BITS Pilani
Pilani | Dubai | Goa | Hyderabad

Probabilistic Graphical Model

Session # 5 : 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

Influence

- Influence means identifying the conditions when one random variable changes the beliefs about another random variable.
- When can X influence Y?

$P(Y/X) \neq P(Y)$

$X \quad \rightarrow$

$Y \quad X$

$\leftarrow \; Y \qquad y$

$X \leftarrow \rightarrow \quad Z \; \rightarrow$

$Y \quad X \; \leftarrow \; Z \; y$

$X \leftarrow Z \rightarrow Y$

$X \rightarrow Z \leftarrow Y$

$$\begin{tikzcd} X \arrow[dr] & & Y \arrow[dl] \\ & Z \arrow[r] & {} \end{tikzcd}$$

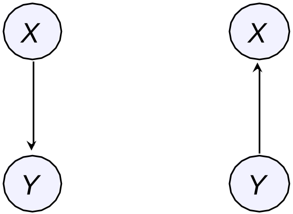
Direct Influence

- X and Y are directly connected.
- Direct parent child relation.
- They influence each other.

$X \rightarrow$

$Y \quad X$

$\leftarrow \quad Y$



Indirect Influence

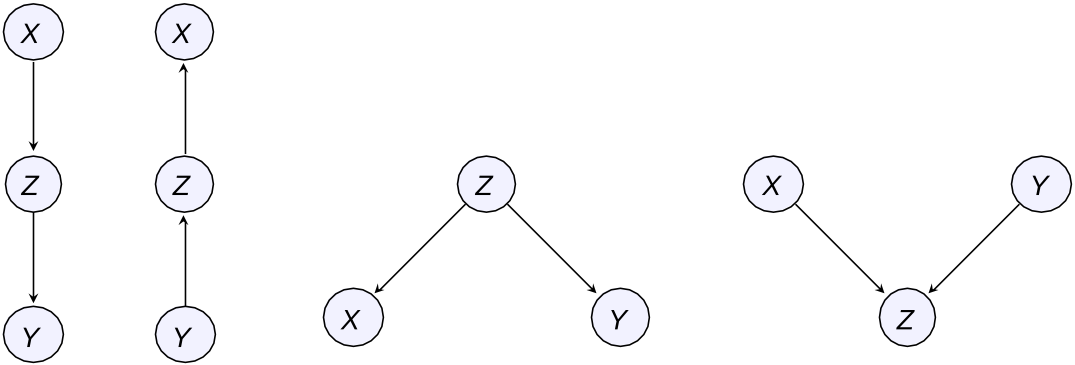
- X and Y are not directly connected, but there is a **trail** between them in the graph.
- X and Y connected by a trail through Z.

$X \rightarrow Z \rightarrow Y$	Causal chain
$Y \rightarrow Z \rightarrow X$ $\cancel{Y \leftarrow X \leftarrow Z}$	Evidential chain
$X \leftarrow Z \rightarrow Y$	Common cause chain
$X \rightarrow Z \leftarrow Y$	Common effect chain

[0] $X_1, \ldots, X_k$ form a trail in the graph, if for every $i = 1, \ldots, k-1$ we have either $X_i = X_{i+1}$.

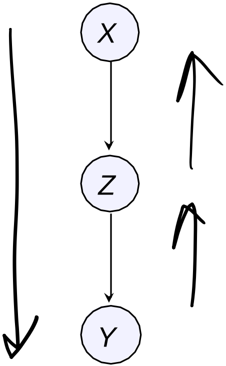
Indirect Influence

Causal | Evidential | Common Cause | Common Effect

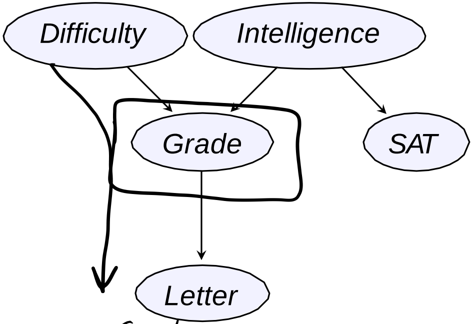


Indirect Causal Effect

L and D are not independent, $P(L/D) \neq P(L)$

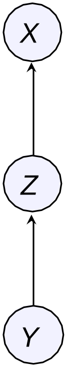


- $X \rightarrow Z \rightarrow Y$
- X can influence Y via Z, if Z is not observed. Eg: $D \rightarrow G \rightarrow L$
- X **cannot** influence Y via Z if Z is observed. Eg: $(L \perp I \mid G)$

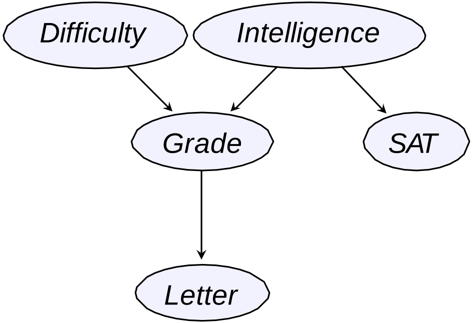


$$P(L/D,G) = P(L/G)$$

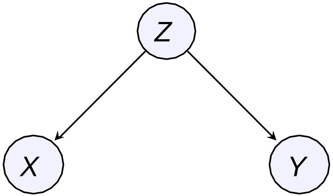
Indirect Evidential Effect



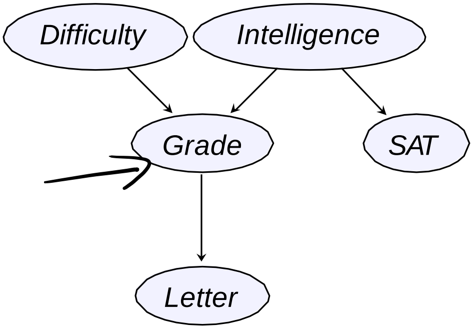
- $X \leftarrow \;\; Z \leftarrow \;\; Y$
- Y can influence X via Z , if Z is not observed. Eg: $D \longrightarrow \;\; G \longrightarrow L$
- Y **cannot** influence X via Z if Z is observed. Eg: $(L \perp I | G)$



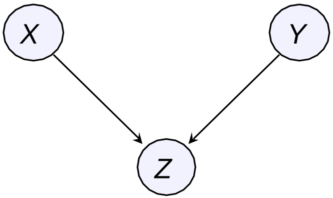
Common Cause Chain



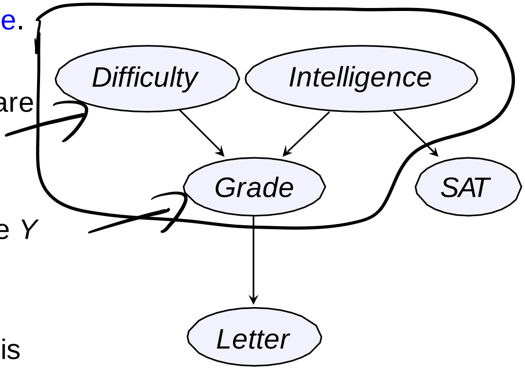
- $X \leftarrow \; Z \rightarrow \; Y$
- X can influence Y via Z , if Z is not observed.Eg: $G \leftarrow \; I \rightarrow S$
- X **cannot** influence Y via Z if Z is observed. Eg: $(S \perp G \mid I)$



Common Effect Chain



- $X \rightarrow Z \leftarrow Y$. This is called v-structure.
- When G is not observed I and D are independent. Eg: $D \rightarrow G \leftarrow I$
- X **cannot** influence Y via Z , if Z is not observed.
- When evidence G is observed, I and D are correlated.



Indirect Influence Flow

- X and Y are not directly connected, but connected by a **trail** through Z .
- If Z is **not** observed,

Causal chain	$: X \longrightarrow Z \longrightarrow Y$	:	active; Yes
Evidential chain	$: X \longleftarrow Z \longleftarrow Y$	:	active; Yes
Common Cause chain	$: X \longleftarrow Z \longrightarrow Y$	:	active; Yes
Common Effect chain	$: X \longrightarrow Z \longleftarrow Y$	:	inactive; NO

- V-structure is active if and only if either Z or one of Z 's descendants are observed.

Influence Flow

$X_1 \leftarrow X_2 \leftarrow X_3$

Definition

In a Bayesian Network G with a trail $X_1 = \ldots = X_n$, let a subset Z of variables be observed. The trail is **active** given Z if

- whenever we have a v-structure $X_{i-1} \rightarrow X_i \leftarrow X_{i+1}$, then X_i or one of its decendants are in Z.
- no other node along the trail is in Z. (not in v-structure)

One node can influence another if there is any trail along which influence can flow.

Influence Flow

Definition

In a Bayesian Network G with a trail $X_1 = \ldots = X_n$, let a subset Z of variables be observed. The trail is **active** given Z if

- whenever we have a v-structure $X_{i-1} \rightarrow X_i \leftarrow X_{i+1}$, then X_i or one of its decendants are in Z.
- no other node along the trail is in Z . (not in v-structure)

Influence Flow

Definition

In a Bayesian Network G with a trail $X_1 = \ldots = X_n$, let a subset Z of variables be observed. The trail is **active** given Z if

- whenever we have a v-structure $X_{i-1} \rightarrow X_i \leftarrow X_{i+1}$, then X_i or one of its decendants are in Z.
- no other node along the trial is in Z. (not in v-structure)

One node can influence another if there is any trail along which influence can flow.

For almost all parameterizations P of the graph G, the d-separation test precisely characterizes the independencies that hold for P.

Table of Contents

Directed Separation (d-separation)

The notion of d-separation provides us with a notion of separation between nodes in a directed graph.

Definition

In a Bayesian Network G, let X , Y , Z , be three sets of nodes. X and Y are d-separated, if there is no active trail between X and Y given Z.

$$I(G) = \{(X \perp Y | Z) : d-sep_G(X;Y|Z)\} \tag{1}$$

This set is also called the set of global Markov independencies.

Factorization & Independence in BN

Theorem

If P factorizes over G and $d-sep_G(X;Y|Z)$, then P satisfies, $(X \perp Y|Z)$.

(restatement of Theorem 3.3 from Daphne Koller's book)

$d\text{-}sep_G(X;Y|Z) \Rightarrow X \perp Y|Z \in \mathcal{I}(G)$ and $\mathcal{I}(G) \subseteq \mathcal{I}(P)$

$\Rightarrow X \perp Y|Z \in \mathcal{I}(P)$

[COMP538: Introduction to Bayesian Networks - Lecture 3: Probabilistic Independence and Graph Separation (hkust.edu.hk)]

Factorization & Independence in BN

Theorem

If P factorizes over G and d $-$ *sep*$_G(X; Y | Z)$, *then P satisfies,* $(X \perp Y | Z)$.

Proof for $D \perp S$

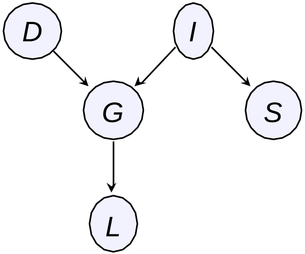
Active Trail $S \leftarrow I \rightarrow G \leftarrow D$

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

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

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

$$= P(D) \sum_{I} P(I)P(S | I) = P(D)P(S) \implies D \perp S$$



Factorization & Independence in BN

Theorem

If P factorizes over G; then any node is d-seperated from its non-descendants given its parents.

Factorization & Independence in BN

Theorem

If P factorizes over G, then any node is d-seperated from its non-descendants given its parents.

For L $\quad$ *descendants* = J

For L $\quad$ *nondescendants* = D, G, I, S

1. Trail $\quad S \leftarrow I \rightarrow G \rightarrow L$

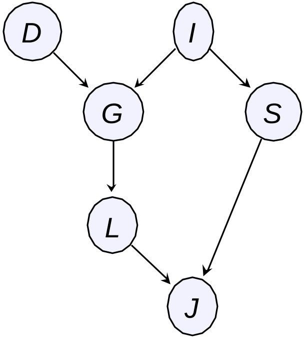
   Trail not active as G is observed.

   G parent of L, blocks the trail.

2. Trail $\quad S \leftarrow J \rightarrow L$

   Trail not active as only G is observed.

   J is descendant and is not observed.



I-map & d-separation

Definition

In a Bayesian Network G, P satisfies the corresponding independence statements.

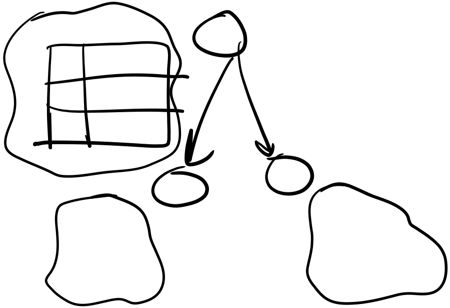
$$I(G) = \{(X \perp Y | Z) : d-sep_G(X; Y | Z)\} \tag{2}$$

If P satisfies $I(G)$, then G is an I-map of P.

Table of Contents

Tabular CPD

- Take all the possible combinations of different states of a variable and represent them in a tabular form.
- Tabular CPD is not the best choice to represent CPDs always.



Tabular CPD

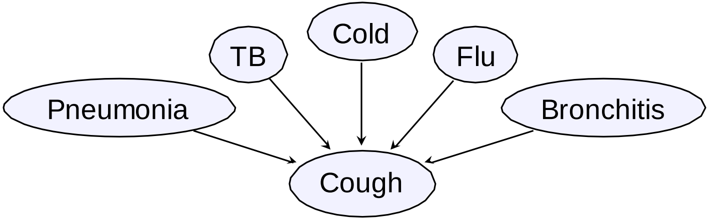
- Take all the possible combinations of different states of a variable and represent them in a tabular form.
- Tabular CPD is not the best choice to represent CPDs always.

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

Tabular CPD

- Take all the possible combinations of different states of a variable and represent them in a tabular form.
- Tabular CPD is not the best choice to represent CPDs always.

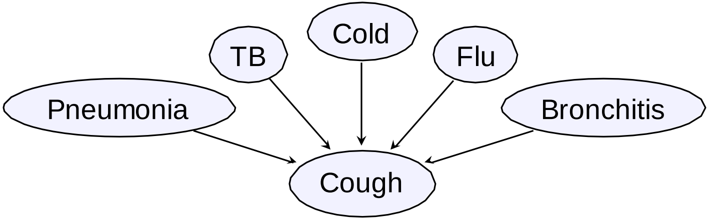
X	Y	$P(X, Y)$
x^0	y^0	0.08
x^0	y^1	0.32
x^1	y^0	0.12
x^1	y^1	0.48



Tabular CPD

- Take all the possible combinations of different states of a variable and represent them in a tabular form.
- Tabular CPD is not the best choice to represent CPDs always.

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



- For binary valued k parents, the size of the tabular CPD will be of $O(2^k)$.

Deterministic CPD

- Deterministic random variable are those, whose value depends only on the values of its parents in the model.

$$P(X|Pa_X) = \begin{cases} 1 & \text{if} \quad x = Val(Pa_X) \\ 0 & \text{otherwise} \end{cases} \tag{3}$$

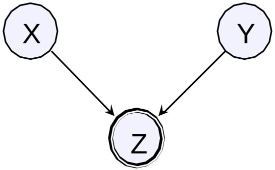
- Denote a deterministic variable by double circles.

Deterministic CPD

- Deterministic random variable are those, whose value depends only on the values of its parents in the model.

$$P(X|Pa_X) = \begin{cases} 1 & \text{if} \quad x = Val(Pa_X) \\ 0 & \text{otherwise} \end{cases} \qquad (3)$$

- Denote a deterministic variable by double circles.
- Eg: A Bayesian network for a logic gate. *X* and *Y* are the inputs, *A* and *B* are the outputs and *Z* is a deterministic variable representing the operation of the logic gate.



Context Specific CPD

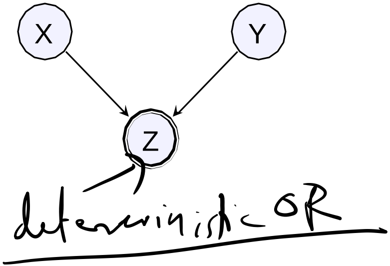
- Context specific independence is a type of independence for random variables X, Y, Z and an assignment c.
- The independence statement only holds for a particular value of conditioning variable c.

$$P \models (X \perp_c Y \mid Z, c) \qquad (4)$$

$$(X \perp Y \mid Z) \quad P((X,Y)/Z) = P(X/Z)\,P(Y/Z)$$

Context Specific CPD

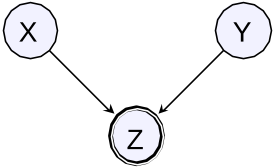
Which of the following Context specific independences hold when Z is a deterministic OR of X and Y?



1. $\boxed{(Z \perp X \mid y^0)}$
2. $(Z \perp X \mid y^1)$
3. $(X \perp Y \mid z^0)$
4. $(X \perp Y \mid z^1)$

Context Specific CPD

Which of the following Context specific independences hold when Z is a deterministic OR of X and Y?



1. $(Z \perp X \mid y^0)$ False
   When $Y = 0$, $Z = X$. So not independent.
2. $(Z \perp X \mid y^1)$ True
   When $Y = 1$, $Z = 1$. So context specific independent.
3. $(X \perp Y \mid z^0)$ True
   When $Z = 0$, $X \perp Y$. So context specific independent.
4. $(X \perp Y \mid z^1)$ False
   When $Z = 1$, $X \not\perp Y$. So not independent.

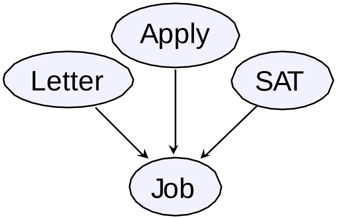
Tree Structured CPD

- Tree Structured CPD encode dependence of a child on a parent.

Tree Structured CPD

- Tree Structured CPD encode dependence of a child on a parent.

Bayesian Network

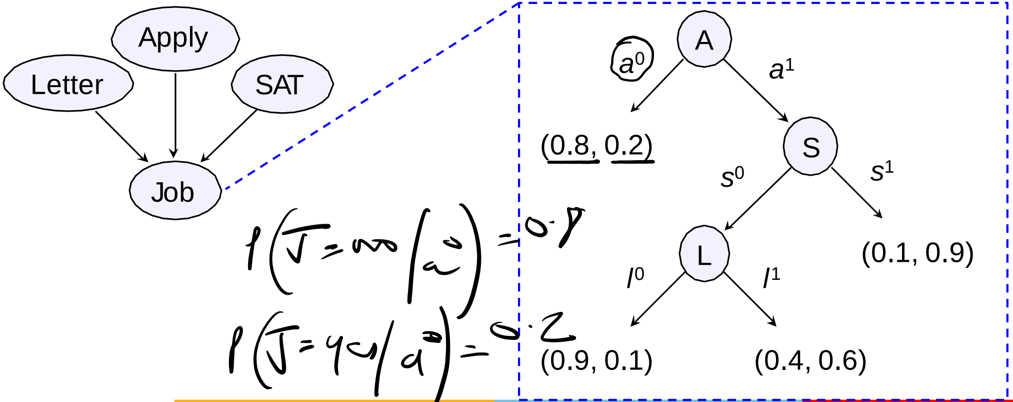


Tree Structured CPD

- Tree Structured CPD encode dependence of a child on a parent.

Bayesian Network

CPD of Job(no, yes)



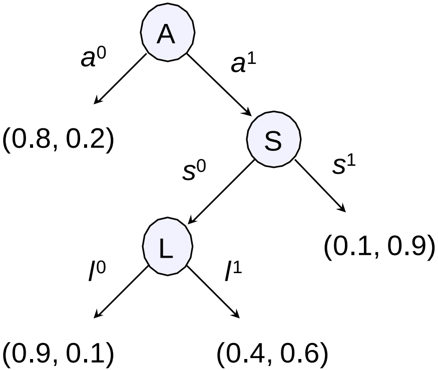
$P(J = no \mid a^0) = 0.8$

$P(J = yes \mid a^0) = 0.2$

Tree Structured CPD

Which of the following Context specific independences hold?

CPD of Job(no, yes)

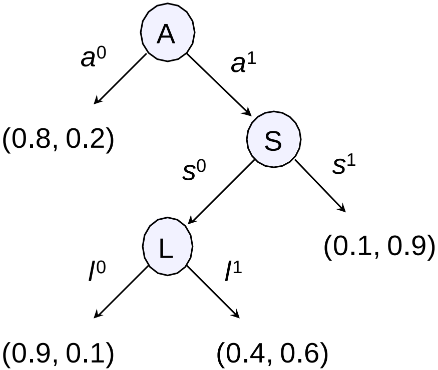


1. $(J \perp_c L \mid a^1, s^1)$
2. $(J \perp_c L \mid a^1)$
3. $(J \perp_c L \mid s^1, A)$
4. $(J \perp_c L, S \mid a^0)$

Tree Structured CPD

Which of the following Context specific independences hold?

CPD of Job(no, yes)



1. $(J \perp_c L \mid a^1, s^1)$ True
   Context specific independent.
2. $(J \perp_c L \mid a^1)$ False
   Not independent.
3. $(J \perp_c L \mid s^1, A)$ True
   Context specific independent.
4. $(J \perp_c L, S \mid a^0)$ True
   Context specific independent.

Table of Contents

Independence

Definition

Independence: $P(\{X_1, X_2, \ldots, X_n\}) = \prod_{i=1}^{n} P(X_i)$

Local Independency: $\mathcal{I}_l(G) : (X_i \perp NonDescendants_{X_i} \mid Pa_{X_i}) \quad \forall X_i$

Independency Map: G is an I-map for P if $\mathcal{I}_l(G) \subset \mathcal{I}(P)$

Theorem

P satisfies $\mathcal{I}_l(G)$ if P is representable as a set of CPDs associated with G.

Bayesian Network

Definition

A Bayesian Network $B = (G, P)$ where P factorizes over G and where P is specified as a set of CPDs associated with G.

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

$$G \quad \text{encodes} \quad I_l(G) \quad \text{if} \quad \forall X_i : \quad (X_i \perp NonDescendants_{X_i} \mid Pa_{X_i})$$

Theorem

If G is an I-map for P, then P factorizes G.

If P factorizes according to G, then G is an I-map for P.

Reasoning Patterns

Causal reasoning Queries that predict the effects of various factors or features are called causal reasoning.

Evidential reasoning Queries that reason from effects to causes are called evidential reasoning.

Intercausal reasoning Explaining away is an instance of intercausal reasoning, where different causes of the same effect can interact.

- Inference by Enumeration

$$P_B(Y = y \mid E = e)$$

Influence Flow

Definition

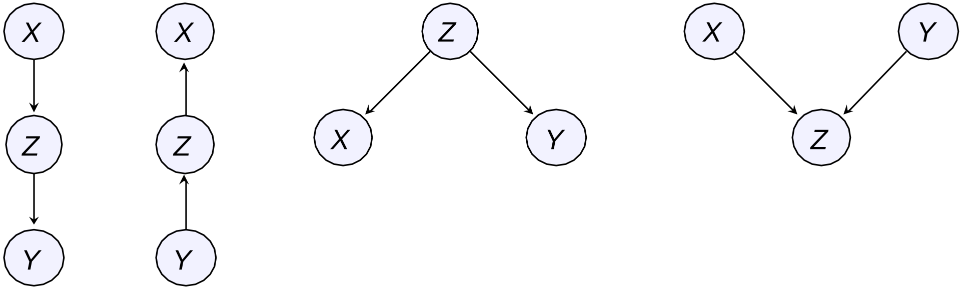
In a Bayesian Network G with a trail $X_1 \rightleftharpoons \ldots \rightleftharpoons X_n$, let a subset Z of variables be observed. The trail is **active** given Z if

- whenever we have a v-structure $X_{i-1} \rightarrow X_i \leftarrow X_{i+1}$, then X_i or one of its decendants are in Z.
- no other node along the trial is in Z. (not in v-structure)

One node can influence another if there is any trail along which influence can flow.

Indirect Influence

Causal | Evidential | Common Cause | Common Effect



Active | Active | Active | Inactive

If Z is **not** observed

Directed Separation (d-separation)

Definition

In a Bayesian Network G, let X , Y , Z , be three sets of nodes. X and Y are d-separated, if there is no active trail between any node given Z .

$$I(G) = \{(X \perp Y | Z) : d - sep_G(X; Y | Z)\} \tag{5}$$

Questions

1. Given a Bayesian Network, find the appropriate factorization of joint distribution.
2. Given a Bayesian Network, identify the active trails.
3. Given a Bayesian Network, identify the I-maps.
4. Given a Bayesian Network, identify the d-seperations.
5. In a Bayesian Network, infer by enumeration, the probability of an event when some evidences are observed.
6. Given a toy application, generate the CPD and Bayesian Network.
7. Given CPDs, generate a Bayesian Network.
8. Given a joint distribution in the factorized form, generate a Bayesian Network.
9. Given a Bayesian Network, identify the conditional independencies.
10. Given a CPD, identify the context specific independencies.

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 !!!