

today: finish DGM
UGM

more properties of DGM

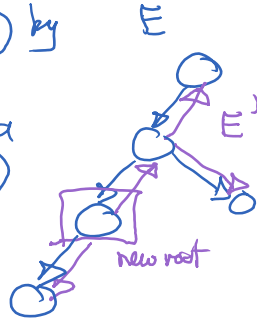
inclusion: $E \subseteq E' \Rightarrow \mathcal{J}(G) \subseteq \mathcal{J}(G')$

reversal: if G is a directed tree (or a forest)
 $\hookrightarrow \forall i: \text{in}_i \leq 2 \text{ out}_i$ ie. there is no v-structure

let E' be another directed tree
 (with same undirected edges) by
 choose a different root

(reverse some edges, while keeping a
 directed tree)

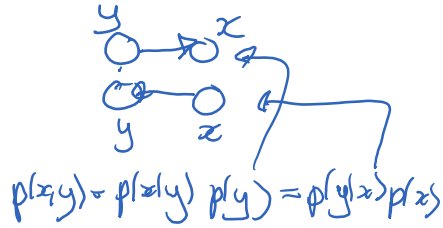
$\Rightarrow \mathcal{J}(G) = \mathcal{J}(G')$



rephrasing:

all directed trees from an
 undirected tree give
same DGM

⊗ why direction of edges
are not causal in DGM



marginalization

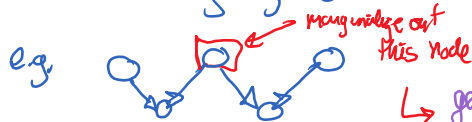
- marginalizing a leaf node n gives us a smaller DGM

let $S = \{ q \text{ is a dist. on } x_{1:n-1} \text{ s.t. } q(x_{1:n-1}) = \sum_{x_n} p(x_{1:n}) \}$

for some $p \in \mathcal{J}(G)$

then $S = \mathcal{J}(G')$ where G' is G with leaf n plucked (removed)

- not true for all marginalizations



\hookrightarrow get a set of dist.
 \neq to any $\mathcal{J}(G')$ for any G'

ie. marginalization is not a "closed" operation on DGMs



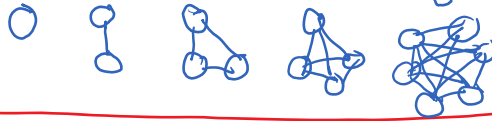
Undirected GM (UGM) (a.k.a. Markov random field or Markov network)

let $G = (V, E)$ be an undirected graph

let \mathcal{C} be the set of cliques of G

clique = fully connected set of nodes

$$(\text{if } C \in \mathcal{C} \Leftrightarrow \forall i, j \in C, \exists (i, j) \in E)$$



UGM associated with G

$$\mathcal{J}(G) \triangleq \left\{ p \text{ is a dist. over } \mathcal{X}_V : p(x_V) = \frac{1}{Z} \prod_{C \in \mathcal{C}} \psi_C(x_C) \right\}$$

for some "potentials" $\psi_C : \Omega_{x_C} \rightarrow \mathbb{R}_+$

and $Z \triangleq \sum_{x_V} \left(\prod_{C \in \mathcal{C}} \psi_C(x_C) \right)$
 normalization constant "partition function"

notes:

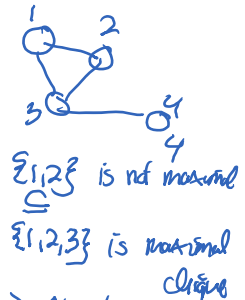
- unlike a DGM $\psi_C(x_C)$ is not directly related to $p(x_C)$ (where could think of C to be (i, π_i) $\psi_C(x_C) = p(x_i | x_{\pi_i})$)

- can rescale any potential without changing joint p i.e. $\psi_C^{\text{new}}(x_C) \triangleq \psi_C^{\text{old}}(x_C) \cdot A$ for some $A > 0$
 (unlike in DGM, where ξ_i 's were uniquely defined from p)

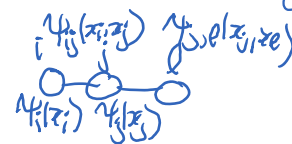
- it is sufficient to consider \mathcal{C}_{max} , the set of maximal cliques
 e.g. $C' \subseteq C$

redefine $\psi_C^{\text{new}}(x_C) \triangleq \psi_C^{\text{old}}(x_C) \cdot \psi_{C'}^{\text{old}}(x_{C'})$

(if C' belongs to more than one C , just choose one)



[we'll see later sometimes convenient to consider "loose parameterization" of trees using both $\psi_i(x_i)$ and $\psi_{ij}(x_i, x_j)$
 $\psi_{ij}(x_i, x_j)$
 \Downarrow
 $\{i, j\}$



properties of UGM:

- as before $E \subseteq E' \Rightarrow \mathcal{J}(G) \subseteq \mathcal{J}(G')$

$E = \emptyset \Rightarrow \mathcal{J}(G) = \text{set of fully factored dist.}$

$E = \text{all pairs (i.e. } V \text{ is just one big clique in } G) \Rightarrow \mathcal{J}(G) = \text{all distributions on } x_V$

- if $\psi_C(x_C) > 0 \forall x_C, \forall C$

then can write $p(x_V) = \exp\left(\sum_{C \in \mathcal{C}} \log \psi_C(x_C) - \log Z \right)$

$\langle \theta_C, T_C(x_C) \rangle$

where $T_C(x_C) = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \triangleq x_C$
 $\dim = |x_C|$

... .. $\phi(x_i) = \dots$ $\phi(x_j) = \dots$

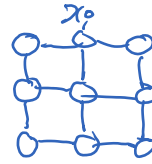
physic link: negative energy set.

$din = |d(x)|$

$G_C(x_C) = \log \Psi_C(x_C)$

e.g. Ising model in physics

$x_i \in \{0, 1\}$



node potential $E_i = \log \Psi_i(x_i = i)$

edge potential $E_{ij} = \log \Psi_{ij}(x_i = 1, x_j = 1)$

other: social network modeling

16h12

conditional independence for UGM:

def: we say that p satisfies global Markov property (with respect to a undirected graph G)

iff $\forall A, B, S \subseteq V$ s.t. S separates A from B in G
disjoint
 then $X_A \perp\!\!\!\perp X_B \mid X_S$



prop:

$p \in \mathcal{S}(G) \Rightarrow p$ satisfies the global Markov property for G

proof: WLOG, we assume $A \cup B \cup S = V$ & A separated from B by S

[why? if not, $\tilde{A} \triangleq A \cup \{a \in V: a \text{ and } A \text{ are not separated by } S\}$

$\tilde{B} \triangleq V \setminus (S \cup \tilde{A}) \Rightarrow \tilde{A} \cup \tilde{B} \cup S = V$

then if we have $X_{\tilde{A}} \perp\!\!\!\perp X_{\tilde{B}} \mid X_S$
 by decomposition $\Rightarrow X_A \perp\!\!\!\perp X_B \mid X_S$

$A \subseteq \tilde{A}$ $B \subseteq \tilde{B}$
 [exercise: show that if $B \in \tilde{B}$
 $\Rightarrow B$ is separated from \tilde{A} by S]

$V = A \cup S \cup B$

let $C \in \mathcal{E}$



cannot have both $C \cap A \neq \emptyset$ and $C \cap B \neq \emptyset$

(because of separation)

thus $p(x_C) = \prod_{\substack{C \in \mathcal{E} \\ C \subseteq A \cup S}} \pi_{C|C}(x_C) \prod_{\substack{C \in \mathcal{E} \\ C \subseteq B \cup S}} \pi_{C|C}(x_C) = f(x_{A \cup S}) g(x_{B \cup S})$
 $\Rightarrow C \subseteq B \cup S$

$$\Rightarrow C' \subseteq B \cup S$$

$$p(x_A | x_S) \propto p(x_A, x_S) = \sum_{x_B} p(x_V) = \sum_{x_B} f(x_{A \cup S}) g(x_{B \cup S})$$

$$= f(x_{A \cup S}) \underbrace{\sum_{x_B} g(x_{B \cup S})}_{\text{constant w.r. to } x_A}$$

$$\Rightarrow p(x_A | x_S) = \frac{f(x_{A \cup S})}{\sum_{x_A} f(x_A, x_S)}$$

similarly, $p(x_B | x_S) = \frac{g(x_{B \cup S})}{\sum_{x_B} g(x_B, x_S)}$

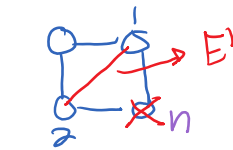
$$p(x_A | x_S) p(x_B | x_S) = \frac{f(x_{A \cup S}) g(x_{B \cup S})}{\sum_{x_A} f(x_A, x_S) \sum_{x_B} g(x_B, x_S)} \quad \left. \begin{array}{l} \} p(x_V) = p(x_A, x_B | x_S) \\ \} p(x_S) \end{array} \right\} \text{ic. } x_A \perp\!\!\!\perp x_B | x_S //$$

thm: (Hammersley-Clifford) if $p(x_V) > 0 \forall x_V$
 then $p \in \mathcal{L}(G) \iff p$ satisfies the global Markov property for G

proof: see ch. 16 of Mike's book; use "Möbius inversion formula" (as exclusion-inclusion principle in prob.)

more properties of UGM:

* closure with respect to marginalization



let $V' = V \setminus \{n\}$

$E' = \text{edges in } G \setminus \{n\}$

+ connect all neighbors of n in G together (now clique)

$$\sum_{x_n} \Psi_{2n}(x_2, x_n) \Psi_{1n}(x_1, x_n)$$

$\{ \text{marginal on } x_{1:n-1} \text{ for some } p \in \mathcal{L}(G) \} = \mathcal{L}(G')$