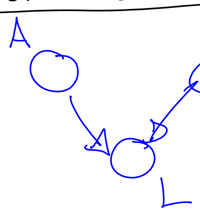
Lecture 11 - scribbles

Friday, October 7, 2016
13:44

today: · DGM
· UGM

V-structure revisited (explaining away / common effect)



$p(a,b,\ell) = \underbrace{p(\ell \mid a,b)}_{\text{CPT}}\, p(a)\, p(b)$

$p(a)$: $A=1$: ε; $A=0$: $1-\varepsilon$

$p(b)$: $B=1$: 0.1; $B=0$: 0.9

CPT:

A	B	$P(L=1 \mid a,b)$
1	1	1
1	0	1
0	1	0.8
0	0	0.1

"A=1"

$A \;\{\; \neg A$

$P(A) = \varepsilon$

$P(A \mid L) = \frac{P(A,L)}{P(L)} \rightarrow \underbrace{\overset{p(B)\,p(A)\,p(L|A,B)}{0.1 \cdot \varepsilon \cdot 1} + \overset{p(\neg B)\,p(A)\,p(L|A,\neg B)}{0.9\varepsilon \cdot 1}}_{} = \varepsilon$

$\rightarrow \varepsilon + 0.8 \cdot (1-\varepsilon)\, 0.1 + 0.1 (1-\varepsilon)\, 0.9$

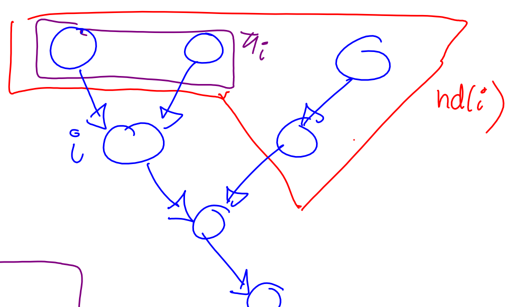
$= \frac{\varepsilon}{\varepsilon + (1-\varepsilon)\,[\overbrace{0.8 \cdot 0.1 + 0.1 \cdot 0.9}^{0.17}]} = \frac{\varepsilon}{0.17 + \varepsilon \cdot 0.83} > \varepsilon$

$P(A \mid L,B) = \frac{P(A,L,B)}{P(L,B)} \rightarrow \cancel{0.1}\varepsilon$

$\rightarrow \cancel{0.1}\varepsilon + 0.8 \cdot \cancel{0.1}(1-\varepsilon)$

$= \frac{\varepsilon}{\varepsilon + (1-\varepsilon)0.8} = \frac{\varepsilon}{0.8 + 0.2\varepsilon} < \frac{\varepsilon}{0.17 + 0.83\varepsilon} = P(A \mid L)$

Cond. indep. in DGM:



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

prop: $p \in \mathcal{L}(G) \iff$

$$\boxed{X_i \perp\!\!\!\perp X_{nd(i)} \mid X_{\pi_i} \quad \forall i}$$

(by decomposition $\Rightarrow X_i \perp\!\!\!\perp X_j \mid X_{\pi_i} \;\; \forall j \in nd(i)$)

proof:

⇒) Key point: Let i be fixed

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

$$\big(nd(i),\ i,\ \overbrace{descendants(i)}^{\text{pluck all these}}\big)$$

(why? · desc. of i have to be to the right of i)

[after marginalizing $x_{descendants(i)}$]

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

$$p(x_i | x_{nd(i)}) = \frac{p(x_i, x_{nd(i)})}{p(x_{nd(i)})} = \frac{p(x_i|x_{\pi_i}) \cancel{\prod_{j\in nd(i)} p(x_j|x_{\pi_j})}}{\underbrace{\sum_{x_i} p(x_i|x_{\pi_i})}_{1} \cancel{\prod_{j} ''}}$$

$$= p(x_i | x_{\pi_i}) \quad \text{i.e. } X_i \perp\!\!\!\perp X_{nd(i)} | X_{\pi_i}$$

⇐) (suppose p satisfies cond. indep. statements)

let $1{:}n$ be top. sort

then $1{:}i-1 \subseteq nd(i)\ \forall i$

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

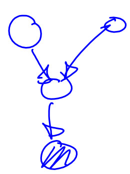
$$p(x_V) = \prod_{i=1}^{n} p(x_i | x_{1:i-1}) \quad \text{(by chain rule)}$$

$$= \prod_{i=1}^{n} p(x_i | x_{\pi_i}) \quad \text{(by C.I.)}$$

$\Rightarrow p \in \mathcal{L}(G)$ //

other C.I. statements?

intuition is no

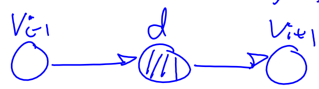


chain: just any (undirected) path from in graph from a to b
from a to b

set of observed / conditioned R.V.

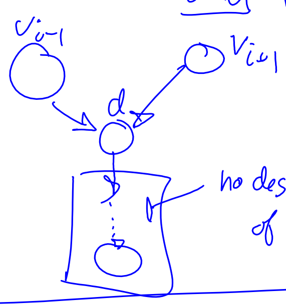
d-separation: set A & B are said to be d-separated by C iff all chains from $a \in A$ to $b \in B$ are "blocked" given C

- where a chain from a to b is blocked at node d if a) either $d \in C$ and (v_{i-1}, d, v_{i+1}) is not a V-structure



or b) $d \notin C$ and (v_{i-1}, d, v_{i+1}) is a v-structure

and no descendent of d is in C

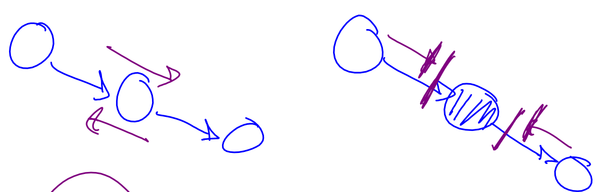


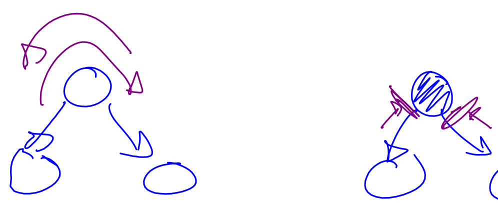
prop: if $p \in \mathcal{L}(G) \iff X_A \perp\!\!\!\perp X_B \mid X_C$

$\forall A, B, C$ s.t. A & B are d-separated by C

"Bayes-Ball" alg.: "intuitive" alg. to check d-separation

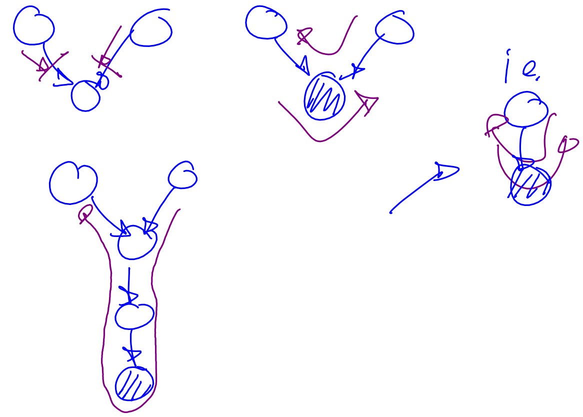
rules for balls / chain being blocked:

1) 

2) 

a b note: a can be

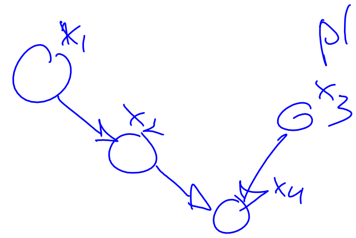
3) V-structure



same as b

i.e.

see Alg. 3.1 in K & F book



$p(x_{1:4}) = \prod_{i=1}^{n} p(x_i)$

$p(x_i \mid x_{\pi_i})$

$p(x_i \mid x_{\pi_i}) = p(x_i) \; \forall x_{\pi_i}$

Properties of DGM :

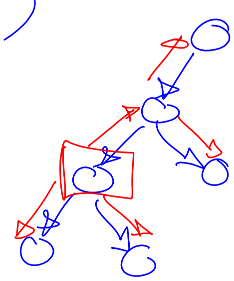
inclusion : $E \subseteq E'$ $\quad \mathcal{L}(G) \subseteq \mathcal{L}(G')$

(adding edges ↗ set of distributions)

reversal : if G is a directed tree (or forest) $(|\pi_i| \leq 1$; i.e. $\nexists$ no v-structure)

then let E' be another directed tree by choosing a different root

(reverse some edges while keeping a directed tree)



$\Rightarrow \quad \mathcal{L}(G) = \mathcal{L}(G')$

$G = (V,E)$, $\quad G' = (V,E')$

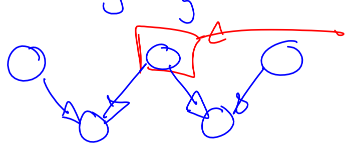
$\Rightarrow$ why directed edges are not causal

rephrasing : all directed trees from undirected tree are equivalent

marginalizing :

- marginalizing a leaf gives us a smaller DGM
- not true for all marginalization

e.g.



marginalize out this

get a set of dist.

$\neq$ to any $\mathcal{L}(G')$ for some G'

Γ marginalization

[marginalization is not a closed operation]

undirected GM

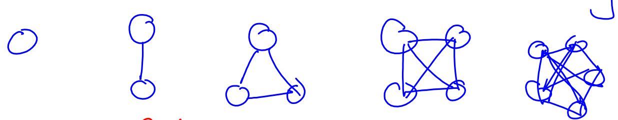
(aka Markov random field, Markov network)

let $G = (V, E)$ be undirected graph

let $\mathcal{C}$ be set of cliques of G

clique = fully connected set of nodes

$$C \in \mathcal{C} \iff \forall i, j \in C,\ \{i, j\} \in E \quad (i \neq j)$$



then $\mathcal{L}(G) = \left\{ p : p(x) = \frac{1}{Z} \prod_{C \in \mathcal{C}} \psi_C(x_C) \right\}$

where $\psi_C(x_C) \geq 0 \ \forall x_C$ "potentials"

and $Z \triangleq \sum_x \left(\prod_{C \in \mathcal{C}} \psi_C(x_C) \right)$ "partition function"

notes:
- can multiply any $\psi_C(x_C)$ by a constant without changing p
- it is sufficient to consider only $\mathcal{C}_{max}$, the set of maximal clique
  e.g. $C' \subseteq C$
  $\tilde{\psi}_C(x_C) = \psi_C(x_C) \cdot \psi_{C'}(x_{C'})$
- unlike for B.N. (where could think of C to be (i, π_i), $\psi_C(x_C) = p(x_i \mid x_{\pi_i})$)
  here, $\psi_C(x_C)$ is not related to $p(x_C)$
- used in statistical physics, computer vision, NLP (CRF)

properties:

as before, $E \subseteq E' \Rightarrow \mathcal{L}(G) \subseteq \mathcal{L}(G')$

$\bar{E} = \emptyset \Rightarrow$ fully factorized dist.

$E =$ all pairs $\Rightarrow$ all distributions
(one big clique)

- for $\psi_C(x_C) > 0 \ \forall x_C$

can write $p(x) = \exp\Big(\overbrace{\sum_{C \in \mathcal{C}} \underbrace{\log \psi_C(x_C)}_{\langle \theta_C, T_C(x_C) \rangle}}^{\text{negative energy function}} - \log Z \Big)$

$\rightarrow$ "exponential family" (we'll see later)