Lecture 11 - DGM

Tuesday, October 10, 2023 2:28 PM

today: · graph theory
· DGM

graphical model

graph. model ⇝ prob theory + C.S.

prob theory ↓ R.V.

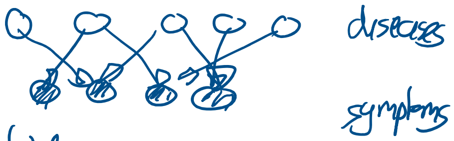
C.S. ↓ graph

graph → efficient data structure

e.g. $X_1, X_2, \dots, X_n$ R.V. QMR

$X_i \in \{0,1\}$ n = 100s

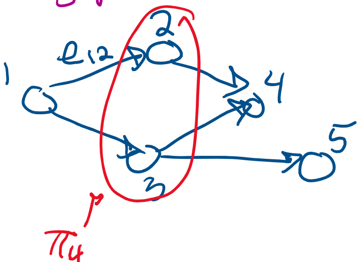
$\Rightarrow 2^{100s}$ #'s table
⇝ intractable



graph theory review:

directed graph $G = (V, E)$ $V = \{1, \dots, n\}$ "nodes/vertices"

"digraph" $E \subseteq V \times V$ "directed edges"



$e_{12} = (1,2)$

directed path $1 \rightsquigarrow 4$ compatible seq. of edges

or (1,2), (2,4)
(1,3), (3,4)

$\pi_i \triangleq \{j \in V : \exists (j,i) \in E\}$

set of parents of i

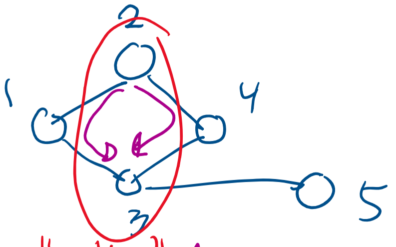
children(i) $\triangleq \{j \in V : \exists (i,j) \in E\}$

(note: no self loop in this class i.e. $|e| = 2$)

undirected graph: $G = (V, E)$ where elements of E are 2-sets
(set of 2 different elements of V)

thus we have $\{i,j\} = \{j,i\}$

vs $(i,j) \neq (j,i)$ [order matters]



"neighbors" of node of 4 or 1

undirected path 2 and 3

$N(i)$ for neighbors of i
$\triangleq \{j \in V : \exists \{i,j\} \in E\}$

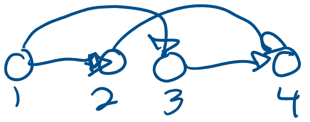
⊗ neighbors replace the parents/children terminology from digraph

def: DAG = directed acyclic graph = digraph with no cycles

def: an ordering $I : V \to \{1, \ldots, |V|\}$ is said to be topological for digraph G
$\overset{}{\text{1:1}}$

iff nodes in π_i appear before i in I $\forall i$

$(j \in \pi_i \Rightarrow I(j) < I(i))$



→ if top. ordering ⇒ all edges go from left to right
[no "back edge"]

prop: digraph G is a DAG ⇔ ∃ a topological ordering of G

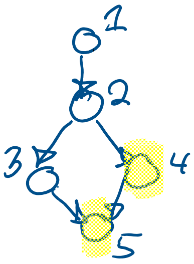
proof: ⇐) trivial: no back edge ⇒ no cycle

⇒) use DFS algorithm to label nodes in decreasing order when have no children

to construct a top. sort.

in $O(|E| + |V|)$

top. sort → finding a topological ordering



notation for graphical model

n discrete R.V. $X_1, \ldots, X_n$

↳ discrete R.V. for simplicity

cond. dist. concept is tricky to formalize for cts R.V.
(see Kolmogorov ~ Kolmogorov paradox

V ← set of vertices
one R.V. per node

shorthand

pencilbox

$V \leftarrow$ set of vertices
one R.V. per node

shorthand

joint $p(X_1=x_1, X_2=x_2, \ldots, X_n=x_n) \overset{\downarrow}{=} p(x_1, \ldots, x_n)$ $\qquad x = x_{1:n}$

$$= p(x_V) \overset{(!)}{=} p(x)$$

for any $A \subseteq V$

$$p(x_A) = P\{X_i = x_i : i \in A\} = \sum_{x_{A^c}} p(x_A, x_{A^c})$$

subset of "subscripts"

summing over all possible values of x_{A^c}
ie $(x_i)_{i \in A^c \triangleq V \setminus A}$

question: is $p(x_1, x_2, x_4) \overset{?}{=} p(x_2, x_1, x_4)$?

yes! usually in this class, we "typed conventions"

revisit cond. indep

let $A, B, C \subseteq V$

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

(F) factorization $\iff p(x_A, x_B \mid x_C) = p(x_A \mid x_C)\, p(x_B \mid x_C)$ $\quad \forall x_A, x_B, x_C$ s.t. $p(x_C) > 0$

(C) conditioning $\quad p(x_A \mid x_B, x_C) = p(x_A \mid x_C)$ $\quad \forall x_A, x_B, x_C$ s.t. $p(x_B, x_C) > 0$

$\circledast$ "marginal independence": $X_A \perp\!\!\!\perp X_B \quad (\mid \emptyset$ or $X_\emptyset)$

$$p(x_A, x_B) = p(x_A)\, p(x_B)$$

2 facts about cond. indep:

1) can repeat variables in statement (for convenience) $\qquad X \perp\!\!\!\perp \underbrace{Y, Z}_{\text{does not do anything}} \mid Z, W$ is fine to say $\qquad \swarrow X \perp\!\!\!\perp Y \mid Z, W$

2) decomposition: $X \perp\!\!\!\perp (Y, Z) \mid W \Rightarrow X \perp\!\!\!\perp Y \mid W$ & $X \perp\!\!\!\perp Z \mid W$

$\circledast$ pairwise indep. $\not\Rightarrow$ mutual indep. → see lecture (3)?

$$Z = X \underset{\text{xor}}{\oplus} Y$$

$\circledast$ chain rule (always true) $p(x_V) = \prod_{i=1}^{n} p(x_i \mid x_{1:i-1})$ ← last cond.

$p(x_n \mid x_{1:n-1})$
table with 2^n entries

assumption in DGM

$$p(x_V) = \prod_{i=1}^{n} p(x_i \mid x_{\pi_i}) \rightarrow \text{tables of } 2^{\max_i |\pi_i| + 1}$$

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

15h30

<u>directed graph. model</u>

let $G = (V, E)$ be a DAG

a <u>directed graphical model (DGM)</u> (associated with G) (aka a Bayesian network)
is a <u>family</u> of distributions over X_V

$$\mathcal{L}(G) = \left\{ p \text{ is a dist. over } X_V : \exists \text{ legal factors } f_i\text{'s st. } p(x_V) = \prod_{i=1}^{n} f_i(x_i \mid x_{\pi_i}) \ \forall x \right\}$$

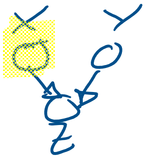
→ s.t. $f_i : \Omega_{X_i} \times \Omega_{X_{\pi_i}} \rightarrow [0,1]$

st. $\underbrace{\sum_{x_i} f_i(x_i \mid x_{\pi_i}) = 1}$ $\forall x_{\pi_i}$

$\Rightarrow f_i$ is like a CPT (conditional prob. table)

<u>terminology</u>: if we can write $p(x_V) = \prod_i f_i(x_i \mid x_{\pi_i})$ → this determines G
then we say that "<u>p factorizes according to G</u>"

<u>one example</u>:



$p \in \mathcal{L}(G) \Leftrightarrow \exists f_x, f_y, f_z$ legal
st. $p(x, y, z) = f_x(x) f_y(y) f_z(z \mid x, y)$

<u>"leaf plucking" property</u> (fundamental property of DGM)

"leaf plucking" property (fundamental property of DGM)

let $p \in \mathcal{L}(G)$; let n be a leaf in G (ie. n has no children)

a) then $p(x_{1:n-1}) \in \mathcal{L}(G - \{n\})$

means remove node n from G

b) moreover, if $p(x_{1:n}) = \prod_{i=1}^{n} f_i(x_i | x_{\pi_i})$ then $p(x_{1:n-1}) = \prod_{i=1}^{n-1} f_i(x_i|x_{\pi_i})$

proof:

$$p(x_n, x_{1:n-1}) = f_n(x_n|x_{\pi_n}) \prod_{i=1}^{n-1} f_i(x_i|x_{\pi_i})$$

no n in any π_i

$$p(x_{1:n-1}) = \sum_{x_n} p(x_n, x_{1:n-1}) = \Big(\underbrace{\sum_{x_n} f_n(x_n|x_{\pi_n})}_{1 \text{ by def'n}}\Big)\Big(\prod_{i=1}^{n-1} f_i(x_i|x_{\pi_i})\Big) \quad //$$

proposition: if $p \in \mathcal{L}(G)$

let $\{f_i\}$ be a factorization w.r.t. G of p

then $\forall i$, $p(x_i|x_{\pi_i}) = f_i(x_i|x_{\pi_i})$ $\forall x_{\pi_i}$

i.e. factors are correct conditionals of p
(and are unique)

proof: WLOG, let $(1, \dots, n)$ be a top. sorting of G (possible since G is a DAG)

- let i be fixed; we want to show that $p(x_i|x_{\pi_i}) = f_i(x_i|x_{\pi_i})$
  - n is a leaf → "pluck it" to get $p(x_{1:n-1}) = \underbrace{\prod_{j=1}^{n-1} f_j(x_j|x_{\pi_j})}_{\text{factorization for a DAG where } x_{n-1} \text{ is now a leaf}}$
  - pluck $n-1$ as well? (as leaf, because of top-sort)
  - keep doing this [in reverse order $n-1, n-2, \dots, i+1$] to get

$$p(x_{1:i}) = f_i(x_i|x_{\pi_i}) \underbrace{\prod_{j<i} f_j(x_j|x_{\pi_j})}_{\triangleq\, g(x_{1:i-1})}$$

i.e. there is no x_i by top. sort. property partition $1{:}i$ as $\{i\} \cup \pi_i \cup A$ $[A \triangleq 1{:}i-1 \setminus \pi_i]$

$$p(x_{1:i}) = p(x_i, x_{\pi_i}, x_A)$$

$$p(x_i | x_{\pi_i}) = \frac{\sum_{x_A} p(x_i, x_{\pi_i}, x_A)}{\sum_{x_i'} \sum_{x_A} p(x_i', x_{\pi_i}, x_A)} = \frac{f_i(x_i|x_{\pi_i}) \cancel{\sum_{x_A} g(x_{1:i-1})}}{\underbrace{\left(\sum_{x_i'} f_i(x_i'|x_{\pi_i})\right)}_{1} \cancel{\sum_{x_A} g(x_{1:i-1})}} = f_i(x_i|x_{\pi_i}) \quad //$$

is there is no x_i' there

⊛ (now), we can safely write

$$\mathcal{L}(G) = \{p : p(x_V) = \prod_{i=1}^n p(x_i | x_{\pi_i})\}$$

properties of DGM:

adding edges ⇒ more distributions

i.e. $G=(V,E)$ and $G'=(V,E')$ with $E' \supseteq E$

then $\mathcal{L}(G) \subseteq \mathcal{L}(G')$

examples:

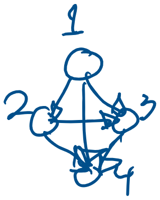
- $E = \emptyset$ (trivial graph) ⇒ $\mathcal{L}(G)$ contains only fully independent distributions

i.e. $p(x_V) = \prod_{i=1}^n p(x_i)$

$p(x_i | x_\emptyset)$

- complete DAG

for all i & j either $(i,j) \in E$ or $(j,i) \in E$



get WLOG

$$p(x_V) = \prod_{i=1}^n p(x_i | x_{1:i-1})$$

(like chain rule)

⇒ all dist. on X_V are in $\mathcal{L}(G_{complete})$

⊛ removing edges impose more factorization constraints (and thus more cond. indep. assumptions) "absence of edge" is meaningful in a DGM