BITS Pilani

Pilani | Dubai | Goa | Hyderabad

Probabilistic Graphical Model

Session # 4 : Bayesian Model

Seetha Parameswaran

seetha.p@pilani.bits-pilani.ac.in

The instructor is gratefully acknowledging the authors who made their course materials freely available online.

Table of Contents

Bayesian Network

Definition (Global Semantics)

A Bayesian Network is a directed acyclic graph $\mathcal{G}$ whose nodes represent the random variables $\{X_1, X_2, \ldots X_n\}$ and represents a joint distribution via the chain rule for the Bayesian Networks.

$$P(X_1, X_2, \ldots, X_n) = \prod_{i=1}^{n} P\left(X_i | Pa(X_i)\right) \tag{1}$$

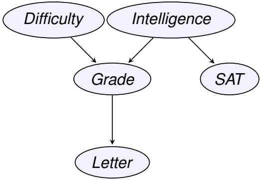
- Each node is associated with a CPD.

$$CPD(X_i) = P\left(X_i | Pa(X_i)\right)$$

Bayesian Network is Legal

A BN is a legal distribution; if

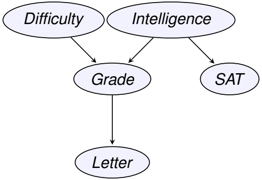
- $P \geq 0$
  - P is a product of CPDs.
  - CPDs are non negative.
- $\sum P = 1$



Bayesian Network is Legal

A BN is a legal distribution; if

- $P \geq 0$
  - P is a product of CPDs.
  - CPDs are non negative.
- $\sum P = 1$



$$\sum P = P(I, D, G, S, L)$$

$$= \sum_{D,I,G,S,L} P(I)P(D)P(G|I,D)P(S|I)P(L|G)$$

$$= \sum_{D,I,G,S} P(I)P(D)P(G|I,D)P(S|I)\sum_{L} P(L|G)$$

$$= \sum_{D,I,G} P(I)P(D)P(G|I,D)\sum_{S} P(S|I)$$

$$= \sum_{D,I} P(I)P(D)\sum_{G} P(G|I,D)$$

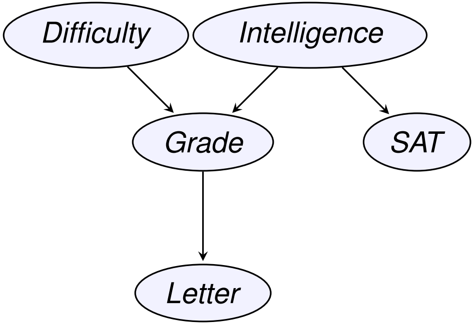
$$= \sum_{I} P(I)\sum_{D} P(D) = 1$$

Table of Contents

Reasoning Patterns

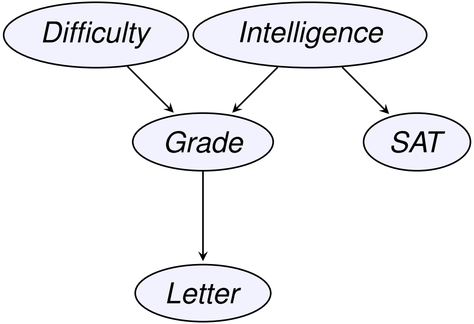
1. Causal reasoning
2. Evidential reasoning
3. Intercausal reasoning

Causal Reasoning



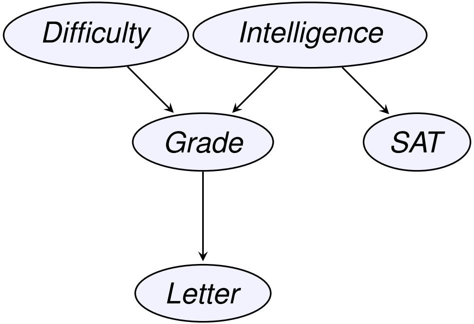
- How likely will a student get a strong recommendation?
  $P(l^1) = ?$
- Given that the student is not so intelligent, what is chance that he gets a strong letter?
  $P(l^1 | i^0) = ?$
- What if the course is easy?
  $P(l^1 | i^0, d^0) = ?$
- Queries that predict the effects of various factors or features are called causal reasoning.

Evidential Reasoning



- Given that a student gets C grade for a course, comment on his intelligence.
  $P(i^1|g^3) = ?$
- Given that the student got a weak letter, comment on his intelligence.
  $P(i^1|l^0) = ?$
- $P(i^1|l^0, g^3) = ?$
- Queries that reason from effects to causes are called evidential reasoning.

Intercausal Reasoning



- Given that a student gets C grade for a course, and a high SAT score, comment on his intelligence. $P(i^1|g^3, s^1) = ?$
- Does this give any idea regarding the difficulty of the course? $P(d^1|g^3, s^1) = ?$
- Explained away the poor grade via difficulty of the class.
- Explaining away is an instance of intercausal reasoning, where different causes of the same effect can interact.

Table of Contents

Dependency in BN

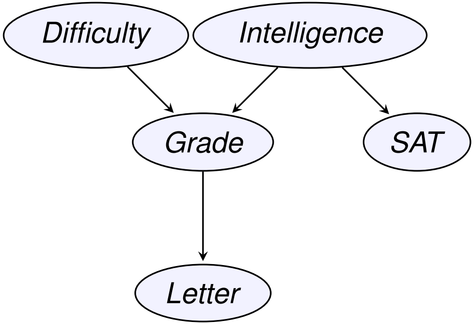
- A node depends directly only on its parents.
- If the student's grade is known, the quality of his recommendation letter is not influenced by information about any other variable. L is conditionally independent of all other nodes in the network given its parent G.

$$(L \perp I, D | G)$$

- The student's SAT score depends only on his intelligence. S is conditionally independent of all other nodes in the network given its parent I.

$$(S \perp D, G, L | I)$$

Dependency in BN



- Given the parents, a node can depend on its descendants.

$$G \text{ dependent } L$$

- Does G depend on S, given I and D?

$$(G \perp S | I, D)$$

- For D, both I and S are non descendants.

$$(D \perp I, S)$$

Bayesian Network Structure

Definition (Local Semantics)

A directed acyclic graph $\mathcal{G}$ whose nodes represent random variables $\{X_1, X_2, \ldots, X_n\}$ and $\mathcal{G}$ encodes a set of conditional independence assumptions.

$$\text{For each variable } X_i : \quad (X_i \perp NonDescendants_{X_i} | Pa_{X_i}) \tag{2}$$

- Pa_{X_i} represent parents of X_i in $\mathcal{G}$.
- $NonDescendants_{X_i}$ represent the random variables that are not descendants of X_i.
- $\mathcal{I}_l(\mathcal{G})$ represents the set of conditional independence assumptions called local independencies.

Bayesian Network Semantics

Local Semantics BN encodes a set of conditional independence assumptions.

$$\text{For each variable } X_i : \quad (X_i \perp NonDescendants_{X_i} | Pa_{X_i})$$

Global Semantics BN represents a joint distribution via the chain rule.

$$P(X_1, X_2, \ldots, X_n) = \prod_{i=1}^{n} P\left(X_i | Pa(X_i)\right)$$

Markov Blanket A node is conditionally independent of all other nodes in the Baysian Network, given its parents, children and children's parents.

$$\text{For each variable } X_i : \quad (X_i \perp \textit{other nodes} | Pa(X_i), Ch(X_i), Pa(Ch(X_i)))$$

Independency Map

Theorem

A distribution P satisfies local independencies $\mathcal{I}_l(\mathcal{G})$ associated with $\mathcal{G}$ if and only if P is representable as a set of CPDs associated with the graph $\mathcal{G}$.

Independency Map or I-map

- P be a distribution over $\mathcal{X}$.
- $\mathcal{I}(P)$ be the set of independence assertions $(X \perp Y|Z)$ that hold in P.
- Any independence that $\mathcal{G}$ asserts must also hold in P.

Definition

$\mathcal{G}$ is called an I-map for P if $\mathcal{I}_l(\mathcal{G}) \subset \mathcal{I}(P)$.

I-Map Example 1

X	Y	$P(X, Y)$
x^0	y^0	0.08
x^0	y^1	0.32
x^1	y^0	0.12
x^1	y^1	0.48

- Is $\mathcal{G}_\phi : X \perp Y$ an I-map of P ?

I-Map Example 1

X	Y	$P(X,Y)$
x^0	y^0	0.08
x^0	y^1	0.32
x^1	y^0	0.12
x^1	y^1	0.48

- Is $\mathcal{G}_\phi : X \perp Y$ an I-map of P ?

- $P(x^1) = 0.48 + 0.12 = 0.60$
- $P(y^1) = 0.32 + 0.48 = 0.80$
- $P(x^1, y^1) = 0.48 = P(x^1)P(y^1)$
- Hence X and Y are independent i.e $(X \perp Y)$
- $(X \perp Y) \in \mathcal{I}(P)$.
- $\mathcal{G}_\phi$ is an I-map of P.

I-Map Example 2

X	Y	$P(X, Y)$
x^0	y^0	0.4
x^0	y^1	0.3
x^1	y^0	0.2
x^1	y^1	0.1

- Is $\mathcal{G}_\phi : X \perp Y$ an I-map of P ?

I-MAP EXAMPLE 2

X	Y	$P(X, Y)$
x^0	y^0	0.4
x^0	y^1	0.3
x^1	y^0	0.2
x^1	y^1	0.1

- Is $\mathcal{G}_\phi : X \perp Y$ an I-map of P ?

- $P(x^1) = 0.2 + 0.1 = 0.1$
- $P(y^1) = 0.3 + 0.1 = 0.4$
- $P(x^1, y^1) \neq P(x^1)P(y^1)$
- Hence X and Y are not independent.
- $(X \perp Y) \notin \mathcal{I}(P)$.
- $\mathcal{G}_\phi$ is not an I-map of P.

Student Example - $\mathcal{B}^{Student}$

We know independence assumptions in $\mathcal{G}$

$$
\begin{aligned}
(D \perp I) &\implies P(D|I) = P(D) \\
(L \perp I, D|G) &\implies P(L|I, D, G) = P(L|G) \\
(S \perp D, G, L|I) &\implies P(S|I, D, G, L) = P(S|I)
\end{aligned}
$$

$$
\begin{aligned}
P(I, D, G, S, L) &= P(I)P(D|I)P(G|I, D)P(L|I, D, G)P(S|I, D, G, L) \\
&= P(I)P(D)P(G|I, D)P(L|I, D, G)P(S|I, D, G, L) \\
&= P(I)P(D)P(G|I, D)P(L|G)P(S|I, D, G, L) \\
&= P(I)P(D)P(G|I, D)P(S|I)P(L|G)
\end{aligned}
$$

A distribution P factorizes $\mathcal{G}$ if P can be represented as a chain rule of CPDs.

P factorizes over $\mathcal{G}$

Definition

Bayesian Network graph $\mathcal{G}$ over the variables $X_1, X_2, \ldots, X_n$ with a distribution P over the same space factorizes according to $\mathcal{G}$ if P can be expressed as a product of its CPDs.

$$P \quad \text{factorizes over } \mathcal{G} \quad \text{if} \quad P(X_1, X_2, \ldots, X_n) = \prod_i P(X_i | Pa^G(X_i)) \tag{3}$$

Theorem

For a Bayesian Network graph $\mathcal{G}$ over the variables $X_1, X_2, \ldots, X_n$ with a distribution P over the same space and $\mathcal{G}$ is an I-map for P, then P factorizes $\mathcal{G}$.

STUDENT EXAMPLE - $\mathcal{B}^{Student}$

By chain rule $\quad P(I, D, G, S, L) = P(I)P(D)P(G|I, D)P(S|I)P(L|G)$

By definition $\quad P(S|I, D, G, L) = \dfrac{P(I, D, G, S, L)}{P(I, D, G, L)}$

Marginalize over S $\quad P(I, D, G, L) = \sum_{S} P(I, D, G, S, L)$

$$= P(I)P(D)P(G|I, D)P(L|G) \sum_{S} P(S|I)$$

$$= P(I)P(D)P(G|I, D)P(L|G)$$

Student Example - $\mathcal{B}^{Student}$

$$
\begin{aligned}
P(S|I, D, G, L) &= \frac{P(I, D, G, S, L)}{P(I, D, G, L)} \\
&= \frac{P(I)P(D)P(G|I, D)P(S|I)P(L|G)}{P(I)P(D)P(G|I, D)P(L|G)} \\
&= P(S|I) \\
&(S \perp D, G, L|I)
\end{aligned}
$$

Independence assumption holds. $\mathcal{G}$ is an I-map for P.

$\mathcal{G}$ is an I-map of P

Theorem

For a Bayesian Network graph $\mathcal{G}$ over the variables $X_1, X_2, \ldots, X_n$ with a distribution P over the same space and if P factorizes according to $\mathcal{G}$, then $\mathcal{G}$ is an I-map for P.

Bayesian Network - another Definition

Definition

A Bayesian network is a pair $\mathcal{B} = (\mathcal{G}, \mathcal{P})$ where P factorizes over $\mathcal{G}$ and where P is specified as a set of CPDs associated with $\mathcal{G}$.

References

1. Probabilistic Graphical Models: Principles and Techniques by Daphne Koller and Nir Friedman. MIT Press. 2009
2. Artificial Intelligence: A Modern Approach (3rd Edition) by Stuart Russell, Peter Norvig
3. Mastering Probabilistic Graphical Models using Python by Ankur Ankan, Abhinash Panda. Packt Publishing 2015.
4. Learning in Graphical Models by Michael I. Jordan. MIT Press. 1999

Thank You !!!