Probabilistic Graphical Models

Tianyu Du

May 23, 2020

1 Graphical Representations

1.1 Factors

Definition 1.1. Let $X_1, X_2, \cdots, X_k$ be a set of random variables, then a **factor** ϕ is a mapping from values of these random variables to $\mathbb{R}$.

$$\phi : Val(X_1, X_2, \cdots, X_k) \to \mathbb{R} \tag{1}$$

The set of random variables $\{X_1, X_2, \cdots, X_k\}$ is defined as the **scope** of ϕ.

Definition 1.2. Let ϕ_1 and ϕ_2 be two factors with scopes $\{A, B\}$ and $\{B, C\}$. Then the **factor product** $\phi_1 \times \phi_2$ is a factor with scope $\{A, B, C\}$ defined as

$$\phi_1 \cdot \phi_2(a, b, c) = \phi_1(a, b) \cdot \phi_2(b, c) \tag{2}$$

Definition 1.3. Let ϕ be a factor with scope $\{A, B, C\}$, then **marginalizing** C **from** ϕ results in a factor ϕ' with scope $\{A, B\}$ defined as the following:

$$\phi'(a, b) = \sum_{c \in Val(C)} \phi(a, b, c) \tag{3}$$

Definition 1.4. The **factor reduction** operation restricts $\phi(A, B, C)$ to take only a specific value of $C = c$, and results in a factor ϕ' with scope $\{A, B\}$.

$$\phi'(a, b) = \phi(a, b, c) \tag{4}$$

1.2 Semantics and Factorization

Definition 1.5. A **Bayesian network** consists of (i) a directed acyclic graph (DAG) G whose nodes correspond to random variables $X_1, \cdots, X_n$ (ii) and a conditional probability distribution $P(X_i | Par_G(X_i))$ for each node X_i. The joint distribution is defined as the factorization

$$P(X_1, \cdots, X_n) = \prod_{i=1}^{n} P(X_i | \mathrm{Par}_G(X_i)) \tag{5}$$

Definition 1.6. Let G be a graph over $X_1, \cdots, X_n$, then the joint probability P **factorizes** over G if and only if

$$P(X_1, \cdots, X_n) = \prod_{i=1}^{n} P(X_i | \mathrm{Par}_G(X_i)) \tag{6}$$

1.3 Pass of Influences in Bayesian Networks

Definition 1.7. A path $X_1 - \cdots - X_k$ in Bayesian network G is **active** if there is no explaining-away structure $X_{i-1} \to X_i \leftarrow X_{i+1}$ in it.

Definition 1.8. Let $Z \subseteq V_G$ be a set of random variables in the Bayesian network, then a path $X_1 - \cdots - X_k$ in G is **active conditioned on** Z if

1. for all explaining-away structure $X_{i-1} \to X_i \leftarrow X_{i+1}$ in the path, X_i or some decedents of X_i are in Z,
2. and no other node in the path is in Z.

Definition 1.9. Let $X, Y, Z \subseteq V_G$, if there is no path from X to Y is active conditioned on Z, then X and Y are **d-separated** by Z in graph G denoted as $\text{d-sep}_G(X, Y|Z)$.

1.4 Independencies and Factorizations

Definition 1.10. Let X, Y, Z be random variables with distribution P, then $X \perp Y$ if and only if $P(X, Y) = P(X)P(Y)$, $X \perp Y|Z$ if and only if $P(X, Y|Z) = P(X|Z)P(Y|Z)$.

Proposition 1.1. Let X, Y, Z be random variables with distribution P, then $X \perp Y$ if and only if $P(X, Y)$ factorizes as the following

$$P(X, Y) \propto \phi_1(X)\phi_1(Y) \tag{7}$$

and $X \perp Y|Z$ if and only if $P(X, Y, Z)$ factorizes as

$$P(X, Y, Z) \propto \phi_1(X, Z)\phi_1(Y, Z) \tag{8}$$

Proof. Relation (7) follows the definition immediately. Suppose $X \perp Y|Z$, then

$$\begin{aligned}
& P(X, Y|Z) = P(X|Z)P(Y|Z) && (9)\\
\iff & P(X, Y, Z) = P(X|Z)P(Y|Z)P(Z) && (10)\\
& P(X, Y, Z) \propto P(X|Z)P(Z)P(Y|Z)P(Z) && (11)\\
& \qquad\qquad\;\; = P(X, Z)P(Y, Z) && (12)\\
& \qquad\qquad\;\; = \phi_1(X, Z)\phi_1(Y, Z) && (13)
\end{aligned}$$

■

Theorem 1.1 (Factorization $\Longrightarrow$ Independence)**.** If P factorizes over G, and $\text{d-sep}_G(X, Y|Z)$ then P satisfies $(X \perp Y|Z)$.

Theorem 1.2. For any random variable X_i in the Bayesian network, X_i is d-separated from all its non-descendants by $\text{Par}_G(X_i)$.

Corollary 1.1. If P factorizes over G, then in P, any variable is independent of its non-descendants given its parents.

Definition 1.11. Let $\mathcal{I}(G)$ denote the collection of independencies implicitly encoded by d-separations in graph G,

$$\mathcal{I}(G) := \{(X \perp Y|Z) : X, Y, Z \in V \;\; s.t. \;\; \text{d-sep}_G(X, Y|Z)\} \tag{14}$$

If a distribution P over V satisfies all independencies in $\mathcal{I}(G)$, then we say that G is an **I-map** (independency map) of P.

That is, the I-map of distribution P is a graphical representation of all (and probably more) independencies of P.

Example 1.1. Let P be a probability distribution and let G be an I-map for P. Let $\mathcal{I}(P)$ and $\mathcal{I}(G)$ denote sets of independencies in P and G. Suppose G is a I-map of P, then all independencies encoded in G are satisfied by P, therefore,

$$\mathcal{I}(G) \subseteq \mathcal{I}(P) \tag{15}$$

Example 1.2. The I-map can be used for two graphs as well. G_1 is a I-map of G_1 if $\mathcal{I}(G_1) \subseteq \mathcal{I}(G_2)$. That is, G_1 is an I-map of G_2 if it does not make independence assumptions that are not true in G_2.

Theorem 1.3 (Independence $\Longrightarrow$ Factorization)**.** If G is an I-map for P, that is, P adheres all independencies encoded in G, then P factorizes over G.

1.5 Template Models

Definition 1.12. A **template variable** $X(U_1, \cdots, U_k)$ is instantiated (duplicated) multiple times in a graph. **Template models** are languages that specify how ground variables (i.e., instantiations of template variables) inherit dependency model from template.

Notation 1.1. Let $X^{(t)}$ denote the variable at time $t\Delta$, where Δ is the time granularity in the discrete timeline. Let $X^{(t:t')} = \{X^{(t)},^{(t+1)}, \cdots, X^{(t')}\}$ denote the set of variables over a period of time.

Definition 1.13. A Bayesian network is said to satisfy the **Markov assumption** if

$$X^{(t+1)} \perp X^{(0:t-1)} | X^{(t)} \tag{16}$$

When Markov assumption holds, we may express the joint distribution of all X as

$$P(X^{(0:T)}) = P(X^{(0)}) \prod_{t=0}^{T-1} P(X^{(t+1)} | X^{(t)}) \tag{17}$$

Definition 1.14. A series of random variables $X^{(0)}, X^{(1)}, \cdots, X^{(T)}$ satisfies the **time invariance** assumption if there exists a template probability model $P(X'|X)$ such that for all t,

$$P(X^{(t+1)} | X^{(t)}) = P(X'|X) \tag{18}$$

Definition 1.15. A **2-time-slice Bayesian network** (2TNB) over $X_1, \cdots, X_n$ (n random variables for each time step) is specified as a Bayesian network fragment such that

- The nodes include $X'_1, \cdots, X'_n$ and a subset of $X_1, \cdots, X_n$,
- and only the nodes $X'_n, \cdots, X'_n$ have parents and a conditional probability distribution.

Further, the 2TBN defines a conditional distribution

$$P(X'|X) = \prod_{i=1}^{n} P(X'_i | \mathrm{Par}(X'_i)) \tag{19}$$

Definition 1.16. A **dynamic Bayesian network** (DNB) over $X_1, \cdots, X_n$ is defined by

- a 2TNB, $\mathrm{BN}_{\rightarrow}$, over $X_1, \cdots, X_n$,
- and a Bayesian network, $\mathrm{BN}^{(0)}$, over $X_1^{(0)}, \cdots, X_n^{(0)}$.

Definition 1.17. For a trajectory over $0, \cdots, T$, the **ground (unrolled) network** of a DNB is a model such that

- the dependency model for $X_1^{(0)}, \cdots, X_n^{(0)}$ is copied from $\mathrm{BN}^{(0)}$,
- and the dependency model for $X_1^{(t)}, \cdots, X_n^{(t)}$ is copied from $\mathrm{BN}_{\rightarrow}$.