ECE 6504: Advanced Topics in Machine Learning
Probabilistic Graphical Models and Large-Scale Learning

Topics:
- Bayes Nets: Representation/Semantics
  - d-separation, Local Markov Assumption
  - Markov Blanket
  - I-equivalence, (Minimal) I-Maps, P-Maps

Readings: KF 3.2, 3.4

Dhruv Batra
Virginia Tech
Recap of Last Time
A general Bayes net

- Set of random variables
- Directed acyclic graph
  - Encodes independence assumptions
- CPTs
  - Conditional Probability Tables

Joint distribution:

\[
P(X_1, \ldots, X_n) = \prod_{i=1}^{n} P(X_i \mid \text{Pa}_{X_i})
\]
Independencies in Problem

World, Data, reality:

True distribution $P$ contains independence assertions

BN:

Graph $G$ encodes local independence assumptions
Bayes Nets

• BN encode (conditional) independence assumptions.
  – I(G) = \{X \text{ indep of } Y \text{ given } Z\}

• Which ones?
• And how can we easily read them?
Local Structures

• What’s the smallest Bayes Net?
Local Structures

Indirect causal effect:

\[ X \rightarrow Z \rightarrow Y \]

Indirect evidential effect:

\[ X \leftarrow Z \leftarrow Y \]

Common cause:

\[ X \leftarrow Z \rightarrow Y \]

Common effect:

\[ X \rightarrow Z \rightarrow Y \]
Bayes Ball Rules

• Flow of information
  – on board
Plan for today

• Bayesian Networks: Semantics
  – d-separation
  – General (conditional) independence assumptions in a BN
  – Markov Blanket
  – (Minimal) I-map, P-map
Active trails formalized

• Let variables \( O \subseteq \{X_1, \ldots, X_n\} \) be observed

• A path \( X_1 \rightarrow X_2 \rightarrow \cdots \rightarrow X_k \) is an active trail if for each consecutive triplet:
  
  - \( X_{i-1} \rightarrow X_i \rightarrow X_{i+1}, \) and \( X_i \) is not observed \( (X_i \notin O) \)
  
  - \( X_{i-1} \leftarrow X_i \leftarrow X_{i+1}, \) and \( X_i \) is not observed \( (X_i \notin O) \)
  
  - \( X_{i-1} \leftarrow X_i \rightarrow X_{i+1}, \) and \( X_i \) is not observed \( (X_i \notin O) \)
  
  - \( X_{i-1} \rightarrow X_i \leftarrow X_{i+1}, \) and \( X_i \) is observed \( (X_i \in O), \) or one of its descendents is observed
An active trail – Example

When are $A$ and $H$ independent?
**d-Separation**

- **Definition**: Variables $X$ and $Y$ are d-separated given $Z$ if
  - no active trail between $X_i$ and $Y_j$
  when variables $Z \subseteq \{X_1, \ldots, X_n\}$ are observed

(C) Dhruv Batra
d-Separation

- So what if X and Y are d-separated given Z?
Theorem:
- If
  - P factorizes over G
  - $d$-sep$_G(X, Y \mid Z)$
- Then
  - $P \vdash (X \perp Y \mid Z)$

Corollary:
- $I(G) \subseteq I(P)$
  - All independence assertions read from G are correct!
More generally: Completeness of d-separation

- **Theorem: Completeness of d-separation**
  - For “almost all” distributions where $P$ factorizes over to $G$
  - we have that $I(G) = I(P)$
  - “almost all” distributions: except for a set of measure zero of CPTs
  - Means that if $X$ & $Y$ are not d-separated given $Z$, then $P \neg (X \perp Y | Z)$
A variable $X$ is independent of its non-descendants given its parents and only its parents

$$(X_i \perp \text{NonDescendants}_{X_i} \mid Pa_{X_i})$$
Markov Blanket

= Markov Blanket of variable $x_8$ – Parents, children and parents of children
A variable is conditionally independent of all others, given its Markov Blanket.
I-map

• Independency map

• Definition:
  – If $I(G) \subseteq I(P)$
  – $G$ is an I-map of $P$
Factorization + d-sep $\Rightarrow$ Independence

• Theorem:
  – If
    • $P$ factorizes over $G$
    • $d$-sep$_G(X, Y \mid Z)$
  – Then
    • $P \vdash (X \perp Y \mid Z)$

• Corollary:
  • $I(G) \subseteq I(P)$
  • $G$ is an I-map of $P$
  • All independence assertions read from $G$ are correct!
The BN Representation Theorem

If $G$ is an I-map of $P$

Obtain

$P$ factorizes to $G$

$P(X_1, \ldots, X_n) = \prod_{i=1}^{n} P\left(X_i \mid \text{Pa}_{X_i}\right)$

Important because:

Every

$P$ factorizes to $G$

Obtain

$G$ is an I-map of $P$

Important because:

Read independencies of $P$ from BN structure $G$

Homework 1!!!!

(C) Dhruv Batra

Slide Credit: Carlos Guestrin
I-Equivalence

• Two graphs $G_1$ and $G_2$ are **I-equivalent** if
  – $I(G_1) = I(G_2)$

• **Equivalence class** of BN structures
  – Mutually-exclusive and exhaustive partition of graphs
Minimal I-maps & P-maps

- Many possible I-maps
- Is there a “simplest” I-map?

- Yes, two directions
  - Minimal I-maps
  - P-maps
Minimal I-map

- $G$ is a **minimal I-map** for $P$ if
  - deleting any edges from $G$ makes it no longer an I-map
P-map

• Perfect map

• $G$ is a P-map for $P$ if
  - $I(P) = I(G)$

• Question: Does every distribution $P$ have P-map?
BN: Representation: What you need to know

• Bayesian networks
  – A compact representation for large probability distributions
  – Not an algorithm

• Representation
  – BNs represent (conditional) independence assumptions
  – BN structure = family of distributions
  – BN structure + CPTs = 1 single distribution
  – Concepts
    • Active Trails (flow of information); d-separation;
    • Local Markov Assumptions, Markov Blanket
    • I-map, P-map
    • BN Representation Theorem (I-map $\iff$ Factorization)
Main Issues in PGMs

• Representation
  – How do we store $P(X_1, X_2, \ldots, X_n)$
  – What does my model mean/imply/assume? (Semantics)

• Learning
  – How do we learn parameters and structure of $P(X_1, X_2, \ldots, X_n)$ from data?
  – What model is the right for my data?

• Inference
  – How do I answer questions/queries with my model? such as
  – Marginal Estimation: $P(X_5 | X_1, X_4)$
  – Most Probable Explanation: argmax $P(X_1, X_2, \ldots, X_n)$
# Learning Bayes nets

<table>
<thead>
<tr>
<th>Data structure</th>
<th>Known structure</th>
<th>Unknown structure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fully observable data</td>
<td>Very easy</td>
<td>Hard</td>
</tr>
<tr>
<td>Missing data</td>
<td>Somewhat easy (EM)</td>
<td>Very very hard</td>
</tr>
</tbody>
</table>

\[
\begin{align*}
X^{(1)} & \quad \ldots \quad X^{(m)} \\
\text{structure} & \quad + \\
\text{CPTs} & \quad P(X_i \mid Pa_{X_i})
\end{align*}
\]

(C) Dhruv Batra

Slide Credit: Carlos Guestrin