

1. A data structure providing a skeleton for compact representation of the factorized joint distribution.
   - compact: Considering the $2^n$ alternative.
   - factorized: We look at random variables, disregarding their possible assignments (when building the graph.)
   - This will haunt us later.

2. A compact representation for a set of conditional independence assumptions about a distribution.

-This is the, perhaps, conceptually harder direction, which involves looking at the anti-edges. We will later see it is quite useful.

* Each node represents a distribution that depends only on its parents.
  It is called "Local Probability Model".

\[
\begin{align*}
X \perp Y \mid Z, \text{ or } \text{independence}\end{align*}
\]

3.2.2 Basic Independencies in Bayesian Networks

Intuition

In the previous section and example we viewed Bayesian networks as a data structure, encoding local, conditional probabilities.

We’re now going to look at Bayesian networks from a different perspective: the graph encodes a set of conditional independence assumptions.
\[ X \perp Y \iff P(X | Y) = P(X) \]

Conditional Independencies, the Student example.

1. \( L \perp I, D, S | G \)
   - Because the scores are independent of the exam results, the exam grades, and the test takers, alone do not affect the grades.

2. \( S \perp D, G, L | I \)
   - For the student, the scores do not depend on his grades, test scores, or exam grades.

3. \( G \perp S | D, I \)
   - Given the grades, the test scores do not depend on the exam grades.

4. \( I? \)
   - The causality between the exam grades and the test takers.

5. \( D \perp I \)
   - The causality between the scores and the exam grades.

\( X \rightarrow Y \rightarrow Z \)

\[ P(x_1) = 0.5 ; P(y_1 = x_1) = 0.25 ; P(z_1 | y_1) = 0.9 ; P(z_1 | y_0) = 0.5 \]

<table>
<thead>
<tr>
<th>( X )</th>
<th>( Y )</th>
<th>( Z )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 0 0</td>
<td>0.0625</td>
<td></td>
</tr>
<tr>
<td>0 0 1</td>
<td>0.0625</td>
<td></td>
</tr>
<tr>
<td>0 1 0</td>
<td>0.375</td>
<td></td>
</tr>
<tr>
<td>0 1 1</td>
<td>0.0375</td>
<td></td>
</tr>
<tr>
<td>1 0 0</td>
<td>0.1875</td>
<td></td>
</tr>
<tr>
<td>1 0 1</td>
<td>0.1875</td>
<td></td>
</tr>
<tr>
<td>1 1 0</td>
<td>0.125</td>
<td></td>
</tr>
<tr>
<td>1 1 1</td>
<td>0.1125</td>
<td></td>
</tr>
</tbody>
</table>

\[ P(y_1 | x_1) = 0.125 / 0.125 + 0.375 ; P(y_1 | x_1 z_1) = 0.1125/0.1125 + 0.1875. \]

3. \( G \perp D | I \)

Conditional Independencies: the case for roots, and Intercausal reasoning

\[ \text{Conditional Independencies: the case for roots, and Intercausal reasoning} \]

\[ \text{Conditional Independencies: the case for roots, and Intercausal reasoning} \]

\[ \text{Conditional Independencies: the case for roots, and Intercausal reasoning} \]

\[ \text{Conditional Independencies: the case for roots, and Intercausal reasoning} \]
The window issue

The window issue: any or both of the windows. If the window is open, the exposure to light is low.

\[ R_1 = 0.2 \; ; \; X_1 = 0.3 \; ; \; P_1 | r_1x_1 = 0.9, \; P_1 | r_1x_0 = 0.9, \; P_1 | r_0x_1 = 0.8, \; P_1 | r_0x_0 = 0. \]

Back to George

The question raised by the open window is whether there is a chance of a chance in the rain, and the probability of a chance in the rain is high.

Sometimes, windows without light in the house, I say:

 egy.

R1 = 0.2 ; X1 = 0.3 ; P1 | r1x1 = 0.9, P1 | r1x0 = 0.9, P1 | r0x1 = 0.8 P1 | r0x0 = 0.

3.2.3 Graphs and Distributions

Let G be a graph implying a set of independencies I(G). We say that G is an I-map for a set of independencies I if I(G) ⊆ I.

I-Maps introduction Independence Set Definition

I-Map: \( I \subseteq \mathcal{X} \times \mathcal{X} \times 2^\mathcal{X} \)

I-Map Definition
I(G) = X \perp \{Y \mid Y \text{ is not a descendant of } X \} \mid \text{Parents of } X

\text{CPD - הנבח ב Зна ההמצרה ה-}

I(G) \subseteq I(P).

I-Maps Example

יתכן, כי השבטים הזהים לא תלויה שאינן נבעות מהגרף. הנברך, יהו שבטים الفلות הזהים נבעות מהגרף. הנברך, ויהו שבטים الفلות הזהים נבעות מהגרף.

I-Map example 1

<table>
<thead>
<tr>
<th>X</th>
<th>Y</th>
<th>CPD</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>0.08%</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>0.32%</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>0.12%</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>0.48%</td>
</tr>
</tbody>
</table>

פִּילְּשָׁה X ב”ח Y: c=1 ו h 80% 1:4 ו C=0 X=1

כפי שנראה, מפות מחסנית את המפות הזהים. X=0, Y=1

I-Map example 2

<table>
<thead>
<tr>
<th>X</th>
<th>Y</th>
<th>CPD</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>0.40%</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>0.30%</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>0.20%</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>0.10%</td>
</tr>
</tbody>
</table>

פִּילְּשָׁה X ב”ח Y: c=1 ו h 3:7 ו C=0 X=1

נ西红ה פִּילְּשָׁה לשנים יידון 1:3 ו C=0 X=1

I(G\_phi) \not\subseteq I(P)
Factorization: Bayesian Networks Chain Rule strikes again

\[ P(X_1, \ldots, X_n) = \prod_{i=1}^{n} P(X_i | \text{Pa} x_i) \]

If G is an I-map for P, then P factorizes according to G. (Theorem 3.1.)
And if P factorizes according to G, then G is an I-map for P. (Th. 3.2.)

I-Maps and Factorization

\[ I_l(G) = X \perp \{ Y \mid Y \text{ is not a descendant of } X \} \mid \text{Parents of } X \]

3.3 Independencies in Graphs

3.3.1 D-separation

\[ X \perp Y \mid Z \]

 بشكل ר ngo;bokש,w_png762,ξω εφιάλης ιστος
Given that $Y$ depends on $X$, and $Z$ is a function of $Y$, it is possible that the answer to the question could be based on this relationship, D stands for Directed, or that there may be a causal relationship.

$X \rightarrow Y$

$X \leftrightarrow Z \leftrightarrow Y$?

Causal Trail: $X \rightarrow Z \rightarrow Y$

Evidential Trail: $X \leftarrow Z \leftarrow Y$

Common Cause $X \leftarrow Z \rightarrow Y$

Common Effect $X \rightarrow Z \leftarrow Y$

General Case

The general case is that for every $X$ that satisfies the equation $X \rightarrow Z \leftarrow Y$, there is a corresponding value of $Z$.

$N \rightarrow Z$?
D-separation to I-Maps

Definition 3.7
\[ I(G) = \{ (X \perp Y \mid Z) : d\text{-sep}(X;Y \mid Z) \} \]

3.2.2 Soundness and Completeness

Soundness

If X and Y are d-sep given some Z then we’re guaranteed they’re conditionally independent given Z over all possible distributions, which factorize over G.

[Lack-of] Completeness
Completeness

D-Separation Conclusion

3.3.3 Linear algorithm for d-separation.

Run example over Figure 3.6
Y -> X
X | Y ??

I-Equivalence

Definition

G is I-equivalent to K ⇔ I(G)=I(K).

Motivation

* From P to Equivalence class.

Skeleton

X -> Z <-> Y
X -> Z -> Y
Theorem 3.7, If G and K have the same skeleton and same set of v structures then they are I-Equivalent

Definition 3.11: Immorality

A v-structure $X \rightarrow Z \leftarrow Y$ is an Immorality if there is no direct edge between X and Y. If there is such an edge, we call it a ‘covering edge’.

The. 3.8: I-Equivalence

G and K are I-Equivalent iff they have the same skeleton and set of Immoralities.
3.4 From Distribution to Graph.

Motivation

Rearranging P in the distribution P to get G, we have P = G. The independence P depends on the variables P, κ, and P, κ, the values and the values of the variables P, κ, and P, κ. The independence P depends on the variables P, κ, and P, κ. The independence P depends on the variables P, κ, and P, κ.

To finally dilgano, this is a step.

Naive Approach

Consider both the graphs, but P does not change G. The independence G is unequivocal. The independence P is unequivocal. The independence P is unequivocal.

3.4.1 Minimal I-Maps

The minimal number of edges that are not minimal.

* A graph K is a minimal I-Map for a set of independencies I if it is an I-Map for I, and the removal of even a single edge from K renders it not an I-map.

Minimal I-Map is not unique

As a result of a minimal I-Map, it can be shown that the minimal I-Map is unique.

Minimal I-Map Fails

The minimal I-Map is not unique.
Minimal I-Map Algorithm

1. For X from i to n,
2. Find the minimal subset \( U \subseteq \{1...i-1\} \) s.t \( X \perp \{i...i-1 \setminus U\} | U \), and set them to be \( X_i \)'s parents.

Algorithm 3.2 Procedure to build a minimal I-map given an ordering

```
Procedure Build-Minimal-I-Map (X_1, \ldots, X_n) // an ordering of random variables in \( \mathcal{X} \)
    \( I \) // Set of independencies
)
1. Set \( G \) to an empty graph over \( \mathcal{X} \)
2. for \( i = 1, \ldots, n \)
3. \( U \leftarrow \{X_1, \ldots, X_{i-1}\} \) // \( U \) is the current candidate for parents of \( X_i \)
4. for \( U' \subseteq \{X_1, \ldots, X_{i-1}\} \)
5. if \( U' \subseteq U \) and \( (X_i \perp \{X_1, \ldots, X_{i-1}\} - U' | U') \in I \) then
6. \( U \leftarrow U' \)
7. // At this stage \( U \) is a minimal set satisfying \( (X_i \perp \{X_1, \ldots, X_{i-1}\} - U | U) \)
8. // Now set \( U \) to be the parents of \( X_i \)
9. for \( X_j \in U \)
10. Add \( X_j \rightarrow X_i \) to \( G \)
11. return \( G \)
```

I-Map Algorithm over the Student example.

כדעת נמשכים בשיטות בר肟ית יותר שהן יותר מינימאליות, כגון \( \) הסדרת הגוף \( \) המחברות לשתי התוכנות \( \).
כדעת נמשכים בשיטות בר昫ית \( \) המחברות \( \) לשתי התוכנות \( \).
\( \) ד"ר, \( G \) וא"ד \( I \) וא"ד \( S \).
Minimal I-Maps Conclusion

Let's consider the distribution $P_{\text{indep}}$, where Minimal I-Map definition is applied. Given a set of independencies $I$, we want to find a graph $G$ that precisely captures these independencies.

### 3.4.2 Perfect Maps

We aim to find a graph $G$ that precisely captures the independencies in a given distribution $P$.

**Definition 3.14: Perfect Map**

We say that $G$ is a P-Map for a set of independencies $I$ if we have that $I(G) = I$.

**Perfect Map Fails: Regularity**

Additionally, the independence structure should be preserved under different mappings. For example:

* $P_{x,y,z} = 1/12$ iff $x \oplus y \oplus z$; 1/6 otherwise.

The key here is to ensure that independencies are preserved under different mappings.

**Perfect Map Fails, Cases where BN is simply not appropriate**

Cases where BN is not appropriate include situations where the underlying independence structure cannot be captured accurately.

---

*Diagram* Figure 3.8: Three minimal I-maps for $P_{\text{indep}}$, induced by different orderings: (a) $D, I, S, G, L$; (b) $L, S, G, I, D$; (c) $L, D, S, I, G$. 
Figure 3.9 Network for the OneLetter example

Perfect Map: Conclusion

Box 3.C: Skill: Knowledge Engineering

Picking Variables
* Relevance Test

The relevance test should focus on the factors affecting the outcome, especially the behavior of the subjects. It is important to consider the impact of the additional variables. To conduct this test, we need to identify the critical factors, and analyze their impact on the outcomes.

Picking Structure

When picking structures, we should consider the critical factors and their impact on the outcomes. We should also consider the potential errors and their impact on the results.

* Butterfly Effect

The butterfly effect is a phenomenon where small changes can lead to significant outcomes. This is important to consider when picking structures.

Picking Probabilities

When picking probabilities, we need to consider the impact of the additional factors, especially the potential errors. We should also consider the potential errors and their impact on the results.
Conclusion [terms]

Key terms Covered
CPD,
Chain Rule,
Intercausal Reasoning.
I-Map
D-Separation
I-Equivalence
Minimal I-Map
Perfect Map

Home work

Bravely Skipped

Dummy/Naive Bayesian Network
3.4.3 Finding perfect maps.