Lecture 12 - DGM/ UGM

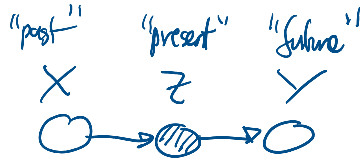
Thursday, October 12, 2023 2:32 PM

today:
- DGM (cont'ed)
- UGM

conditional independence in DGM

basic 3 nodes graphs

1) Markov chain



$p(x,y,z) = p(x)\,p(z|x)\,p(y|z)$

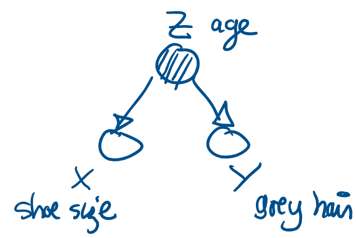
(exercise) $\Rightarrow p(x,y|z) = p(x|z)\,p(y|z)$

here $Y \perp\!\!\!\perp X \mid Z$ (future, past, present)

but $X \not\perp\!\!\!\perp Y$ (for some $p \in \mathcal{L}(G)$)

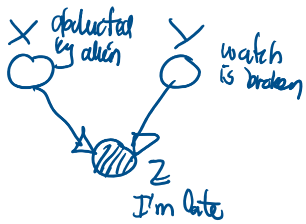
2) latent cause

(hidden variable) → Z is not observed



same statements as above

3) explaining away / competing effect (v structure)



$X \perp\!\!\!\perp Y$

but $X \not\perp\!\!\!\perp Y \mid Z$ (for some p)

non-monotonic property of conditioning:

p(alien) tiny

p(alien | late) > p(alien)

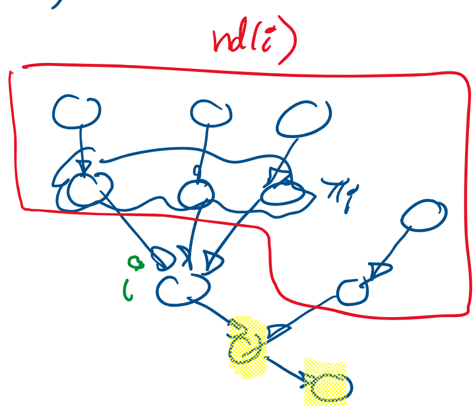
p(alien | late, watch broken) < p(alien | late)

more cond. indep statements in a DGM

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

"non-descendents" of i

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



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

proof:

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

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

i.e. $(nd(i), i, \underbrace{descendants(i)}_{\text{place all these?}})$

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

$$p(x_i \mid x_{nd(i)}) = \frac{p(x_i, nd(i))}{p(nd(i))} = \frac{p(x_i \mid x_{\pi_i}) \cancel{\prod_{j \in nd(i)} p(x_j \mid x_{\pi_j})}}{\left(\sum_{x_i} p(x_i \mid x_{\pi_i})\right) \underbrace{\cancel{\prod_j \; "}}_{\text{no } x_i \text{ here}}} = p(x_i \mid x_{\pi_i}) \quad //$$

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

let $1{:}n$ be (WLOG) a top. sort of G

then have $1{:}i{-}1 \subseteq nd(i) \;\; \forall i$ [by top sort property]

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

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

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

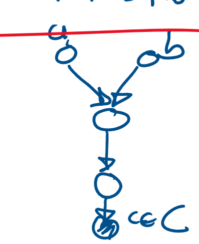
$$\Rightarrow p \in \mathcal{L}(G) \quad //$$

other cond. indep. statements?

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

d-separation:

intuition: do not want to have this situation

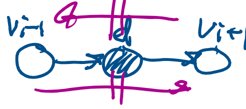
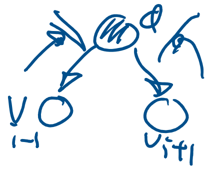


def: set A & B are said to be d-separated by C (in G) ← set of R.V. conditioned on

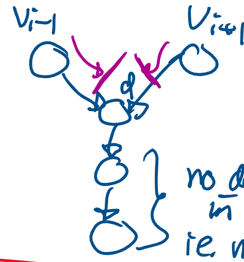
iff all chains from $a \in A$ to $b \in B$ are "blocked" given C

where a chain from a to b is "blocked" given C at a node d ← subpath of chain where a chain from a to b is "blocked" given C at a node d

if a) either $d \in C$ and (v_{i-1}, d, v_{i+1}) is not a v-structure
(subpath of chain)

(e.g.  or 

b) $d \notin C$ and (v_{i-1}, d, v_{i+1}) is a v-structure
and no descendent of d is in C



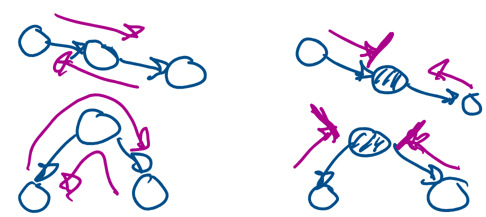
prop: $p \in \mathcal{L}(G) \Leftrightarrow X_A \perp\!\!\!\perp X_B \mid X_C \quad \forall A,B,C \subseteq V$
s.t. A & B are d-separated given C

this covers all cond. indep. statements which are true $\forall p \in \mathcal{L}(G)$

"Bayes ball" alg. "intuitive" alg to check d-separation
rules balls/chains being blocked

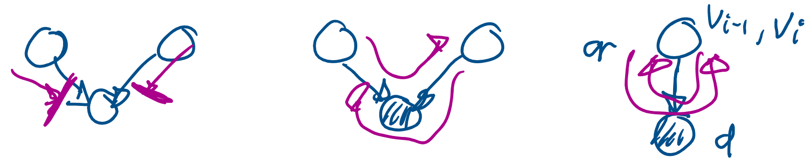
a) in d-separation

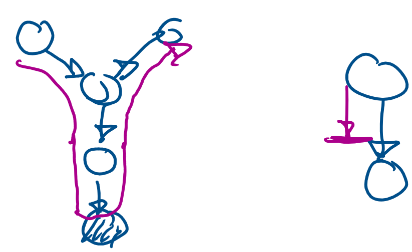
i) 2)



b)

3)





see Alg. 3.1 in K&F book

15h32

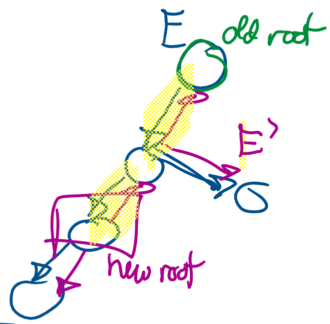
more properties of DGM

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

reversal: if G is a directed tree (or a forest) there is no v-structure

$\hookrightarrow |\pi_i| \leq 1 \ \forall i$

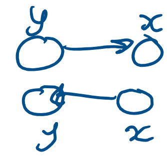
let E' be another directed tree (with same undirected edges) by choosing a different root



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

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

⊗ why direction of edges are not causal in DGM



marginalization:

- marginalizing a leaf node n gives a smaller DGM

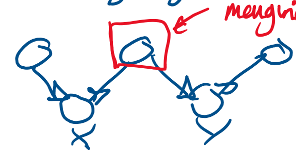
  let $S = \{ q \text{ is a dist. on } x_{1:n-1} \text{ s.t. } q(x_{1:n-1}) = \sum_{x_n} p(x_{1:n}) \}$

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

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

- not true for all marginalizations

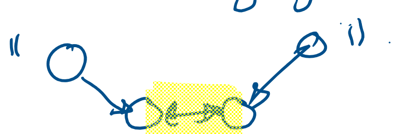
  e.g.

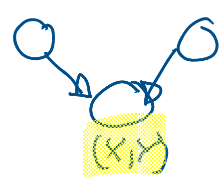


marginalize this node

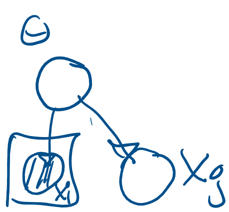
$\hookrightarrow$ get a set of dist. $\neq$ to any $\mathcal{L}(G')$ for any G'

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





← still bigger set unfortunately...



$$p(x_{new} \mid x_{1:n}) = \int_\theta p(x_{new} \mid \theta)\, p(\theta \mid x_{1:n})$$

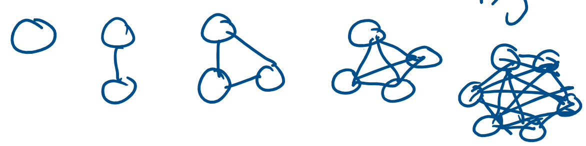
Undirected GM (UGM) (aka. Markov random field or Markov network)

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

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

clique = fully connected set of nodes

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



UGM associated with G

$$\mathcal{L}(G) \triangleq \{ p \text{ is a dist over } X_V : p(x_V) = \frac{1}{Z} \prod_{C \in \mathcal{C}} \psi_C(x_C) \}$$

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 (where 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 $p(x_C)$)

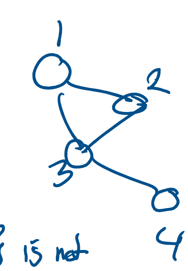
- can rescale any potential without changing joint p ie $\psi_C^{new}(x_C) \triangleq \psi_C^{old}(x_C) \cdot A$ for some $A > 0$
  (unlike in DGM, where f_i's were uniquely defined p)

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

  e.g. $C' \subseteq C$

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

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



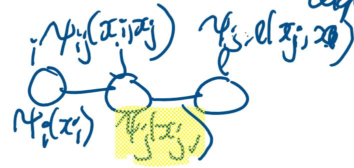
$\{1,2\}$ is not maximal
$\subseteq$
$\{1,2,3\}$ is maximal clique

$\psi_{ij}(x_i, x_j)$ $\psi_{ijl}(x_i, x_j, x_l)$

[we'll see later sometimes convenient to consider "over-parameterization" of trees {i,j} is maximal clique

[we'll see later sometimes convenient to consider "over-parameterization" of trees using $\psi_i(x_i)$

$\psi_{ij}(x_i, x_j)$ → $\{i,j\}$



properties of UGM:

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

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

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

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

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

  physics link: negative energy fct.

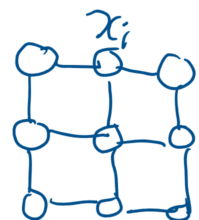
  $T_C(x_C) = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \\ 0 \\ 0 \end{pmatrix}$ ← "x_C"

  $\Theta_C(x_C) = \log \psi_C(x_C)$

  $\dim T_C = |\Omega_{X_C}|$

e.g. Ising model in physics

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



node potential $E_i = \log \psi_i(x_i = 1)$

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

other: social network modeling