

today : • graph theory
• DGM

Graphical model

graph. model \rightarrow prob. theory + C.S.
 \downarrow \downarrow
 R.V. graph

graph \rightarrow efficient data structure

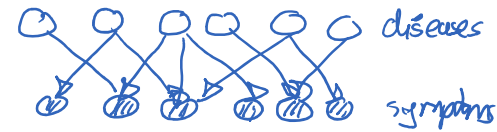
eg. X_1, \dots, X_n R.V.s

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

$n = 100$ s

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

GMR



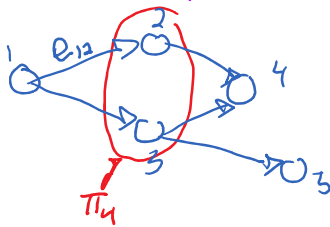
graph theory review:

directed graph
"digraph"

$G = (V, E)$

$V = \{1, \dots, n\}$ "nodes/vertices"

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



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

directed graph

$1 \rightarrow 4$
 $(1, 2), (2, 4)$
 or
 $(1, 3), (3, 4)$

\rightarrow compatible seq. of edges

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

set of parents of i

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

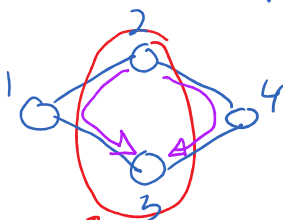
undirected graph : $G = (V, E)$ where elements of E are 2-sets

(set of 2 elements)

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

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

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



"neighbors" of node 4 or 1

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

undirected path $2 \rightarrow 3$

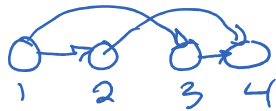
* neighbors \rightarrow replace the parents/children terminology from a digraph

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

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

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

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



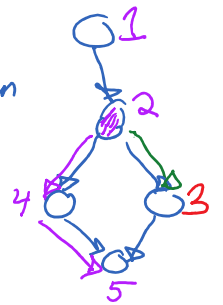
→ if top. ordering
 \Rightarrow all edges go from left to right
 [no "back edge"]

prop: digraph G is DAG $\Leftrightarrow \exists$ a topological ordering of G

proof: \Leftarrow) trivial: no back edge \Rightarrow no cycle

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

to construct a top. sort
 in $O(|E| + |V|)$



Notation for graph models:

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

↳ discrete R.V. for simplicity

conditional dist. concept is tricky to formalize

V ← set of vertices
 one ^{R.V.} variable per node

shorthand see Borel-kolmogorov paradox

$$\text{joint } p(x_1=x_1, \dots, x_n=x_n) \stackrel{\text{shorthand}}{=} p(x_1, \dots, x_n) \quad x = x_{i \in n}$$

$$= p(x_V) \stackrel{(\text{P})}{=} p(x)$$

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 $\{x_i\}_{i \in V \setminus A}$ can take x_{A^c}

$$x_{\{2,1,2,4\}} \Leftrightarrow \{x_1, x_2, x_4\}$$

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

yes! usually in this class, we use "topol convention"

revisit cond. indep.:

let $A, B, C \subseteq V$

⊛ $X_A \perp\!\!\!\perp X_B \mid X_C$

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

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

⊛ "marginal independence": $X_A \perp\!\!\!\perp X_B \mid \emptyset$
 or X_B
 $p(x_A, x_B) = p(x_A) p(x_B)$

2 facts about C.I.:

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

1) can repeat variables in statement $X \perp\!\!\!\perp Y, Z \mid Z, W$ is fine to say
 (for convenience)
 ↑
 does not do anything

2) decomposition: $X \perp\!\!\!\perp (Y, Z) \mid W \Rightarrow \begin{matrix} X \perp\!\!\!\perp Y \mid W \\ X \perp\!\!\!\perp Z \mid W \end{matrix}$

⊛ pairwise indep. $\not\Rightarrow$ mutual indep. \rightarrow see lecture 3

$Z = X \oplus Y$
 XOR

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

assumption in DGM

↖ last cond. $p(x_n \mid x_{1:n-1})$ ↗ table with $\max_i |\pi_i|$ entries $\approx 2^n$

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

15h38

Directed graph. model

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

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

is a family of distributions over X_V

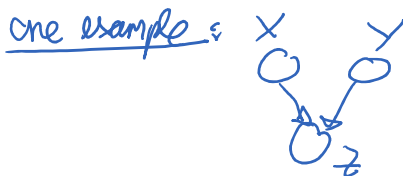
$$\mathcal{L}(\mathcal{G}) = \left\{ p \text{ is a dist. over } X_V : \exists \text{ legal factors } f_i \text{'s st. } p(x_V) = \prod_{i=1}^n f_i(x_i | x_{\pi_i}) \right\}$$

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

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

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

terminology: if we can write $p(x_V) = \prod_i f_i(x_i | x_{\pi_i})$ then we say that " p factors according to \mathcal{G} " \rightarrow this determines \mathcal{G}



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

"leaf plucking" property (fundamental property of DAG)

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

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

$$p(x_n, x_{1:n-1}) = f_n(x_n | x_{\pi_n}) \prod_{j \neq n} f_j(x_j | x_{\pi_j})$$

$$p(x_{1:n-1}) = \sum_{x_n} p(x_n, x_{1:n-1}) = \left(\sum_{x_n} f_n(x_n | x_{\pi_n}) \right) \left(\prod_{j \neq n} f_j(x_j | x_{\pi_j}) \right)$$

no n in any x_j
 by def

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

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

then $p(x_V) = \prod_i f_i(x_i | x_{\pi_i})$

Let $\{f_i\}$ be a factorization w.r.t. G for p

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

i.e. factors are correct conditions
(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 \rightarrow "pluck it" to get $p(x_{i:n-1}) = \prod_{j=1}^{n-1} f_j(x_j | x_{\pi_j})$

factorization for a DAG
where x_{n-1} is now a leaf

• pluck $n-1$ as well? (as leaf because of top order)

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

$$p(x_{i:i}) = f_i(x_i | x_{\pi_i}) \prod_{j=i}^{n-1} f_j(x_j | x_{\pi_j})$$

$$\triangleq g(x_{i:i-1})$$

i.e. no x_i by top. sort properly

partition $1:i$ as $\Sigma_i \cup \pi_i \cup A$ [$A \triangleq 1:(i-1) \setminus \pi_i$]

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

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

$$\sum_{x_i} \sum_{x_A} p(x_i, x_{\pi_i}, x_A)$$

$$\left(\sum_{x_i} f_i(x_i | x_{\pi_i}) \right) \left(\sum_{x_A} g(x_{i:i-1}) \right)$$

x_i

1

there is no x_i there

* (now) we can safely write

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

properties:

adding edges \Rightarrow more distributions

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

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

ie $G = (V, E)$ and $G' = (V, E')$ with $E \supseteq E'$
 then $\mathcal{L}(G) \subseteq \mathcal{L}(G')$

examples:

• $E = \emptyset$ (trivial graph) $\Rightarrow \mathcal{L}(G)$ contains only fully independent distributions
 ie. $p(x_v) = \prod_{i=1}^n p(x_i)$

• complete digraph (DAG)
 for all $i \neq j$ either $(i, j) \in E$
 or $(j, i) \in E$



get wlog

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

(like chain rule)

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

⊗ removing edges impose more factorization constraints
 (and thus ^{more} cond. indep. assumptions)