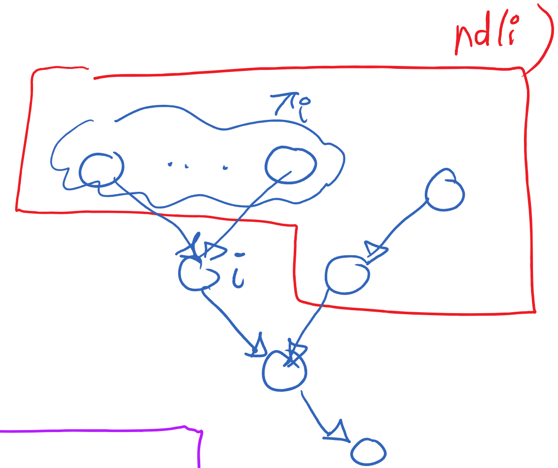
Lecture 12 - scribbles

Friday, October 12, 2018 13:34

today: finish DGM
start UGM



cond. indep. statements in DGM:

let $nd(i) \triangleq \{ j : \text{no path } i \rightsquigarrow j \}$

"non-descendants" of i

prop: $\boxed{p \in \mathcal{L}(G) \iff \underset{\forall i}{X_i \perp\!\!\!\perp X_{nd(i)} \mid X_{\pi_i}}}$

by decomposition $\Rightarrow X_i \perp\!\!\!\perp X_j \mid X_{\pi_i} \quad \forall j \in nd(i)$

proof:

$\Rightarrow$) key point: let i be fixed,

then $\exists$ a top. ordering s.t. $nd(i)$ are exactly before i

i.e. $(nd(i), i, \underbrace{\text{descendant}(i)}_{\text{pluck all these}})$

$\Rightarrow p(x_i, x_{nd(i)}) = p(x_i \mid x_{\pi_i}) \prod_{j \in nd(i)} p(x_j \mid x_{\pi_j})$

$$p(x_i \mid x_{nd(i)}) = \frac{p(x_i, x_{nd(i)})}{p(x_{nd(i)})} = \frac{p(x_i \mid x_{\pi_i}) \prod_{j \in nd(i)} \cancel{p(x_j \mid x_{\pi_j})}}{\underbrace{\left(\sum_{x_i} \cancel{p(x_i \mid x_{\pi_i})}\right)}_{1} \underbrace{\prod_{j} \cancel{p(x_j \mid x_{\pi_j})}}_{\text{no } x_i \text{ here (because } j \text{ is } nd(i) \Rightarrow i \notin \pi_j)}} = p(x_i \mid x_{\pi_i})$$

//

$\Leftarrow$) (suppose p satisfies cond. indep. statements)

let $1{:}n$ be a top. sort

then $1{:}i-1 \subseteq nd(i)$ $\forall i$ [say $j < i$ is a descendent of i $\Rightarrow$ back edge $\cancel{\exists}$ contradiction]

$\Rightarrow X_i \perp\!\!\!\perp X_{1:i-1} \mid X_{\pi_i}$ (by decomposition)

$$p(x_V) = \prod_{i=1}^{n} p(x_i \mid x_{1:i-1}) \quad \text{(by chain rule)}$$

this includes π_i

$p(x_i \mid x_{\pi_i}, x_{1:i-1}) = p(x_i \mid x_{\pi_i})$ by cond. indep.

$$= \prod_{i=1}^{n} p(x_i \mid x_{\pi_i}) \quad \text{(by C.I.)}$$

$\Rightarrow p \in \mathcal{L}(G)$

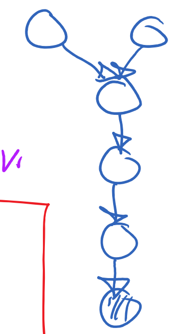
//

other C.I. statements?

chain : just any (undirected) path in graph from a to b

<u>chain from a to b</u> : just any (<u>undirected</u>) path in graph from a to b

<u>d-separation</u> : (intuition, is do not want this situation



def: sets A & B are said to be <u>d-separated by C</u>

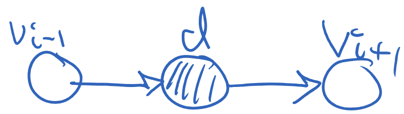
set of observed R.V./ conditioned R.V.

iff all chains from $a \in A$ to $b \in B$ are "blocked" given C

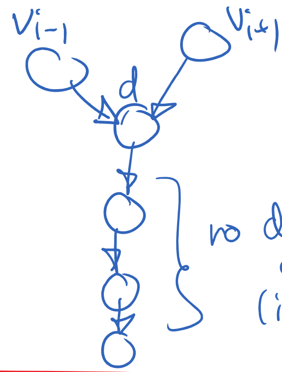
where a chain from a to b is "<u>blocked</u>" at node d

subpath of chain

if a) either $d \in C$ and (v_{i-1}, d, v_{i+1}) is <u>not</u> a v-structure

e.g.: 

b) $d \notin C$ and (v_{i-1}, d, v_{i+1}) is a v-structure



<u>and</u> no descendent of d is in C

no descendent of d in C (ie. no conditioning)

prop: if $p \in \mathcal{L}(G) \Longleftrightarrow X_A \perp\!\!\!\perp X_B \mid X_C$

prop. If $p \in \mathcal{L}(G) \iff X_A \perp\!\!\!\perp X_B \mid X_C$

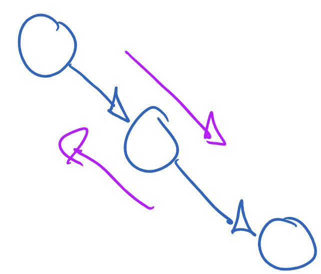
$\forall A,B,C$ s.t. A & B are d-separated by C

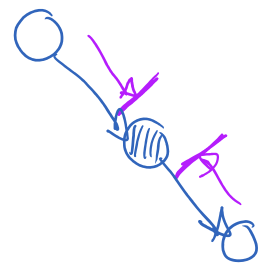
"Bayes-Ball" alg. : "intuitive" alg. to check d-separation

rules for balls/chains being blocked

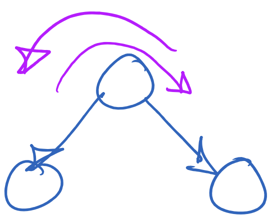
(a) in d-separation

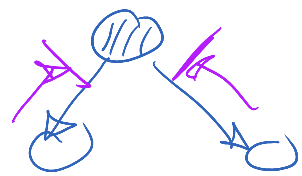
1)



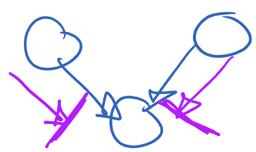


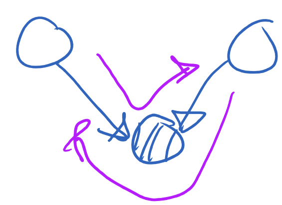
2)



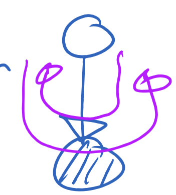


3) V-structure



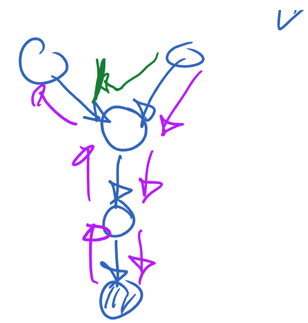


or



see Alg. 3.1

see Alg. 3.1
in K&F book



14h21

properties of DGM':

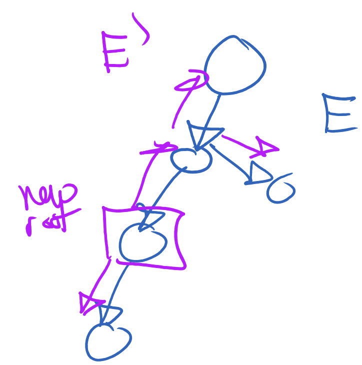
inclusion: $E \subseteq E' \Rightarrow \mathcal{L}(G) \subseteq \mathcal{L}(G')$

(adding edges increase # of distributions)

reversal: if G is a directed tree (or forest)

↳ ($|\pi_i| \leq 1$; i.e. there is no v-structure)

then let E' be another directed tree by choosing a different root
(reverse some edges, while keeping a directed tree)



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

↳ $[G' = (V, E')]$

⊛ this is why direction of edges are not causal

↳ $\mathcal{L}(G' = (V, E'))$ }

(*) this is why direction of edges are not causal

[e.g. $p(x,y) = p(x|y)p(y) = p(y|x)p(x)$]

$y \to x$ $\qquad$ $y \leftarrow x$

rephrasing: $\boxed{\text{all directed trees from a undirected tree give same DGM}}$

---

marginalizing:

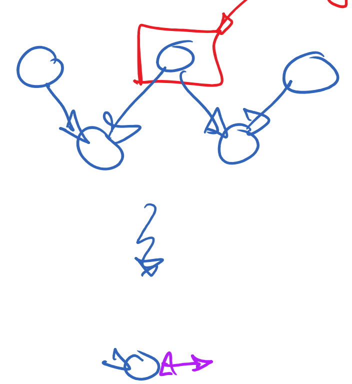
- marginalizing a leaf node n gives us a smaller DGM

  let $S = \{ q \text{ is dist on } x_{1:n-1} \text{ s.t. } q(x_{1:n-1}) = p(x_{1:n-1}) \text{ for } p \in \mathcal{L}(G) \}$

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

- not true for all marginalization

e.g.



marginalizing out this node

→ get a set of distributions $\neq$ to any $\mathcal{L}(G')$ for G'

i.e. marginalization is not a "closed operation" on DGM

14h47

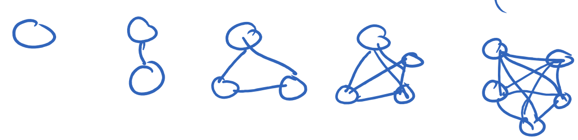
undirected GM (UGM) (a.k.a. Markov random field or Markov network)

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. $C \in \mathcal{C} \iff \forall i,j \in C,\ i\neq j,\ \{i,j\}\in E$)



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.

$$\psi_C(x_C) \geq 0 \quad \forall x_C$$

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

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

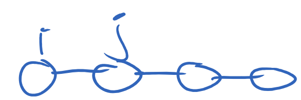
$\psi_C(x_C)$ is not directly related to $p(x_C)$

• can multiply any $\psi_C(\cdot)$ by a constant without changing p

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

e.g. $C' \subseteq C$ (choose one C when C' included in many C's)

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

[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.

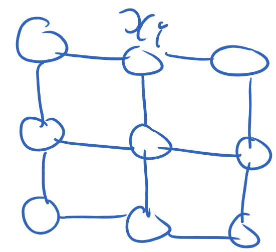
$E =$ all pairs (ie. G is one big clique) $\Rightarrow$ all distributions

- 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 potentials $\rightarrow E_i$

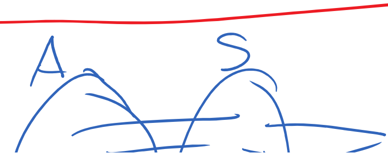
edge potentials $\rightarrow E_{i,j}$

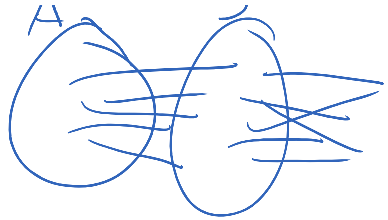
think also
social network modeling

conditional independence for UGM:

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

iff $\forall A, B, S \subseteq V$ s.t. S separates A from B in G
disjoint





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

proof: WLOG, we assume $A \cup B \cup S = V$

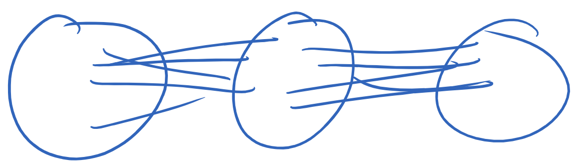
[why? if not, let $\tilde{A} \triangleq A \cup \{a \in V : a \text{ and } A \text{ are not separated by } S\}$]

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

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

[exercise: show that $b \in \tilde{B} \Rightarrow b$ 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 $$p(x) = \frac{1}{Z} \prod_{\substack{C \in \mathcal{C} \\ C \subseteq A \cup S}} \psi_C(x_C) \prod_{\substack{C' \in \mathcal{C} \\ \underbrace{C' \not\subseteq A \cup S}_{\Rightarrow C' \subseteq B \cup S}}} \psi_{C'}(x_{C'}) = f(x_{A \cup S})\, g(x_{B \cup S})$$

$$p(x_A \mid x_S) \propto p(\underbrace{x_A, x_S}_{x_{A\cup S}}) = \sum_{x_B}^{\text{and } C' \not\subseteq S} p(x_V) = \sum_{x_B} f(x_{A\cup S})\, g(x_{B\cup S})$$

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

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

similarly, $p(x_B \mid x_S) = \dfrac{g(x_B, x_S)}{\sum_{x'_B} g(x'_B, x_S)}$

$$p(x_A \mid x_S)\, p(x_B \mid x_S) = \frac{f(x_{A\cup S})\, g(x_{B\cup S}) \;\}\; p(x_V)}{\sum_{x'_A} \sum_{x'_B} f(x'_A, x_S)\, g(x'_B, x_S) \;\}\; p(x_S)} = p(x_A, x_B \mid x_S)$$

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

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

then $p \in \mathcal{L}(G) \Longleftrightarrow p$ satisfies global Markov property then $\quad p \in \mathcal{F}(G) \iff p$ satisfies global Markov property

(H.C.) $\Leftarrow$

<u>proof</u>: see ch. 16 of Mike's book; use "Möbius inversion formula"

(as exclusion-inclusion principle in proba)