PGMs – An Odyssey

Book 1: Bayesian & Markov networks

Ben Teo
4/25/2022

Odyssey:

> 1: a long wandering or voyage usually marked by many changes of fortune
>
> 2: an intellectual or spiritual wandering or quest

Motivation:

Draw upon the extensive PGM literature to extend Phylogenetic Comparative Methods and software implementations.

Charting a (tentative) course:

Koller, Daphne, and Nir Friedman. *Probabilistic graphical models: principles and techniques*. MIT press, 2009

- Chapters 3, 4, 7, 9, 10, 11, 14

Bayesian networks (BN)

1. Distribution, P
2. Directed graph, $\mathcal{G}$: Represents a set of (conditional) independencies, $I(\mathcal{G})$.

Definition. Bayesian network, $\mathcal{B} = (\mathcal{G}, P)$ such that P satisfies $I(\mathcal{G})$. Also written as $I(\mathcal{G}) \subseteq I(P)$.

- If $I(\mathcal{G}) \subseteq I(P)$, we call $\mathcal{G}$ an *I-map* for P.

1(i). P can be factored as a product of *conditional probability distributions* (CPDs) via the chain rule. E.g.:

$$\begin{aligned} P(X_1, X_2, X_3) &= P(X_1 \mid X_2, X_3)P(X_2 \mid X_3)P(X_3) \\ &= P(X_2 \mid X_1, X_3)P(X_3 \mid X_1)P(X_1) \end{aligned}$$

2(i). P factorizes over $\mathcal{G}$ if

$$P(X_1, \ldots, X_n) = \prod_i P(X_i \mid \mathrm{Pa}^{\mathcal{G}}_{X_i})$$

- That is, P satisfies the *local independencies* of $\mathcal{G}$, $I_\ell(\mathcal{G})$.
- $I_\ell(\mathcal{G})$: $\text{node} \perp\!\!\!\perp \text{non-descendants} \mid \text{parents}$

Theorem. *$\mathcal{G}$ is an I-map for P $\Leftrightarrow$ P factorizes over $\mathcal{G}$.*

- Bayesian network, $\mathcal{B} = (\mathcal{G}, P)$ where P factorizes over $\mathcal{G}$.

So far $I(\mathcal{G}) = I_\ell(\mathcal{G})$. $\mathcal{B} = (\mathcal{G}, P)$ must satisfy $I_\ell(\mathcal{G}) \subseteq I(P)$.

- But what other (non-local) independencies does $I_\ell(\mathcal{G})$ imply? By definition, $I(\mathcal{G})$ should also include those.

3. D-separation: Define *active trails* between two sets of nodes given a conditioning set.

- Absence of an active trail indicates a *global independency*.
- Conditioning on Z "blocks" the following paths:
  - $X \to Z \to Y$
  - $X \leftarrow Z \leftarrow Y$
  - $X \leftarrow Z \to Y$
- $X \to Z \leftarrow Y$ is "active" if we condition on any $Z^* \in Z \cup \mathrm{Desc}_Z^{\mathcal{G}}$
  - A v-structure without a covering edge (e.g. $X \leftrightarrow Y$) is called an *immorality*.

So we define $I(\mathcal{G}) = \{(\mathbf{X} \perp\!\!\!\perp \mathbf{Y} \mid \mathbf{Z}) : \text{d-sep}_{\mathcal{G}}(\mathbf{X}; \mathbf{Y} \mid \mathbf{Z})\}$.

Theorem. *D-separation is sound, but not complete.*

- $\text{d-sep}_{\mathcal{G}}(\mathbf{X}; \mathbf{Y} \mid \mathbf{Z}) \notin I(\mathcal{G})$ does not imply $\mathbf{X} \not\perp\!\!\!\perp \mathbf{Y} \mid \mathbf{Z}$!

4. Minimal I-maps & P-maps

- $\mathcal{G}$ is a *minimal* I-map for P if no edge can be removed without $I(\mathcal{G}) \nsubseteq I(P)$.
  - E.g. construct such a $\mathcal{G}$ given a topological ordering of the X_is by selecting minimal $\mathrm{Pa}_{X_i}^{\mathcal{G}}$ s.
- $\mathcal{K}$ is a P-map for P if $I(\mathcal{K}) = I(P)$.

Not every P can be perfectly represented by some $\mathcal{G}$. E.g.

- $$P(x, y, z) = \begin{cases} 1/12, & x \oplus y \oplus z = \text{false} \\ 1/6, & x \oplus y \oplus z = \text{true} \end{cases}$$
  - X, Y, Z are pairwise independent ($\Rightarrow$ no edges), but not jointly independent ($\Rightarrow$ some edges) $-\!\!\times\!\!-$
- Finer-grained independencies (i.e. that hold for specific variable values)
- Symmetric independencies (e.g. $\diamond$ structure)

Markov networks (MN)

1. Gibbs distribution, $P_\Phi \propto \prod_i \phi_i(\mathbf{D}_i)$

- Φ denotes $\{\phi_1(\mathbf{D}_1), \ldots, \phi_k(\mathbf{D}_k)\}$

2. Undirected graph, $\mathcal{H}$

- $\mathbf{X}$ is *separated* from $\mathbf{Y}$ given $\mathbf{Z}$ if every path from $\mathbf{X}$ to $\mathbf{Y}$ passes through $\mathbf{Z}$.
- $I(\mathcal{H}) = \{(\mathbf{X} \perp\!\!\!\perp \mathbf{Y} \mid \mathbf{Z}) : \mathrm{sep}_{\mathcal{H}}(\mathbf{X}; \mathbf{Y} \mid \mathbf{Z})\}$

Definition. Markov network, $\mathcal{M} = (\mathcal{H}, P_\Phi)$ such that $I(\mathcal{H}) \subseteq I(P_\Phi)$.

2(i). P_Φ factorizes over $\mathcal{H}$ if each $\mathbf{D}_i$ is a clique of $\mathcal{H}$.

Theorem. *P_Φ factorizes over $\mathcal{H} \Rightarrow I(\mathcal{H}) \subseteq I(P_\Phi)$.*

Theorem. *Suppose $P_\Phi > 0$. P_Φ factorizes over $\mathcal{H} \Leftrightarrow I(\mathcal{H}) \subseteq I(P_\Phi)$.*

- Generally ($P_\Phi \geq 0$), factorization is sound but *not complete* for $I(\mathcal{H}) \subseteq I(P_\Phi)$.
- Considering only $P_\Phi > 0$, Markov network, $\mathcal{M} = (\mathcal{H}, P_\Phi)$ where P_Φ factorizes over $\mathcal{H}$.

For BNs, to check if $P \models I(\mathcal{G})$, it suffices to check $P \models I_\ell(\mathcal{G})$. Can we similarly find local criteria for $P \models I(\mathcal{H})$?

3. Pairwise & local independencies

- Pairwise: $I_p(\mathcal{H}) = \{(X \perp\!\!\!\perp Y \mid \mathcal{X} \setminus \{X, Y\}) : X - Y \notin \mathcal{H}\}$
  - Two non-adjacent nodes are independent conditional on the other nodes.
- Local: $I_\ell(\mathcal{H}) = \{(X \perp\!\!\!\perp \mathcal{X} \setminus \{X\} \setminus \mathrm{MB}_X^{\mathcal{H}} \mid \mathrm{MB}_X^{\mathcal{H}}) : X \in \mathcal{X}\}$
  - A node is independent of its non-neighbors conditional on its neighbors.

Theorem. *Suppose $P > 0$.*

$$P \models I_p(\mathcal{H}) \Leftrightarrow P \models I_\ell(\mathcal{H}) \Leftrightarrow P \models I(\mathcal{H})$$

- *Generally, $I_p(\mathcal{H}) \subseteq I_\ell(\mathcal{H}) \subseteq I(\mathcal{H})$.*
- *If $P > 0$, then $P \models I_p(\mathcal{H}) \Rightarrow P \models I(\mathcal{H})$.*

4. Representing BNs as MNs, and vice versa

- BN$\rightarrow$MN: Two views
  - Given $\mathcal{B} = (\mathcal{G}, P)$, $\mathcal{G}$ implies a factorization for P. Construct $\mathcal{H}$ such that each $\{X_i\} \cup \mathrm{Pa}_{X_i}^{\mathcal{G}}$ is a clique.
  - Construct $\mathcal{H}$ such that $I(\mathcal{H}) \subseteq I(\mathcal{G})$ using $\mathrm{MB}_{X_i}^{\mathcal{G}}$s identified from $\mathcal{G}$.

Theorem. *The moral graph $\mathcal{H} = \mathcal{M}[\mathcal{G}]$ is a minimal I-map for P, and is a P-map for P if $\mathcal{G}$ is moral.*

- BN$\leftarrow$MN

Theorem. *If $\mathcal{H}$ is nonchordal, $\nexists \mathcal{G}$ with $I(\mathcal{G}) = I(\mathcal{H})$. If $\mathcal{H}$ is chordal, such a $\mathcal{G}$ exists.*

Theorem. *A set of independencies I can be perfectly represented by both some $\mathcal{G}$ and some $\mathcal{H}$ iff $I = I(\mathcal{H})$ for some chordal $\mathcal{H}$.*

The existence proofs on the previous slide use *clique trees*.

5. **Definition.** Tree $\mathcal{T}$ is a clique tree for $\mathcal{H}$ if

- Each node is a clique in $\mathcal{H}$, and each maximal clique in $\mathcal{H}$ is a node in $\mathcal{T}$.
- Each edge $\mathbf{C}_i - \mathbf{C}_j$ in $\mathcal{T}$ is associated with the sepset $\mathbf{S}_{i,j} = \mathbf{C}_i \cap \mathbf{C}_j$ (intersection of neighbor cliques).
- Each sepset $\mathbf{S}_{i,j}$ separates $\mathbf{W}_{<(i,j)}$ and $\mathbf{W}_{<(j,i)}$ in $\mathcal{H}$.
  - $\mathbf{W}_{<(i,j)}$ is the union of variables in any clique (including $\mathbf{C}_i$) on the $\mathbf{C}_i$ side of the $\mathbf{C}_i - \mathbf{C}_j$ edge.

Theorem. *Every chordal $\mathcal{H}$ has a clique tree $\mathcal{T}$.*