Lecture 11 - DGM

Tuesday, October 11, 2022 3:20 PM

today : • graph theory
• DGM

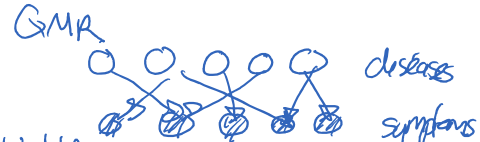
Graphical model

graph. model $\rightsquigarrow$ prob. theory + C.S.

prob. theory $\downarrow$ R.V.   C.S. $\downarrow$ graph

graph $\to$ efficient data structure

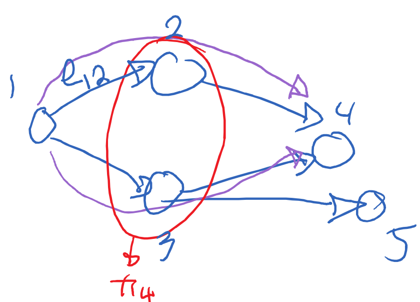
e.g. $X_1, \dots, X_n$ R.V. $n = 100$s
$X_i \in \{0,1\}$
$\Rightarrow 2^{100}$ #'s table
$\rightsquigarrow$ intractable



graph theory review :

directed graph "digraph" $G = (V, E)$

$V = \{1, \dots, n\}$ "nodes/vertices"
$E \subseteq V \times V$ "directed edges"
$e_{12} = (1,2)$

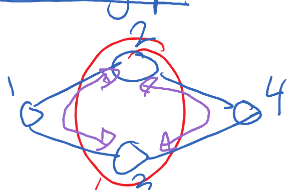


directed path $1 \rightsquigarrow 4$
$(1,2), (2,4)$ or $(1,3), (3,4)$ seq. of compatible edges

$\pi_i \triangleq \{j \in V : \exists (j,i) \in E\}$
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 (sets of 2 elements)
thus we have $\{i,j\} = \{j,i\}$
Vs: $(i,j) \neq (j,i)$ [order matters]



"neighbors" of node 4 or 1

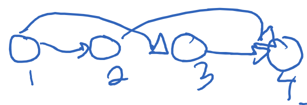
undirected path $2 \rightsquigarrow 3$

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

⊛ neighbors replace the parent/children terminology from a digraph def: DAG = directed acyclic graph = digraph with no cycles

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

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"]

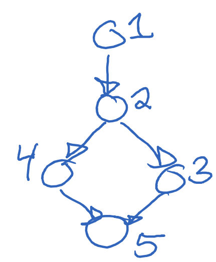
prop.: digraph G is a DAG ⇔ ∃ a topological ordering of G

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

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

to construct a top. sort ordering

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



top. sort → finding a topological ordering of graph

notation for graph models

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

↳ discrete R.V. for simplicity (cond. distribution concept is tricky to formalize for cts. R.V. (see Borel-Kolmogorov paradox))

V → set of vertices
one R.V. per node

joint $p(X_1 = x_1, \dots, X_n = x_n) \overset{\text{shorthand}}{=} p(x_1, \dots, x_n)$ $\quad x = x_{1:n}$

$= p(x_V) = p(x)$

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

subset of "subscripts"

summing over all possible values of x_{A^c} can take

i.e. $(x_i)_{i \in A^c = V \setminus A}$

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

yes? usually in this class, we used "typed convention"

<u>revisit cond. independence</u>:

let $A, B, C \subseteq V$

(*) $X_A \perp\!\!\!\perp X_B \mid X_C$

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

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

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

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

<u>2 facts about cond. indep.</u>:

1) can <u>repeat</u> variables in statement (for convenience) $X \perp\!\!\!\perp Y, \underset{\text{does not do anything}}{\underline{Z}} \mid \underline{Z}, W$ is fine to say

is equivalent $X \perp\!\!\!\perp Y \mid Z, W$

2) decomposition: $X \perp\!\!\!\perp (Y, Z) \mid W \Rightarrow X \perp\!\!\!\perp Y \mid W$ <u>and</u> $X \perp\!\!\!\perp Z \mid W$

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

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

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

<u>always true</u>

last conditional

$p(x_n \mid x_{1:n-1})$

→ table with 2^n entries

assumption in DGM

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

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

16h08

<u>Directed graph. model</u>

let $G = (V, E)$ be a <u>DAG</u>

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

a directed graphical model (DGM) (associated with G) (aka Bayesian network)

is a family of distributions over X_V

$$\mathcal{L}(G) = \{ p \text{ is a dist. over } X_V : \exists \text{ legal factors } f_i\text{'s st. } p(x_V) = \prod_i f_i(x_i \mid x_{\pi_i}) \ \forall x_V \}$$

st. $f_i : \Omega_{X_i} \times \Omega_{X_{\pi_i}} \to [0,1]$

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

→ f_i is like a local CPT (conditional probability table)

terminology: if we can write $p(x_V) = \prod_i f_i(x_i \mid x_{\pi_i})$ → this determines G

then we say that "p factorizes according to G"

one example:

$p \in \mathcal{L}(G) \Leftrightarrow \exists f_x, f_y, f_z$ legal

st. $p(x,y,z) = f_x(x) f_y(y) f_z(z \mid x,y)$

"leaf plucking" property (fundamental property of DGM)

let $p \in \mathcal{L}(G)$; let n be a leaf in G (i.e. 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 $p(x_{1:n}) = \prod_i f_i(x_i \mid x_{\pi_i})$ then $p(x_{1:n-1}) = \prod_{i=1}^{n-1} f_i(x_i \mid x_{\pi_i})$

proof: $p(x_n, x_{1:n-1}) = f_n(x_n \mid x_{\pi_n}) \prod_{i=1}^{n-1} (f_i(x_i \mid x_{\pi_i}))$ ← no n in any π_i

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

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

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

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

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 $\longrightarrow$ "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} is now a <u>leaf</u>

- pluck $n-1$ as well? (as leaf., because of top. sort.)

- keep doing this [in reverse order $n-1, n-2, n-3, \dots, i+1$] to get

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

no x_i in any π_j by top. sort property

partition: $1:i$ as $\{i\} \cup \{\pi_i\} \cup A$ $\quad [A \triangleq 1:i-1 \setminus \pi_i]$

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

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

there is no x_i

$\circledast$ (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:

$$\boxed{\begin{array}{l}\text{adding edges} \Rightarrow \text{more distributions} \\ \text{i.e. } G=(V,E) \text{ and } G'=(V,E') \text{ with } E' \supseteq E \\ \text{then } \mathcal{L}(G) \subseteq \mathcal{L}(G')\end{array}}$$

<u>examples</u>:

- $E = \emptyset$ (trivial graph) $\Rightarrow$ $\mathcal{L}(G)$ contains only <u>fully independent</u> distributions

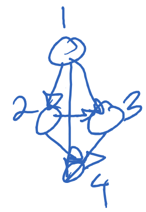
  i.e. $p(x_V) = \prod_i \underbrace{p(x_i)}_{p(x_i|\emptyset)}$

- complete digraph (DAG)

$p(x_i|\emptyset)$)

- complete digraph (DAG)

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



get WLOG $p(x_V) = \prod_i p(x_i | x_{1:i-1})$

(like chain rule)

$\Rightarrow$ <u>all dist. on x_V</u> are in $\mathcal{L}(G_{\text{complete}})$

(*) removing edges imposes <u>more</u> factorization constraints

(and thus <u>more</u> cond. indep. assumptions)