Lecture 13 - scribbles

Friday, October 18, 2019 13:57

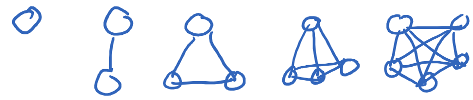
today: • undirected graph model
• inference

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

(i.e. $C\in\mathcal{C} \Leftrightarrow \forall i,j\in C,\ i\neq j,\ \{i,j\}\in E$)



UGM associated with G

$$\mathcal{L}(G) \triangleq \left\{ p : p(x_V) = \frac{1}{Z}\prod_{C\in\mathcal{C}} \psi_C(x_C) \right\}$$

for some "potentials" ψ_C s.t. $\psi_C(x_C)\geq 0\ \forall x_C$

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

$\sum_{x_V} \to \sum_{x_1}\sum_{x_2}\cdots\sum_{x_n}$

note: • unlike in DGM (where we could think of C to be (i,π_i), $\psi_C(x_C)=p(x_i|x_{\pi_i})$)

$\psi_C(x_C)$ is not directly related to $p(x_C)$

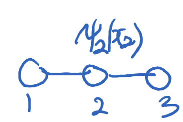
• can rescale any potential without changing p i.e. $\psi_C^{new}(x_C) \triangleq \psi_C^{old}(x_C)\cdot c'$, $c'>0$

• it is sufficient to consider $\mathcal{C}_{max}$, the set of maximal cliques

e.g. $C'\subseteq C$

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

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



[we'll see later it is convenient to]

consider "overparameterization" of trees using both $\psi_i(x_i)$ and $\psi_{ij}(x_i, x_j)$

properties:

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

  $E = \emptyset \Rightarrow \mathcal{L}(G) =$ fully factorized dist.

  $E =$ all pairs (i.e. G is just one big clique) $\quad \mathcal{L}(G) =$ all distributions

- if $\psi_C(x_C) > 0 \;\; \forall x_C$

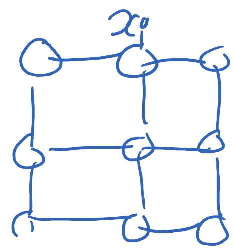
  can write $p(x_V) = \exp\Big(\underbrace{\sum_{C \in \mathcal{C}} \overbrace{\log \psi_C(x_C)}^{\langle \theta_C, T_C(x_C) \rangle}} - \log Z \Big)$

  physics link: negative energy function

e.g. Ising model in physics

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

node potential $\rightarrow E_i$
edge potential $\rightarrow E_{ij}$

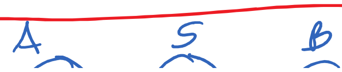


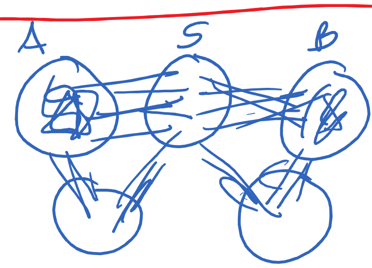
$E_i = \psi_i(x_i = 1)$
$E_{ij} = \psi_{ij}(x_i = 1, x_j = 1)$

other example: social network modeling

conditional independence for UGM

$\boxed{\begin{array}{l} \text{def: we say that } p \text{ satisfies global Markov property (with respect to an undirected graph } G) \\ \text{if } \forall A, B, S \subseteq V \text{ (disjoint) s.t. } S \text{ separates } A \text{ from } B \text{ in } G \\ \text{then } X_A \perp\!\!\!\perp X_B \mid X_S \end{array}}$





$a \in A$
$A \subseteq V$
$C \in \mathcal{C}$

prop: $\boxed{p \in \mathcal{L}(G) \Rightarrow p \text{ satisfies the global Markov property for } G}$

proof:

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

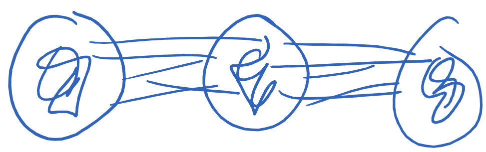
[why? if not, let $\tilde{A} \triangleq A \cup \{a \in V : a \text{ and } A \text{ are not separated by } S\}$

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

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

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

$V = A \cup S \cup B$



let $C \in \mathcal{C}$

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

thus $$p(x) = \frac{1}{Z} \prod_{\substack{C \in \mathcal{C} \\ C \subseteq A \cup S}} \psi_C(x_C) \prod_{\substack{C' \in \mathcal{C} \\ C' \nsubseteq A \cup S}} \psi_{C'}(x_{C'}) = f(x_{A \cup S})\, g(x_{B \cup S})$$

$\underbrace{C' \nsubseteq A \cup S}_{\Rightarrow C' \subseteq B \cup S \text{ and } C' \nsubseteq S}$

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

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

$$\Rightarrow p(x_A \mid x_S) = f(x_A, x_S)$$

constant with respect to x_A

$$\Rightarrow p(x_A | x_S) = \frac{f(x_A, x_S)}{\sum_{x_A'} f(x_A', x_S)}$$

similarly, $$p(x_B | x_S) = \frac{g(x_B, x_S)}{\sum_{x_B'} g(x_B', x_S)}$$

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

i.e. $X_A \perp\!\!\!\perp X_B \mid X_S$ //

thm: (Hammersley-Clifford) if $p(x_V) > 0 \;\forall x_V$

then $p \in \mathcal{L}(G) \Leftrightarrow p$ satisfies the global Markov prop. for G

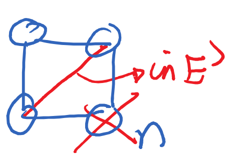
$\xLeftarrow{\text{H-C.}}$

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 $V' = V \setminus \{n\}$ $\quad E' =$ edges in $G \setminus \{n\}$
+ connect all neighbors of n in G together
(new clique)

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

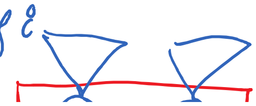
DGM vs UGM

def: Markov blanket for i (for graph G) is the smallest set of nodes M

s.t. $X_i \perp\!\!\!\perp X_{V} \mid X_M$

"rest"

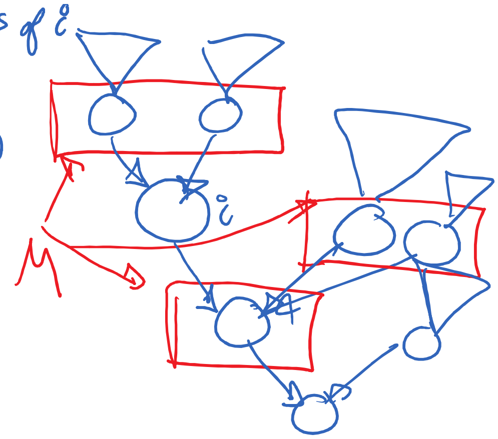
• for UGM: $M = \{j : \{i,j\} \in E\}$ = set of neighbors of i



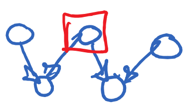
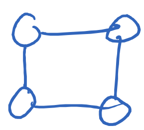
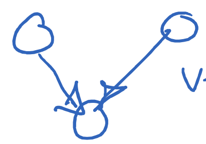
• for DGM: $M = $ [illegible]

• for UGM: $M = \{j : \{i,j\} \in E\}$ ← set of neighbors of i

• for DGM: $M = \underbrace{\pi_i \cup \text{children}(i)}_{\text{like UGM}} \cup \bigcup_{j \in \text{children}(i)} \pi_j$



<u>recap</u>

	DGM	UGM
factorization:	$p(x) = \prod_i p(x_i \mid x_{\pi_i})$	$\frac{1}{Z} \prod_C \psi_C(x_C)$
cond. indep.:	d-separation	separation
marginalization:	not closed in general eg  <u>but</u> is fine for <u>leaves</u>	closed (connect all neighbors of removed node)
cannot exactly capture some families:		 v-structure

<u>Moralization</u>:

let G be a DAG; when can we transform to equivalent UGM

<u>Def</u>: for G a DAG, we call $\bar{G}$ the <u>moralized graph</u> of G

where $\bar{G}$ is an undirected graph with same V

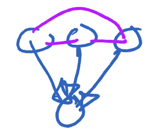
and $\bar{E} = \{\{i,j\} : (i,j) \in E\}$ } undirected version of G

$\cup \{\{k,l\} : k \neq l \in \pi_i \text{ for some } i\}$ } "moralization"

connect all the parents of i with i in big clique

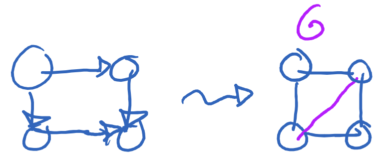
"marrying the parents" (???)

"connect all the parents of i with i in big clique" marrying the parents [???]



only needed when $|\pi_i| > 1$
i.e. v-structure

e.g.



prop: for a DAG G with no v-structure [forest]
then $\mathcal{L}(G) = \mathcal{L}(\bar{G})$
DGM UGM

but in general, can only say $\underset{\text{DAG}}{\mathcal{L}(G)} \subseteq \mathcal{L}(\bar{G})$

[note that $\bar{G}$ is the minimal undirected graph s.t. $\underset{\text{DAG}}{\mathcal{L}(G)} \subseteq \mathcal{L}(G')$]

15h27

general themes in this class

A) representation → DGM
→ UGM

parameterization → exponential family

} modeling high dim. dist.

B) inference $p(x_Q \mid x_E)$
"query" "evidence"
→ today's elimination alg.
→ next: sum-product / belief propagation (for trees)

C) statistical estimation / learning → MLE
maximum entropy
method of moments

Inference:

present alg. for UGM [for simplicity and more generality]

but note that sometimes more efficient to work with DGM directly

make DGM $\rightsquigarrow \subseteq$ UGM via moralization

i.e. DGM : $p(x) = \prod_i p(x_i | x_{\pi_i})$ moralize: $C = \{i\} \cup \pi_i$

$= \frac{1}{1} \prod_i \psi_C(x_C)$ $\psi_C(x_C) \triangleq p(x_i | x_{\pi_i})$

$Z = 1$

inference : want to compute

a) marginal : $p(x_F)$ for some $F \subseteq V$

b) conditional : $p(x_F | x_E)$

"query" "evidence"

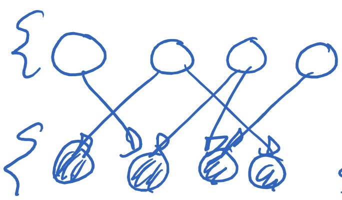
c) for UGM : partition function $Z = \sum_{x_V} \prod_{C \in \mathcal{C}} \psi_C(x_C)$

why?

- missing data $p(x_{\text{unobs}} | x_{\text{obs}})$
- prediction e.g. $p(x_{\text{future}} | x_{\text{past}})$
- "latent cause"
  e.g. $p(x_{\text{cause}} | x_{\text{obs}})$

e.g. F {

E {

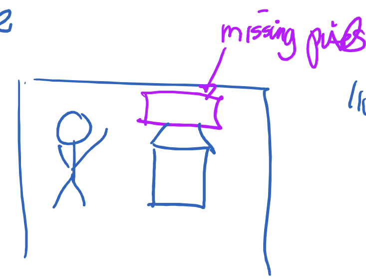


diseases

symptoms

in vision

image



"infilling task"

* also related to inference

$\arg\max_{x_F} p(x_F | x_E)$

F could be big here (e.g. speech recognition)

* inference is also needed during estimation (parameter fitting MLE)
[e.g. during E-step $p(z|x)$]