Probabilistic Graphical Models

Graph separation $\Longleftrightarrow$ $I(G) = \bigcap_{P \text{ factorizes as per } G} I(P)$

$\Uparrow$

Graph $\Rightarrow$ Factorization $\Rightarrow$ conditional independence

G P $I(P)$

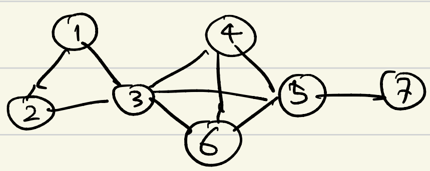
* Undirected Graphical Models

Def.: Undirected Graph $G = (V, E)$

$\hookrightarrow (i,j) = (j,i)$

- a set of nodes $C = \{i_1, i_2, \cdots, i_k\}$ is a clique in $G=(V,E)$ if all pairs in C are connected in G.

- a clique C is a maximal clique if you cannot add any node to C and make a bigger clique



$\{1,2,3\}, \{3,4,5,6\}, \{5,7\}$ are maximal cliques

- let $\mathcal{C}$ denote the set of all maximal cliques in G.

Undirected Graphical Model on $G=(V,E)$ is a family of distributions on $X=[x_1 \cdots x_n]$ that factorizes as

$$P(x_1 \cdots x_n) = \frac{1}{Z} \prod_{c \in \mathcal{C}} f_c(x_c)$$

where $\mathcal{C}$ is the set of maximal cliques in G,

$f_c : \mathcal{X}^{|c|} \to \mathbb{R}_+$ are non-negative function called factors or compatibility functions

$Z \triangleq \sum_{x \in \mathcal{X}^n} \left[\prod_{c \in \mathcal{C}} f_c(x_c) \right]$ is the normalization to ensure the probability sums to one, called Partition function.

Warm up examples >

G:

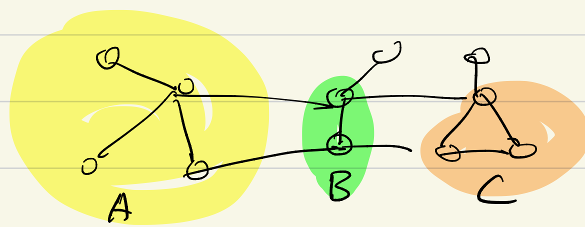
Factorization $P(X_1,X_2,X_3)$ $\quad P(x)=\frac{1}{Z}f_{12}(x_1,x_2)f_{13}(x_1,x_3)$ $\quad P(x)=\frac{1}{Z}f_{12}(x_1,x_2)f_3(x_3)$ $\quad P(x)=f_1(x_1)f_2(x_2)f_3(x_3)$

$\Downarrow$ HW1

$X_2 \perp\!\!\!\perp X_3 \mid X_1$ $\qquad (X_1,X_2) \perp\!\!\!\perp X_3$ $\qquad X_1, X_2, X_3$ mutually independent.

Def. [Graph Separation] $\qquad$ *Casually we will write it as A–B–C.

$B\subset V$ separates $A, C\subset V$ if every path from any node $a\in A$ to $c\in C$ passes through at least one node in B.



Def. [Global Markov Property (G)]

A distribution $P(X)$ satisfies the global Markov property w.r.t. G if for any $A, B, C\subset V$ s.t. A & C are separated by B,

$$X_A \perp\!\!\!\perp X_C \mid X_B$$

we will show that for all $p(x)$ that factorizes according to G, it also satisfies global Markov Property.

Q. give an example of Bayesian Network that cannot be represented by a MRF.

Q. give an example of MRF that cannot be represented by a BN.

Def. [local Markov Property (L)] w.r.t G

$P(x)$ satisfies local Markov Property if

$$X_i \perp\!\!\!\perp X_{rest} \mid X_{\partial i}$$

where $\partial i \triangleq$ neighborhood of $i = \{ j \in V \mid (i,j) \in E \}$

and $rest = V \setminus (\{i\} \cup \{\partial i\})$

Def [Pairwise Markov Property (P)] w.r.t G

$P(x)$ satisfies pairwise Markov Property if

$X_i \perp\!\!\!\perp X_j \mid X_{rest}$ for all i, j not connected in G.

where $rest = V \setminus \{i, j\}$

Claim: $(G) \Longrightarrow (L) \Longrightarrow (P)$

proof 1: $(G) \Rightarrow (L)$

set $A = \{i\}$, $B = \partial i$, $C = rest$

proof 2: $(L) \Rightarrow (P)$

$$X_i \perp\!\!\!\perp X_{V \setminus (\{i\} \cup \partial i)} \mid X_{\partial i}$$

$$\underset{j}{\Psi}$$

$$\overset{(a)}{\Longrightarrow} X_i \perp\!\!\!\perp X_{V \setminus (\{i\} \cup \partial i)} \mid X_{V \setminus \{i,j\}}$$

$$\overset{(b)}{\Longrightarrow} X_i \perp\!\!\!\perp X_j \mid X_{V \setminus \{i,j\}}.$$

(a) follows from $X \perp\!\!\!\perp Y \mid Z \Rightarrow X \perp\!\!\!\perp Y \mid (Z, h(Y))$

(b) follows from $X \perp\!\!\!\perp Y \mid Z \Rightarrow X \perp\!\!\!\perp h(Y) \mid Z$

claim: (P) $\Rightarrow$ (G) if $P(x) > 0$, $\forall x$

Lemma [Intersection Lemma]

For $P(x) > 0$, if $X_A \perp\!\!\!\perp X_B \mid (X_C, X_D)$ & $X_A \perp\!\!\!\perp X_C \mid (X_B, X_D)$
then $X_A \perp (X_B, X_C) \mid X_D$.

proof of claim:

suppose " $X_i \perp\!\!\!\perp X_j \mid X_{rest}$ ", then $X_A \perp\!\!\!\perp X_C \mid X_B$.

we show this by induction on all B of size $|B| = n-2, n-3, \cdots$

① initial condition: $|B| = n-2$.

true by assumption.

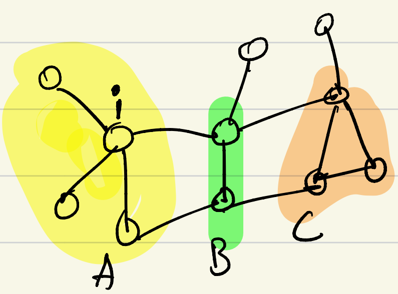
② suppose by induction hypothesis that (G) holds for all $|B| \geq S$.

we will prove (G) for a $|B| = S-1 < n-2$

either $|A| \geq 2$ or $|C| \geq 2$, so w.l.o.g let $|A| \geq 2$.

for any $i \in A$, we have from $A - B - C$ separation

$A \setminus \{i\} - B \cup \{i\} - C$ separation



$$\Downarrow$$

$$X_C \perp\!\!\!\perp X_{A \setminus \{i\}} \mid (X_B, X_i) \quad (1)$$

$\{i\} - B \cup A \setminus \{i\} - C$ separation

$$\Downarrow$$

$$X_C \perp\!\!\!\perp X_i \mid (X_B, X_{A \setminus \{i\}}) \quad (2)$$

Applying intersection lemma to (1) & (2) we get

$$X_C \perp\!\!\!\perp X_A \mid X_B$$

By induction (G) holds for all B. s.t. $A - B - C$.

* If $P \not> 0$, then (G) $\not\Rightarrow$ (P). as we show in HW1.

Def. [Factorization (F)]

we say $P(x)$ factorizes accordly to G if

$$P(x) = \frac{1}{Z} \prod_{c \in \mathcal{C}} f_c(X_c)$$

where $\mathcal{C}$ is the set of maximal cliques in G.

Claim. (F) $\Rightarrow$ (G).

proof. for any $A-B-C$ separated in G

there is no clique that includes $a \in A$ and $c \in C$, hence.

$$P(x) = \frac{1}{Z} F_1(X_A, X_B) \cdot F_2(X_B, X_C)$$

by HW1. this implies

$$X_A \perp\!\!\!\perp X_C \mid X_B.$$

Theorem. [Hammerseley-Clifford theorem]

If $P(x) > 0$ for all x, then (G) $\Rightarrow$ (F).

Implication: for $P(x) > 0$, (F) $\Leftrightarrow$ (G) $\Leftrightarrow$ (L) $\Leftrightarrow$ (P)

Counter example: If $P(x) \not> 0$, then (G) $\not\Rightarrow$ (F) (HW1).