BITS Pilani
Pilani | Dubai | Goa | Hyderabad

Probabilistic Graphical Model

Session # 7 : Undirected Graphical 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

$X \quad Y \quad Z$

$X \perp Y \mid Z$

$Z + \alpha = Z'$

$X \perp Y \mid Z' \rightarrow$ yes $\rightarrow$ no

$X \rightarrow \alpha \leftarrow Y$

Active Trail in Markov Network

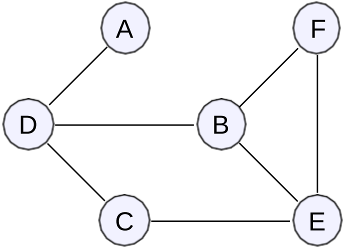
Definition

Let H be a Markov network structure and let $X_1 -, \ldots, -X_n$ be a path in H.
Let $Z \subseteq X$ be a set of observed variables.
The path $X_1 -, \ldots, -X_n$ is **active** given Z if none of the X_i is in Z.

- Influence has to flow through unobserved variables along the trail.
- Once a variable is observed along the trail, the influence is blocked.

Markov Network - Example

- Find the active trails given B is observed.



$$A - D - C - E - F$$

Separation in Markov Network

Definition

A set of nodes Z separates X and Y in H, a Markov network structure, if there is no active path between any node in X and Y given Z.

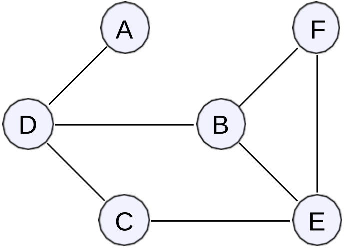
- Denote Separation as $sep_H(X;Y|Z)$
- Global Independencies

$$I_g(H) = \{(X \perp Y|Z) : sep_H(X;Y|Z)\} \quad (1)$$

- The Independencies in $I(H)$ are precisely those that are guaranteed to hold for every distribution P over H.

Markov Network - Example

- Find the global Independencies.



$$
\begin{aligned}
I_g(H) &= \{(A \perp B \mid D) : sep_H(A;B \mid D)\} \\
&= \{(A \perp C \mid D) : sep_H(A;C \mid D)\} \\
&= \{(A \perp E \mid D) : sep_H(A;E \mid D)\} \\
&= \{(A \perp F \mid D) : sep_H(A;F \mid D)\}
\end{aligned}
$$

Factorization implies Independence

Theorem

Let P be a distribution over X and H a Markov Network structure over X.
*If P is a Gibbs distribution that factorizes over H ~~and~~ H is an **I-map** for P.*

then

Proof

Let X, Y and Z be any 3 disjoint sets in $\mathcal{X}$ such that Z separates X and Y in $\mathcal{H}$

Consider two cases:

ⓐ $X \cup Y \cup Z = \mathcal{X}$ (set of all vertices in the graph)

We need to show that $P \models (X \perp Y \mid Z)$

(X is independent of Y given Z)

Proof

Since Z separates X and Y, there are no direct edges between X and Y

Therefore any clique in H is completely contained in $X \cup Z$ or $Y \cup Z$

Let I_X be the indexes of the cliques completely contained in $X \cup Z$. Let I_Y be the remaining cliques.

innovate achieve lead

Proof

Since P is a Gibbs distribution that factorizes over H we can write

$$P(X_1, X_2 \ldots X_n) = \frac{1}{Z} \prod_{i \in I_X} \phi_i(D_i) \cdot \prod_{i \in I_Y} \phi_i(D_i)$$

None of the factors in the first product involve any variable in Y and none in the second product involve any variable in X.

Proof

We can rewrite $P(X_1, X_2, \dots X_n)$ such that

$$P(X_1, X_2 \dots X_n) = \frac{1}{Z} f(X, Z)\, g(Y, Z)$$

$\to$ this Z is the partition function

Thus $(X \perp Y \mid Z)$

Case 1: $X \cup Y \cup Z \subset \mathcal{X}$ (proper subset of $\mathcal{X}$)

Let $U = \mathcal{X} - X \cup Y \cup Z$

Proof

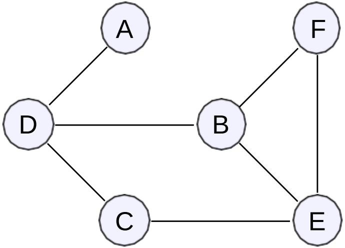
Partition U into disjoint sets U_1 and U_2 such that Z separates $X \cup U_1$ from $Y \cup U_2$ in $\mathcal{H}$.

Using the previous argument we can conclude $X, U_1 \perp Y, U_2 | Z$

Using the decomposition property we can conclude that $X \perp Y | Z$

Markov Network - Example

- Identify a possible factorization for the Markov Network.



$$P_G = \phi_1(A, D)\phi_2(B, E, F)$$
$$\phi_3(D, B)\phi_4(D, C)\phi_5(C, E)$$

Independence implies Factorization

Theorem

Hammersley-Clifford theorem:

*Let P be a **positive** distribution over $\mathcal{X}$ and $\mathcal{H}$ a Markov Network structure over $\mathcal{X}$.*

*If $\mathcal{H}$ is an **I-map** for P, then P is a Gibbs distribution that factorizes over $\mathcal{H}$.*

Need for positive distribution

Look at Example 4.4 from Daphne Koller's book

Distribution P over X_1, X_2, X_3, X_4 which has a value $\frac{1}{8}$ for 8 combinations which are

$(0\,0\,0\,0)\ (1\,0\,0\,0)\ (1\,1\,0\,0)\ (1\,1\,1\,0)$
$(0\,0\,0\,1)\ (0\,0\,1\,1)\ (0\,1\,1\,1)\ (1\,1\,1\,1)$ and 0 for all other combinations.

Need for positive distribution

The distribution is not a true distribution as some entries are 0s

Let H be the graph

$$X_1 - X_2 - X_3 - X_4 - X_1$$

P satisfies the global independencies with respect to H

Need for positive distribution

The graph asserts that $X_1 \perp X_3 \mid X_2, X_4$

Does P satisfy this relationship?

We have $P(x_1 = 1 / x_2 = 1, x_4 = 0) = 1$

So for this assignment of x_2, x_4 we can see that x_1 is independent of x_3

Similarly for other assignments & other independencies

Need for positive distribution

Does this distribution factorize over H?

Can we express the distribution P as

$$\phi_1(X_1, X_2)\,\phi_2(X_2, X_3)\,\phi_3(X_3, X_4)\,\phi_4(X_4, X_1)\ ?$$

X_1	X_2	$\phi_1(X_1, X_2)$
0	0	α_1
0	1	α_2
1	0	α_3
1	1	α_4

X_2	X_3	$\phi_2(X_2, X_3)$
0	0	β_1
0	1	β_2
1	0	β_3
1	1	β_4

X_3	X_4	$\phi_3(X_3, X_4)$
0	0	γ_1
0	1	γ_2
1	0	γ_3
1	1	γ_4

X_4	X_1	$\phi_4(X_4, X_1)$
0	0	ρ_1
0	1	ρ_2
1	0	ρ_3
1	1	ρ_4

Need for positive distribution

We need to find values for the parameters

$(\alpha_1\ \alpha_2\ \alpha_3\ \alpha_4) \cdots (\delta_1\ \delta_2\ \delta_3\ \delta_4)$

$P(0\,0\,0\,0) = \frac{1}{8}$ $\qquad \alpha_1 \beta_1 \gamma_1 \delta_1 \propto \frac{1}{8}$

$P(0\,0\,1\,0) = 0$ $\qquad \alpha_1 \underline{\beta_2 \gamma_3} \delta_1 \propto 0 \Rightarrow$ one of

β_2, γ_3 must be 0

$P(0\,0\,1\,1) = \frac{1}{8}$ $\qquad \alpha_1 \beta_2 \gamma_4 \delta_3 \propto \frac{1}{8} \Rightarrow \beta_2$ cannot be 0

Need for positive distribution

So γ_3 must be a 0.

But $P(1110) = \frac{1}{8} \Rightarrow \alpha_4 \beta_4 \gamma_3 \rho_2 \propto \frac{1}{8}$

So γ_3 cannot be a 0

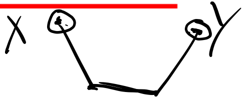
[contradiction]

Proof of Hammersley-Clifford Theorem

Take a look at
https://vision.in.tum.de/_media/teaching/ss2017/pgmcv/in2329-02_gm.pdf

The statement of the theorem here is different from Koller's book (but equivalent) and this is a full proof → focus on Forward Direction in the above document

Pairwise Independencies



Definition

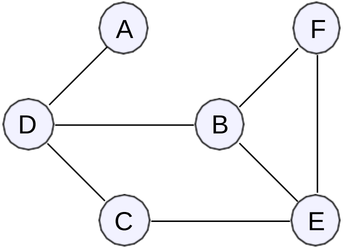
Let H be a Markov network structure. Pairwise Independencies associated with H is defined as

$$I_p(H) = \{(X \perp Y | X - \{X, Y\}) : edge(X - Y) \notin H\} \tag{2}$$

X and Y are independent given all the remaining nodes in the network.

Markov Network - Example

- Find the pairwise independencies.



For node A

$$
\begin{aligned}
I_p(H) &= (A \perp C \mid D, B, E, F : edge(A - C) \notin H) \\
&= (A \perp B \mid D, C, E, F : edge(A - B) \notin H) \\
&= (A \perp E \mid D, B, C, F : edge(A - E) \notin H) \\
&= (A \perp F \mid D, B, E, C : edge(A - F) \notin H)
\end{aligned}
$$

Markov Blanket

Definition

For a given graph H, the Markov blanket of X in H is defined as neighbours of X in H.

$$MB_X = \cancel{\{Pa(X), Ch(X), Pa(Ch(X))\}} \tag{3}$$

this defn does not apply here

- Markov blanket is the set of nodes containing parents, children, and children's parents.

Local Independencies

Definition

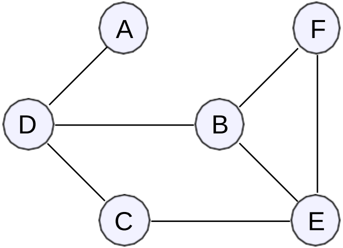
Let H be a Markov network structure. Local Independencies associated with H is defined as

$$I_l(H) = \{(X \perp \mathcal{X} - \{X + MB_H(X)\} \mid MB_H(X)) : X \in \mathcal{X}\} \quad (4)$$

X is independent of the rest of the nodes in the graph given its immediate neighbours

Markov Network - Example

- Find the local independencies and Markov blanket.



For node ~~A~~ D

$$MB_{\cancel{A}\,D} = \{A, D, B, C\}$$

$$I_l(H) = (\cancel{A}\,D \perp E, F \mid MB_{\cancel{A}\,D})$$

Independencies

Definition

For any Markov network H and any distribution P,

$$\text{if} \quad P \models \mathcal{I}_l(H) \quad \text{then} \quad P \models \mathcal{I}_p(H) \tag{5}$$

$$\text{if} \quad P \models \mathcal{I}_g(H) \quad \text{then} \quad P \models \mathcal{I}_l(H) \tag{6}$$

- $\mathcal{I}_p(H)$ is strictly weaker that $\mathcal{I}_l(H)$ which is strictly weaker that $\mathcal{I}_g(H)$
- For a positive distribution P,

$$P \models \mathcal{I}_p(H)$$
$$P \models \mathcal{I}_l(H)$$
$$P \models \mathcal{I}_g(H)$$

Pairwise and Local Independencies

Theorem: For any Markov network H and any distribution P, we have that if $P \models I_\ell(H)$ then $P \models I_p(H)$

Let us look at its proof

Proof

First we need to prove the Weak Union property

$$(X \perp Y, W \mid Z) \implies (X \perp Y \mid W, Z)$$

If $P(W = w, Z = z) = 0$ then the implication follows trivially

Proof

Assume that $P(Z) \neq 0$ and $P(Z,W) \neq 0$. Then

$$P(X,Y \mid Z,W) = \frac{P(X,Y,W \mid Z)}{P(W \mid Z)}$$

Since $(X \perp Y,W \mid Z)$ we can write $P(X,Y,W/Z)$

$$= P(X/Z)\,P(Y,W/Z)$$

Proof

Thus we can write $P(X, Y/Z, W) = \dfrac{P(X/Z)\,P(Y, W/Z)}{P(W/Z)}$

$$= \frac{P(X/Z)\,P(Y/W, Z)\,P(W/Z)}{P(W/Z)} = P(X/Z)\,P(Y/Z, W)$$

According to the Decomposition Rule

$$(X \perp Y, W \,|\, Z) \Rightarrow (X \perp W \,|\, Z)$$

Proof

Using the Decomposition Rule we have

$$(X \perp Y, W / Z) \Rightarrow (X \perp W / Z)$$ which means

$$P(X / W, Z) = P(X / Z)$$

We had $P(X, Y / Z, W) = P(X / Z)\, P(Y / Z, W)$ from which

we have $P(X, Y / Z, W) = P(X / W, Z)\, P(Y / Z, W)$

$$\Rightarrow (X \perp Y / Z, W)$$

Proof

Armed with the weak union property, we can prove the result we want:

Let $Z = N_H\{X\} \rightarrow$ neighbor set of X

Let $W = \mathcal{X} - \{X, Y\} - Z$

Markov Assumption $(\mathcal{I}_\ell(H))$ tells us that

$(X \perp \{Y\} \cup W \mid Z)$

innovate achieve lead

Proof

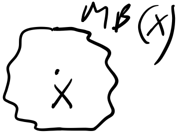
Applying the just proved Weak Union property we have

$$(X \perp \{Y\} \cup W \mid Z) \Rightarrow (X \perp Y \mid W, Z)$$

i.e Markov Property $\Rightarrow$ Pairwise Independence

Local and Global Independencies

Proposition 4.4 For any Markov network H and any distribution P we have that if $P \models \mathcal{I}(H)$ then $P \models \mathcal{I}_\ell(H)$



$$\mathcal{I}(H) = \{(X \perp Y \mid Z) : \mathrm{sep}_H(X ; Y \mid Z)\}$$

$$\mathcal{I}_\ell(H) = \{X \perp \mathcal{X} - \{X\} - MB_H(X) \mid MB_H(X)\}$$

When are $I(H)$, $I_l(H)$, $I_p(H)$ all equivalent?

For positive distributions

Theorem 4.4 Let P be a positive distribution. If P satisfies $I_p(H)$ then P satisfies $I(H)$

Earlier we showed that

$$P \models I(H) \Rightarrow P \models I_l(H) \Rightarrow P \models I_p(H)$$

for any distribution P

Proof of Theorem 4.4

We want to show that $P \models I(H)$ given that $P \models I_\ell(H)$ for all disjoint sets X, Y and Z

$P \models I(H)$ means that whenever it is true that $sep_H(X ; Y / Z)$, then $P \models (X \perp Y / Z)$

Proof of Theorem 4.4

Proof is by induction on the size of Z

When $|Z| = n-2$, the statement follows trivially since $I_P(H)$ and $I(H)$ then mean the same thing (X and Y are individual nodes in this case)

Assume that $\mathrm{Sep}_H(X; Y | Z) \Rightarrow P \models (X \perp Y | Z)$ holds for every Z' whose $|Z'| = K$.

Proof of Theorem 4.4

Let Z be any set such that $|Z| = k-1$

There are two cases:

Case (a): $X \cup Y \cup Z = \chi$ (set of all vertices)

As $|Z| < n-2$ we must have either $|X| > 2$ or $|Y| > 2$. Assume that $|Y| > 2$

There exists $A \in Y$. Let $Y' = Y - \{A\}$

Proof of Theorem 4.4

Since $\mathrm{Sep}_H(X; Y \mid Z)$ we must also have

$\mathrm{Sep}_H(X; Y' \mid Z)$ and $\mathrm{Sep}_H(X; A \mid Z)$

Separation is monotonic so we have

$\mathrm{Sep}_H(X; Y' \mid Z \cup \{A\})$ and $\mathrm{Sep}_H(X; A \mid Z \cup Y')$

Each of the sets $Z \cup \{A\}$ and $Z \cup Y'$ has size at least K

Proof of Theorem 4.4

Therefore the induction hypothesis applies and we must have P satisfies:

① $X \perp Y' \mid Z \cup \{A\}$ and ② $X \perp A \mid Z \cup Y'$

Since P is a positive distribution the intersection property applies and we have

$P \models X \perp Y' \cup \{A\} \mid Z$, ie $X \perp Y \mid Z$ (done)

Proof of Theorem 4.4

Case (b): $X \cup Y \cup Z \subseteq \mathcal{X}$

In this case there is a node A that does not belong to $X \cup Y \cup Z$.

We have $sep_H(X; Y \mid Z)$. From monotonicity of separation we have $sep_H(X; Y \mid Z \cup \{A\})$

Proof of Theorem 4.4

Since X and Y are separated given Z, there cannot exist a path between X and A and between Y and A given Z. At most one of these paths can exist $\rightarrow$ assume that there is no path between X and A given Z.

By monotonicity, we must have $\mathrm{Sep}_{\mathcal{H}}(X; A \mid Z \cup Y)$

Proof of Theorem 4.4

As before $Z \cup \{A\}$ and $Z \cup Y$ have size at least K

Therefore the <u>induction hypothesis</u> applies and we must have

$$X \perp Y \mid Z \cup \{A\} \text{ and } X \perp A \mid Z \cup Y$$

Use intersection to get $X \perp Y, A \mid Z$ and Decomposition to get $X \perp Y \mid Z$

Constructing Graphs from Distributions

- A fully connected graph has no independence conditions and, hence, it can be an I-Map of any probability distribution.
- To encode the Independencies in a given distribution *P* using a graph structure, use minimal I-map.
- Two approaches for constructing a minimal I-map
  1. using the pairwise Markov Independencies.
  2. using the local Independencies.

Constructing Graphs from Distributions

Pairwise Markov independencies

- Let P be a positive distribution.
- Let H be defined by introducing an edge $\{X, Y\}$.
- If the edge $\{X, Y\}$ is not in H, then X and Y must be independent given all other nodes in the graph.
- To guarantee that H is an I-map, add direct edges between all pairs of nodes X and Y such that

$$P \not\models (X \perp Y \mid X - \{X, Y\})$$

X Y

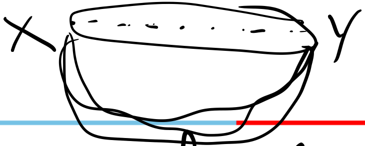
- Then Markov network H is the unique minimal I-map for H.
- To guarantee that H is an I-map, add direct edges between all pairs of nodes X and Y, such that they are dependent even on observing all the other variables in the network.

innovate achieve lead

Minimal I-Map

Theorem: Let P be a positive distribution and let H be obtained by introducing an edge $X - Y$ whenever $P \not\models X \perp Y \mid \mathcal{X} - \{X, Y\}$. Then the Markov network is the unique minimal I-map for P.

by construction



Proof

Why is H an I-map for P? By construction P satisfies $I_P(H)$. Since P is a positive distribution

pairwise independency

$I_P(H)$ is equivalent to $I(H)$

Why is H a minimal I-map? If we eliminate some edge $X - Y$ from H it would mean that $X \perp Y \mid \mathcal{X} - \{X, Y\}$ which we know to be false for P.

Proof

Why is it a unique minimal I-map

Let us say that there is another I-map H' which is also minimal.

H' must contain all the edges of H; otherwise it would imply some independencies that don't exist in P. If H' contains any additional edges then it is no longer minimal

Constructing Graphs from Distributions

Local Independencies

- Let P be a positive distribution.
- For each variable X , define the neighbors of X to be a minimal set of nodes Y that render X independent of the rest of the nodes. i.e. Markov Blanket of X.
- A set U is a Markov blanket of X in a distribution P if $X \notin U$ and if U is a minimal set of nodes such that

$$(X \perp X - \{X+U\} \mid U) \in \mathcal{I}(P)$$

- Then define a graph H by introducing an edge $\{X, Y\}$ for all X and all $Y \in MB_P(X)$
- Then Markov network H is the unique minimal I-map for H. [invalid argument]
- For each variable X , find the minimal set of nodes. Observing these makes the variable independent of all the variables. Then, add an edge between the variable and all the nodes in the set.

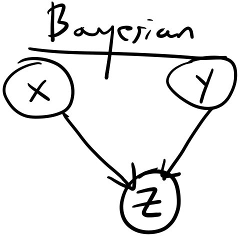
Table of Contents

Bayesian Network and Markov Network

$P(X_i \mid Pa(X_i))$

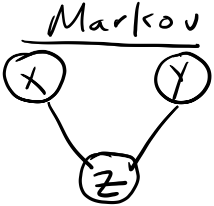
is it not a factor?

$\phi(a, b, c)$



Both

- Parametrize a probability distribution using a graphical model.
- Encode the Independencies among the random variable.



Convert Bayesian Network to Markov Network

$$B \rightarrow A \leftarrow C$$

$$P(A/B,C) = ?$$

Two perspectives

1. Parameterization perspective – represent the probability distribution of the Bayesian model using a fully parameterized Markov model.
2. Independencies perspective – represent the independence constraints encoded by the Bayesian model using the Markov model.

Convert Bayesian Network to Markov Network

Parameterization perspective

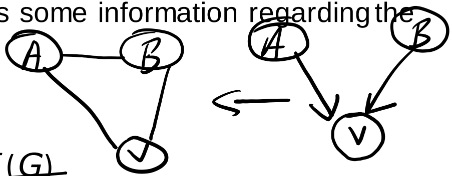
- Probability distribution P_B , B is a parameterized Bayesian network over a graph G.
- The parameterization of the Bayesian network B, can also be viewed as a parameterization of a Gibbs distribution.
- Each CPD $P(X_i \mid Pa_{X_i})$ is a factor with scope $\{X_i, Pa_{X_i}\}$. $\phi(X_i, Pa(X_i))$
- This set of factors defines a Gibbs distribution with the partition function equal to 1.

Convert Bayesian Network to Markov Network

Independencies perspective

1. Replace all the directed edges between the nodes with undirected edges.
2. Add additional undirected edges between nodes that are parents of the node.

- This new structure is called moral graph of Bayesian network.
- $M[G]$ is an I-Map for distribution P_B ,
- Moral graph $M[G]$ of a Bayesian model G loses some information regarding the Independencies.
- $M[G]$ is a minimal I-Map for G.



$I(M[G]) \subseteq I(G)$

Why is this true?

$I(M(G)) \subseteq I(G)$

Consider a Z that does not separate X and Y in the Bayesian network G. Could it happen that Z will separate X and Y in $M(G)$ [the Markov network] and thus prevent $I(M(G)) \subseteq I(G)$?

$X \rightarrow \cdots \; a \rightarrow v \leftarrow b \; \cdots \; Z$ in G

$X - - a - v - b - Z$ in $M(G)$

$I(M(G)) \subseteq I(G)$

The only danger lies in paths having converging nodes

Specifying a converging node v in G activates the path between X and Y in G but inactivates that path in M(G)

Fortunately there is another path in M(G) that comes to the rescue since v's parents a and b are joined by an edge

Why is M(G) a minimal I-map for G?

M(G) = G + edges between parents of a given node.

Can we get rid of some edges in M(G) such that it will continue to remain an I-map for G?

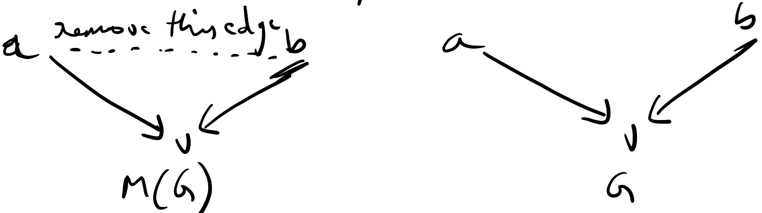
If we get rid of an edge in M(G) that has a directed equivalent in G, we will have an independency introduced in M(G) that is not there in G, i.e

$X \perp Y \mid$ all nodes in M(G)

$X \rightarrow Y$

Why is $M(G)$ a minimal I-map for G?

If we remove an edge that was added between the parents of a node in G, we introduce an independency $a \perp b \mid v$ in $M(G)$ that is not there in G, so $I(M(G)) \not\subseteq I(G)$



$\therefore$ we cannot remove any edges in $M(G)$

Moral Graph

Definition

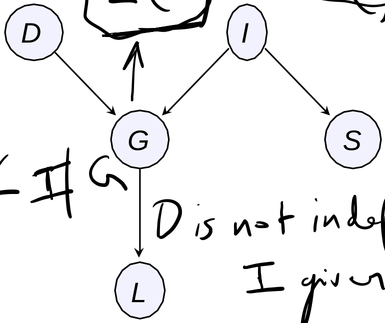
The moral graph $M[G]$ of a Bayesian network structure G over X is the undirected graph over X that contains an undirected edge between X and Y if:

(a) there is a directed edge between them (in either direction) or

(b) X and Y are both parents of the same node.

Convert Bayesian Network to Markov Network

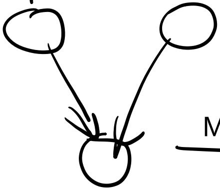
$I(M) \le I(B) \le P$



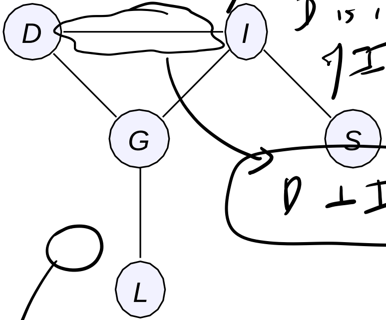
$D \not\perp I \mid G$

D is not independent of I given G

Bayesian Network



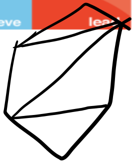
if this edge is removed, D is independent of I given G



$D \perp I \mid G$

Markov Network

Convert Markov Network to Bayesian Network

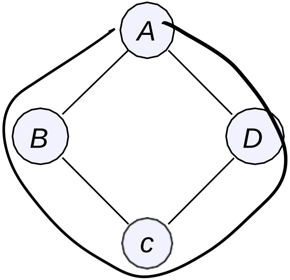


Independencies perspective

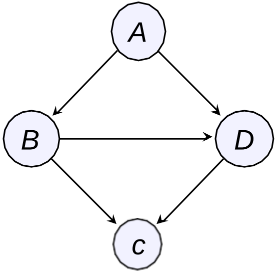
1. Replace all the undirected edges between the nodes with directed edges.
2. Partition all loops into triangles. Add edges to the network to make it chordal.

- Any Bayesian network I-map for the given Markov network must add triangulating edges into the graph, so that the resulting graph is chordal. This process is called triangulation.
- A triangulated or chordal graph is a graph in which each of its cycles of four or more vertices has a chord.
- By simply converting edges of a non-triangulated graph into directed ones, introduces immoralities. An immorality is a v-structure $X \rightarrow Z \leftarrow Y$, if there is no directed edge between X and Y.

Convert Markov Network to Bayesian Network



Markov Network



Bayesian Network

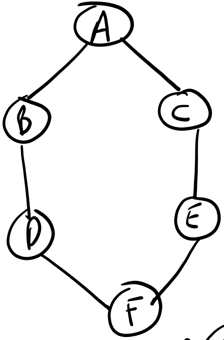
$A \quad \underline{C} \quad \underline{B}$

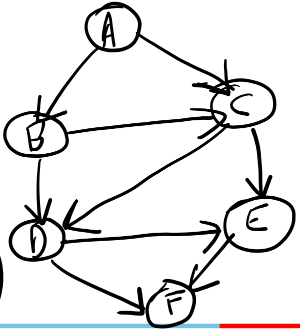
Markov Network to Bayesian network

Example 4.17

Enumerate nodes in the order

$A, \underline{B}, \underline{C}, D, E, F$





$P(A)\, P(B/A)\, P(C/A,B)\, P(D/A,B,C) \ldots$

$P(D/A,B,C) = P(D/B,C)$

innovate achieve lead

Reasoning

A is the first node in the ordering → it has no parents

B can only have A as its parent

Consider C :

We can consider only A as C's parent

Is C independent of B given A? → No! There is a path C, E, F, D, B

Reasoning

So we add an edge B → C

Consider node D:

- We must have B as a parent of D
- Is D independent of C given B? No
- So we add C as a parent to D.
- Now D is independent of A given B and C

Finally E's parents must be C and D.

$P(x_1, x_2, x_3) = P(x_3/x_1, x_2)\, P(x_2/x_1)\, P(x_1)$

Why does this procedure give a minimal I-map for P?

. We are given distribution P. Pick a particular ordering of variables

- We construct a graph G such that each node X_i has as parents a subset U of $\{X_1, X_2, \ldots X_{i-1}\}$ such that $(X_i \perp X_1, X_2, \ldots X_{i-1} \mid U)$

- This ensures that $P(X_i \mid X_1, X_2, \ldots X_{i-1}) = P(X_i / U)$

- Thus P factorizes over G [Theorem 3.1 Daphne Koller's book]

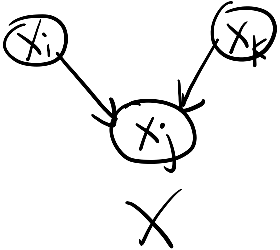
Why does this procedure work?

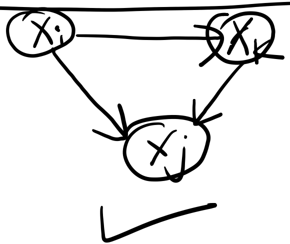
Theorem 3.2 then says that if P factorizes over G then G is an I-Map for P.

G is <u>minimal</u> by construction, removing a single edge will cause it to have some independency that does not belong to P.

Theorem

Let H be a Markov network structure and G be a Bayesian network minimal I-map for H. Then G can have <u>no immoralities</u>





innovate achieve lead

Proof Sketch

By contradiction

Assume that an immorality exists and X_i and X_k can both be parents of X_j without an edge between X_i and X_k

Proof Sketch

Since there is an edge $X_i \to X_j$ in G we conclude that X_i cannot be d-separated from X_j by all of X_j's other parents [Note we assume $i < k < j$]

H contains a path between X_i and X_j that is not cut off by any other parents of X_j. Similarly there exists a path between X_j and X_k

Proof Sketch

Let U be the parent set chosen for X_k.

Why was X_i $(i < k)$ not chosen as a parent

$\exists$ X_k leading to the immorality?

Since there are one or more paths from X_i to X_k

via X_j, all these paths are cut by U

U can separate X_i from X_j or X_j from X_k

Proof Sketch

Let U separate X_j from X_K.

Consider the choice of parent set for X_j. This is a minimal subset of $X_1, X_2, \ldots X_{j-1}$ which separates X_j from other nodes

Since U separates X_K from X_j, X_K cannot be the first node encountered on some uncut path from $X_j \Rightarrow X_K$ cannot be adjacent to X_j

$\Rightarrow$ Contradiction

Questions

1. Given a Markov Network, find the appropriate factorization of joint distribution.
2. Given a Markov Network, identify the active trails.
3. Given a Markov Network, identify the I-maps.
4. Given a Markov Network, identify the d-separations.
5. Given a toy application, generate Markov Network and the factors associated with it.
6. Given factors, generate a Markov Network.
7. Given a joint distribution in the factorized form, generate a Markov Network.
8. Given a Markov Network, identify the conditional 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 !!!