Lecture 11 - DGM

Thursday, October 10, 2024 2:33 PM

today:
- graph theory
- DGM

graphical model

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

$\downarrow$ R.V. $\quad\quad$ $\downarrow$ graph

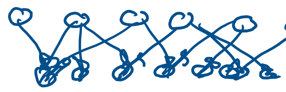
graph $\to$ efficient data structure

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

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

$\Rightarrow$ 2^{100} #'s table $\rightsquigarrow$ untractable

QMR



graph theory review:

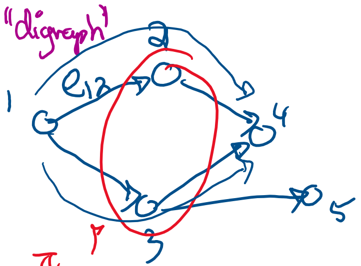
directed graph $G = (V, E)$

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

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

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

"digraph"



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

$(1,2), (2,4)$

or

$(1,3), (3,4)$

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

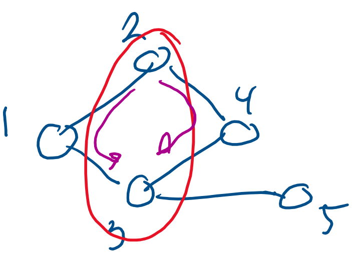
set of parents of i

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

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 4 or 1

undirected path $2 \leftrightsquigarrow 3$

$N(i)$ for neighbors of i

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

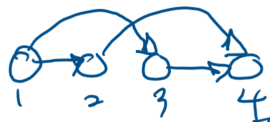
(*) neighbors replace the parent/children terminology for digraphs

def: DAG = directed acyclic graph $\triangleq$ digraph with no cycles

def: an ordering $I: V \to \{1, 2, \ldots, |V|\}$ ($|V| = n$) is said to be topological for digraph G

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

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

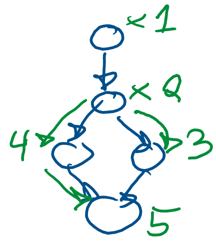


→ if top. ordering ⇒ all edges from left to right
[no "back edge"]

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

proof: $\Leftarrow$) trivial: no back edge ⇒ no cycle

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



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

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 formulate for cts. R.V.
(see Borel-Kolmogorov paradox)

V ← set of vertices
one R.V. per node

joint $p(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n) \overset{\text{shorthand}}{\triangleq} p(x_1, \ldots, x_n)$ $\quad x = x_{1:n}$

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

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

yes? usually in this class we "typed convention"

for any $A \subseteq V$

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

x_A ← subset of "subscripts"

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

e.g. $p(x_1, x_2, x_4) = p(x_{\{1,2,4\}})$

revisit cond. independence

let $A, B, C \subseteq V$

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

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

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

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

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

2 facts about cond. indep.

1) can repeat variable in statement (for convenience)

$X \perp\!\!\!\perp Y, \underset{\text{does not do anything}}{Z} \mid Z, W$ is fine to say

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

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

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

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

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

(always true)

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

assumption in DGM

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

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

15h24

directed graph. model

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

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

is a family of distribution over X_V

$\mathcal{L}(G) = \{ p$ is a dist. over X_V : $\exists$ legal factors f_i's s.t. $p(x_V) = \prod_{i}^{n} f_i(x_i \mid x_{\pi_i}) \;\forall x \}$

$\mathcal{L}(G) = \{ p$ is a dist. over X_V : $\exists$ legal f_i's s.t. $\}$

$$p(x_V) = \prod_{i=1}^{n} f_i(x_i \mid x_{\pi_i}) \quad \forall x$$

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

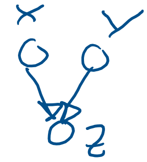
s.t. $\underbrace{\sum_{x_i} f_i(x_i \mid x_{\pi_i}) = 1}$ $\forall x_{\pi_i}$

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

<u>terminology</u>: if we can write $p(x_V) = \prod_i f_i(x_i \mid x_{\pi_i})$ $\rightarrow$ this defines G

then we say that "<u>p factorizes according to G</u>" & $p \in \mathcal{L}(G)$

<u>one example</u>



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

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

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

Let $p \in \mathcal{L}(G)$; let n be a <u>leaf</u> 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 $p(x_{1:n}) = \prod_{i=1}^{n} f_i(x_i \mid x_{\pi_i})$ then $p(x_{1:n-1}) = \prod_{i=1}^{n-1} f_i(x_i \mid x_{\pi_i})$

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

$$p(x_{1:n-1}) = \sum_{x_n} p(x_n, x_{1:n-1}) = \underbrace{\left[\sum_{x_n} f_n(x_n \mid x_{\pi_n})\right]}_{1 \text{ by defn}} \left(\prod_{i=1}^{n-1} f_i(x_i \mid x_{\pi_i})\right)$$

//

<u>proposition</u>: if $p \in \mathcal{L}(G)$

let $\{f_i\}$ be <u>a</u> 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 <u>unique</u>)

<u>proof</u>: 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 $p(x_i \mid x_{\pi_i}) = f_i(x_i \mid x_{\pi_i})$
  - n is a leaf $\rightarrow$ "pluck it" to get $p(x_{1:n-1}) = \prod^{n-1} f_i(x_i \mid x_{\pi_i})$

• 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 $p(x_{1:n-1}) = \underbrace{\prod_{j=1}^{n-1} f_j(x_j|x_{\pi_j})}$

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

• pluck $n-1$ as well? (as leaf, because of top. 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 $\{i\} \cup \pi_i \cup A$ $\quad [A \triangleq 1:i-1 \setminus \pi_i]$

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

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

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 $\Rightarrow$ more distributions

i.e. $\boxed{G=(V,E) \text{ and } G'=(V,E') \text{ with } E' \supseteq E \text{ then } \mathcal{L}(G) \subseteq \mathcal{L}(G')}$

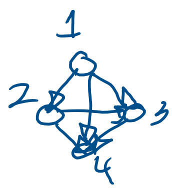
examples:

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

i.e. $p(x_V) = \prod_{i=1}^{n} p(x_i)$ $\quad \triangleq p(x_i|x_\emptyset)$

• complete DAG

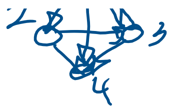
for all $i \neq j$ either $(i,j) \in E$ or $(j,i) \in E$



get WLOG

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

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



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

(like chain rule)

$\Rightarrow$ 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