Graphical model

Graph model $\sim$ prob. theory + CS, R.V. graph

Graph $\rightarrow$ efficient data structure

E.g., $X_1, \ldots, X_n$ R.V. $n=100$

$X_i \in \mathbb{S}^d$, $d \sim 100$, # is table $\sim$ interactive

Graph theory review:

Directed graph $G = (V, E)$

"digraph"

$V = \{v_1, \ldots, v_n\}$ "nodes/vertices"

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

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

Directed path: $1 \rightarrow 4$

$(1, 2), (2, 4)$ set of compatible edges on $(1, 2), (3, 4)$

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

set of parents of $i$

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

(Note: no self loop in the class $\Rightarrow |e_i| = 2$)

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

Thus we have $\exists i, j, i = 2 \Rightarrow i, j, j \neq (i, j)$ (order matters)

$\text{Unirected path} \quad 2 \rightarrow 3$

$N(i)$ for neighbors of $i$

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

Neighbors replace the parent/children terminology.
**def:** DAG = directed acyclic graph = digraph with no cycles

**def:** an ordering \( I: V \rightarrow \mathbb{E}_1, \ldots, \mathbb{E}_n \) is said to be topological for digraph \( G \)

iff nodes in \( \Pi_i \) appear before \( i \) in \( I \) i.e.

\[ I(j) < I(i) \]

\[ \rightarrow \text{if top. ordering } \Rightarrow \text{ all edges go from left to right} \]

[no “back edge”]

**prop.:** digraph \( G \) is a DAG \( \Leftrightarrow \exists \text{ a topological ordering of } G \)

**proof: 1)** trivial: no back edge \( \Rightarrow \) no cycle

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

[Note: 

To construct a top. sort ordering in \( O(\#E + \#V) \)

*top. sort* \( \Rightarrow \) finding a topological ordering of graph.

**notation for graph models**

\( \mathbb{X} = \text{discrete R.V. } X_1, \ldots, X_n \)

\( \text{to discrete R.V. for simplicity} \)

[comb. distribution concept is tricky to formulate for dir. r.v.

(see Benel-Kendall paper)]

\( V \rightarrow \) set of vertices

\( \text{one R.V. per node} \)

\( \text{joint} \quad p(X_1=x_1, \ldots, X_n=x_n) = p(x_1, \ldots, x_n) \)

\( \text{shortened} \quad p(x) \)

\( \text{shorthand} \quad p(x_V) = \prod p(x) \)

\( \text{for any } A \subseteq V \)

\[ p(x_A) = \sum_{x_i \in A} p(x) \quad \text{subject of subsets} \]

\( \text{it of possible values of } x_A \) can take

\( \text{the } x_i \text{ that } \{x_i \} \text{ is } A - C \text{ can take} \)

\( \text{question: is } p(x_1, x_2, x_4) \equiv p(x_2, x_1, x_4) ? \)
revisit cond. independence:

\( \text{let } A, B, C \subseteq \mathcal{X} \)

\( \Leftrightarrow \quad X_A \perp X_B \mid X_C \)

\( (F) \Leftrightarrow p(x_A, x_B | x_C) = p(x_A | x_C) p(x_B | x_C) \quad (\text{by } p(x_C) > 0) \)

\( (C) \Leftrightarrow p(x_A | x_B, x_C) = p(x_A | x_B) \quad (\text{by } p(x_B, x_C) > 0) \)

\( \text{"marginal independence": } X_A \perp X_B \mid \emptyset \)

\[ p(x_A, x_B) = p(x_A) p(x_B) \]

2 facts about cond. indep.:

1) Can repeat variables in statement for convenience:

\( X \perp Y \mid \emptyset \)

2) Decomposition:

\( X \perp (Y, Z) \mid W \Rightarrow X \perp Y \mid W \quad \text{and} \quad X \perp Z \mid W \)

\( \text{pairwise indep. } \not\Rightarrow \text{ mutual indep. } \rightarrow \text{ see lectures} \)

\( Z \equiv X \perp Y \)

\( \text{chain rule } \quad p(x_V) = \prod_{i=1}^{n} p(x_i | x_{1:i-1}) \quad \text{event conditional} \)

\( p(x_0 | x_{1:n-1}) \quad \text{always true} \)

\( \text{assumption in DBMs} \)

\[ p(x_V) = \prod_{i=1}^{n} p(x_i | x_{1:i-1}) \rightarrow \text{table of } \sum_{x_{1:n-1}}^{x_V} x \]

\[ X_i \perp X_{1:i-1} | X_{1:p} \]

16/08

Directed graph model:

\( \text{let } (V, E) \text{ be a DAG} \)
Let $G = (V, E)$ be a DAG.

A directed graphical model (DGM) (associated with $G$) (aka Bayesian network) is a family of distributions over $X$.

$$
\mathcal{G}(G) = \{ P \mid P \text{ is a dist. over } X \} \text{ s.t. } \exists \text{ legal factors } f_i \text{'s st: } \\
\mathbf{p}(\mathbf{x}) = \prod_{i} f_i(\mathbf{x}_i | \mathbf{x}_{\neg i}) \forall \mathbf{x}
$$

s.t. $f_i : \Omega_{x_i} \times \Omega_{\mathbf{x}_{\neg i}} \rightarrow [0, 1]$

s.t. $\sum_{\mathbf{x}_i} f_i(\mathbf{x}_i | \mathbf{x}_{\neg i}) = 1 \forall \mathbf{x}_{\neg i}$

$\Rightarrow f_i$ is like a local CPT (conditioned probability table).

Terminology: if we can write $\mathbf{p}(\mathbf{x}) = \prod_{i} f_i(\mathbf{x}_i | \mathbf{x}_{\neg i})$

then we say that $\mathbf{p}$ factors according to $G$.

One example: $X \rightarrow Y \leftarrow Z$

$p \in \mathcal{G}(G) \iff f_X, f_Y, f_Z \text{ legal s.t. } p(x, y, z) = f_X(x)f_Y(y)f_Z(z)$

"leaf placing property" (fundamental property of DGMs)

Let $p \in \mathcal{G}(G)$; let $n$ be a leaf in $G$ (i.e., $n$ has no children).

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

$\Rightarrow$ means remove node $n$ from $G$.

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

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

$p(x_{1:n-1}) = \sum_{x_n} p(x_n, x_{1:n-1}) = \left( \sum_{x_n} f_n(x_n | x_{\neg n}) \right) \prod_{i=1}^{n-1} f_i(x_i | x_{\neg i})$

by def.

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

Let $f_i, f_j$ be a factorization w.r.t. $G$ of $p$

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

i.e. factors are correct conditionals of $p$ (and are unique)
proof: WLOG, let \((1, \ldots, n)\) be a top. sorting of \(G\) (possible since \(G\) is a DAG)

\[ p(x_1|z_{n-1}) = \frac{\prod_{i=1}^{n-1} f_i(x_i|x_{i+1})}{f(x_{n-1})} \]

\(n-1\) is a leaf \(\implies\) "pluck it" to get \(p(x_{1:n-1}) = \prod_{i=1}^{n-1} f_i(x_i|x_{i+1})\)

pluck \(n-1\) as well? (as leaf, because top sort)

keep doing this \(\vdots\) in reverse order, \(n-1, n-2, n-3, \ldots, c+1\) to get

\[ p(x_{1:c}) = \prod_{i=1}^{c} f_i(x_i|x_{i+1}) \]

\(x_{n-1}\) is now a leaf

... no \(x_i\) in any \(x_j\) by top sort property

partition: \(1: c \sqcup (c+1 : n) \sqcup \{n\} \sqcup A \sqcup \emptyset\)

\[ p(x_{1:i}) = p(x_1, x_i, x_{c+1:n}) \]

\[ p(x_1|x_{c+1:n}) = \frac{\sum_{x_A} p(x_1, x_i, x_A)}{\sum_{x_A} p(x_i, x_{c+1:n})} \]

\[ = \prod_{i=1}^{c} f_i(x_i|x_{i+1}) \frac{\sum_{x_A} f_1(x_1|x_{1,c-1})}{\sum_{x_A} f_i(x_i|x_{i+1})} \]

\[ = \prod_{i=1}^{c} f_i(x_i|x_{i+1}) \frac{\sum_{x_A} f_1(x_1|x_{1,c-1})}{\sum_{x_A} f_i(x_i|x_{i+1})} \]

\(\Rightarrow\) (now), we can safely write

\[ \mathcal{S}(G) = \bigotimes p(x) = \prod_{i=1}^{n} p(x_i|x_{i+1}) \]

**Properties of DGM:**

adding edges \(\Rightarrow\) more distributions

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

then \(\mathcal{S}(G) \subseteq \mathcal{S}(G')\)

**Examples:**

- \(E = \emptyset\) \(\Rightarrow \mathcal{S}(G)\) contains only fully independent distributions
  (trivial graph)
  \[ p(x) = \prod_{i=1}^{n} p(x_i) \]
  i.e. \(p(x) = \prod_{i=1}^{n} p(x_i|x_{i+1})\)

- Complete digraph (DA-6)

\[ \mathcal{S}(G) \]
• complete digraph (DA6)
  for all its other
  \((i,j) \in E\)
  \((j,i) \in E\)

\[ p(x; i, j) \]

\[ p(x) = \frac{1}{Z} p(x_1 | x_{i-1}) \]

\[ = \text{all dist. on } x_i \text{ are in } \Gamma \text{(connect)} \]

- removing edges imposes more factorization constraints
  (and thus more condi. indep. assumptions)