

today: DGM

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

$$\mathcal{J}(G) = \left\{ p \text{ is a dist. over } X_V : \exists \text{ legal factors } f_i \text{ s.t. } p(x_V) = \prod_i f_i(x_i | x_{\pi_i}) \forall x_V \right\}$$

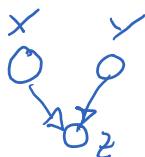
s.t. $f_i: \Omega_{X_i} \times \Omega_{X_{\pi_i}} \rightarrow [0, 1]$

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

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

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

one example:



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

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

"leaf plucking" property (fundamental property of DGM)

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

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

b) moreover, if $p(x_{1:n}) = \prod_{i=1}^n f_i(x_i | x_{\pi_i})$ means remove node n from G 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

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

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

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

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

Let $\{f_i\}$ be a factorization w.r. to G of p

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

ie. 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 \rightarrow "pluck it" to get $p(x_{1:n-1}) = \prod_{j=1}^{n-1} f_j(x_j | x_{\pi_j})$

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

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

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

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

$\cong g(x_{1:i-1})$ ie. no x_j by top. sort properly

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

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

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

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

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

1 \rightarrow \emptyset there is no x_i ? there

\otimes (now), we can safely write

17h33 $\mathcal{J}(G) = \{ p \circ p(x_v) = \prod_{i=1}^n p(x_i | x_{\pi_i}) \}$

properties of DGM:

adding edges \Rightarrow more distributions

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

examples:

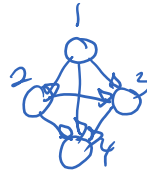
• $E = \emptyset$ (trivial graph) $\Rightarrow \mathcal{J}(G)$ contains only fully independent distributions

ie. $p(x_v) = \prod_{i=1}^n p(x_i)$ \emptyset $p(x_i | \emptyset)$

• Complete digraph (DAG)

• Complete digraph (DAG)

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



get WLOG

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

(like chain rule)

\Rightarrow all dist. on X_i
one in $\mathcal{S}(\text{complete})$

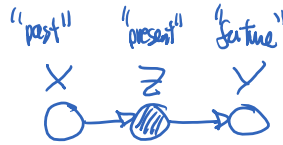
⊗ removing edges impose more factorization constraints

(and thus more conditional indep. assumptions)

conditional indep. in DAG

basic 3 nodes graphs

1) Markov chain



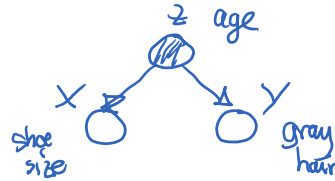
"future" "past" "present"
here $Y \perp\!\!\!\perp X \mid Z$

$$p(x,y,z) = p(y|z)p(z|x)p(x)$$

but $X \not\perp\!\!\!\perp Y$
(for some $p \in \mathcal{S}(G)$)

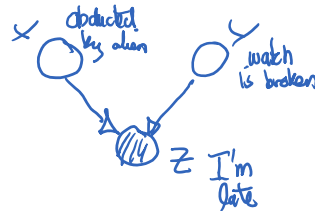
(exercise) $\Rightarrow p(x,y|z) = p(x|z)p(y|z)$

2) latent cause
(hidden variable)



same statements as above

3) explaining away
competing effect
(V structure)



$X \perp\!\!\!\perp Y$
but $X \not\perp\!\!\!\perp Y \mid Z$
(for some p)

non-monotonic property of conditioning:

$$p(\text{alien}) \text{ tiny}$$

$$p(\text{alien} \mid \text{late}) > p(\text{alien})$$

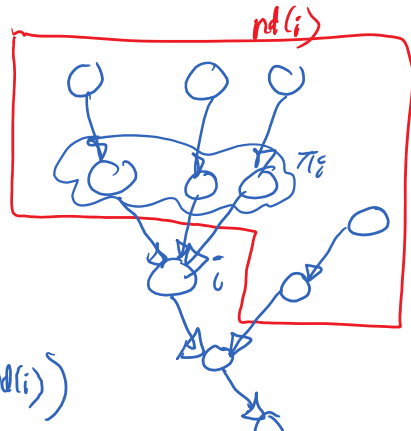
$$p(\text{alien} \mid \text{late}, \text{watch broken}) < p(\text{alien} \mid \text{late})$$

more cond. indep. statements in a DAG

let $nd(i) \triangleq \{j \neq i : \text{no path } i \rightarrow j\}$
"non-descendants" of i

prop.: $p \in \mathcal{S}(G) \Leftrightarrow X_i \perp\!\!\!\perp X_{nd(i)} \mid X_{pa(i)}$

Bay decomposition: $\Rightarrow X_i \perp\!\!\!\perp X_j \mid X_{pa(i)} \cup X_{pa(j)}$



proof: \Rightarrow) key point: let i be fixed

then \exists a top. ordering of G s.t. $nd(i)$ are exactly before i

i.e. $(nd(i), i, \text{descendant}(i))$

$$p(x_i | x_{nd(i)}) = p(x_i, x_{nd(i)}) = p(x_i | x_{\pi_i}) \prod_{j \in nd(i)} p(x_j | x_{\pi_j})$$

place all these ↑

$$p(x_i | x_{nd(i)}) = \frac{p(x_i, x_{nd(i)})}{p(x_{nd(i)})} = \frac{p(x_i | x_{\pi_i}) \prod_{j \in nd(i)} p(x_j | x_{\pi_j})}{\prod_{j \in nd(i)} p(x_j | x_{\pi_j})} = p(x_i | x_{\pi_i})$$

no x_j ' here

\Leftarrow) (suppose p satisfies all ^{these} cond. ind. statements)

let $1:n$ be (wlog) a top sort of G

then $1:i-1 \subseteq nd(i) \forall i$ [key top sort property]

$\Rightarrow X_i \perp\!\!\!\perp X_{1:i-1} | X_{\pi_i}$ by decomposition

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

$$= \prod_{i=1}^n p(x_i | x_{\pi_i}) \quad (\text{key cond. indep.})$$

$\Rightarrow p \in \mathcal{P}(G)$

Other cond. indep. statements?

chain from a to b : just any (undirected) path in graph from a to b

d-separation: (intuition: do not want to have this situation)

def: sets A & B are said to be d-separated by C (in G) ^{set of observed R.V. conditioned on}

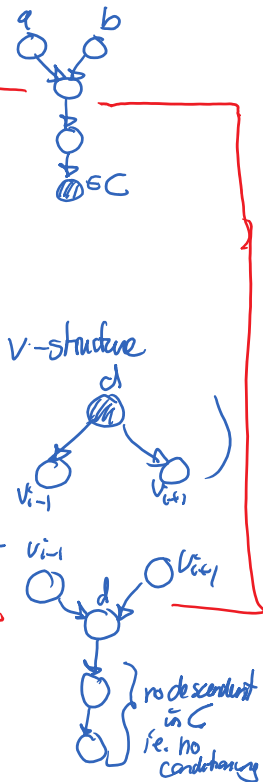
iff all chains from $a \in A$ to $b \in B$ are "blocked" given C

where a chain from a to b is "blocked" given C at node d

if a) either $d \in C$ and (v_{i-1}, d, v_{i+1}) ^{subpath of chain} is not a v-structure



b) $d \notin C$ and (v_{i-1}, d, v_{i+1}) is a v-structure and no descendant of d is in C



prop: $p \in \mathcal{P}(G) \Leftrightarrow X_A \perp\!\!\!\perp X_B | X_C \forall A, B, C \subseteq V$

s.t. A & B are d-separated given C

\uparrow these cover all cond. indep. statements

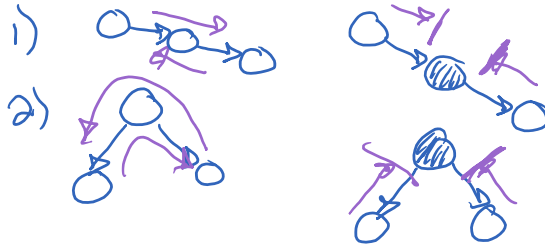
These cover all cond. indep. statements which hold $\forall \theta$

"Bayes Ball alg."

"intuitive" alg. to check d-separation

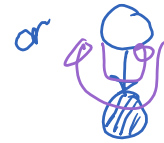
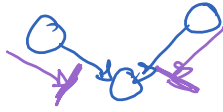
rules balls/chains being blocked

a) in d-separation



b)

3) v-structure



see Alg. 3.1 in K&F book

