1/28/14

BAYES NETS: Factorization, v-structures, Active Trails

Notation Clarification

$$\underbrace{P}_{\text{a distribution } P} \; \underbrace{\models}_{\text{encodes/satisfies}} \; \underbrace{X \perp Y}_{X \text{ indep. of } Y}$$

$\overset{\text{iff}}{\iff} P(X=x, Y=y) = P(X=x)P(Y=y)$ $\forall x,y$

$\overset{\text{iff}}{\iff} P(X=x \mid Y=y) = P(X=x)$ $\forall x,y$

$\overset{\text{iff}}{\iff} P(Y=y \mid X=x) = P(Y=y)$ $\forall x,y$

$$P \models X \perp Y \mid Z \quad \text{(conditionally)}$$

"P satisfies/encodes X is conditionally indep of Y given Z"

$Q \not\models X \perp Y \mid Z$

or $Q \models X \not\perp Y \mid Z$

In $Q(X,Y,Z)$ X is not conditionally indep of Y given Z

ie $Q(X=x, Y=y \mid Z=z) \neq Q(X=x \mid Z=z)\, Q(Y=y \mid Z=z)$

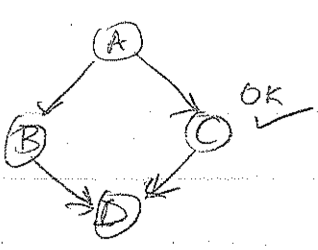
for some (x, y, z)

(1) Bayes Nets

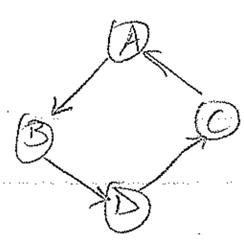
→ Set of RVs $\{X_1, \ldots X_n\}$

→ DAG $G = (V, E)$ $V = \{X_1, \ldots, X_n\}$

Directed Acyclic Graph



OK ✓



X Not Okay

→ Conditional Prob. Tables

$P(X_i \mid Pa(X_i))$

X_i \ $Pa(X_i)$	000	001	010	
0	0.3	0.2		- - -
1	0.7	0.8		

$\#\text{Params} = 2^{\#\text{parents}} \cdot (2-1)$

→ BN Factorization

$$P(x_1, \ldots, x_n) = \prod_i P(X_i \mid Pa(X_i))$$

→ # parameters comparison

⇒ n vars; each k states

→ $P(x_1 \ldots x_n) \equiv k^n - 1$

→ at most d parents in BN

⇒ $|Pa(X_i)| \leq d \quad \forall i$

⇒ $\#\text{params} = n \cdot k^d \cdot (k-1)$

$\approx n k^{d+1}$

Big difference

② Independence Assumption

→ A BN structure (DAG) encodes indep. assumptions

$I(G) = \{$ set of indep assumptions encoded in $G \}$

↑

notice just G; no CPTs

What is this set & how do we "read" it from G?

Examples: Smallest BNs

(2.1) 1 Node X CPT $P(X)$; Nothing to do

(2.2) 2 Nodes X, Y

Case 1: X Y

BN factorization $P(x, y) = P(x)P(y) \quad \forall x, y$

$$\Rightarrow \underbrace{X \perp Y \in I(G)}$$

Irrespective of CPTs

Case 2: $X \rightarrow Y$ CPTs $P(X)\ P(Y|X)$

$X \leftarrow Y$ $P(Y)\ P(X|Y)$

$$X \perp Y \notin I(G)$$

G does not make ANY indep assumption

!However! what if $P(y|x)$:

$Y \backslash X$	0	1
0	0.3	0.3
1	0.7	0.7

$P(x)$:

$$\begin{array}{c|c} X & \\ 0 & 0.5 \\ 1 & 0.5 \end{array}$$

So $P(y|x) = P(y) \quad \forall x, y$

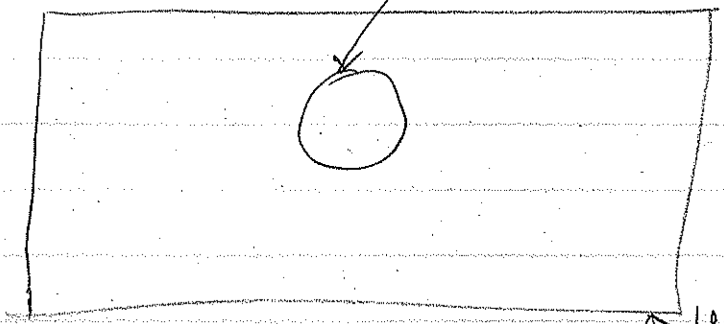
So there is a way to set CPT

$$P_{CPT} \models X \perp Y$$

but our graph does not make this indep. assumption

Intuitively

Prob distributions where $X \perp Y$ / $I(G)$ holds
[corresponds to (X) (Y)]



Universe of all prob distributions on X, Y
$P(X,Y)$

[corresponds to $X \rightarrow Y$
or $Y \leftrightarrow X$]

(2.3) 3 Nodes, 2 edges

→ Case 1
- $X \rightarrow Z \rightarrow Y$
- $X \leftarrow Z \leftarrow Y$
- $X \leftarrow Z \rightarrow Y$

→ Case 2
- $X \rightarrow Z \leftarrow Y$

→ Question 1 $X \perp Y \in I(G)$?

→ Question 2 $X \perp Y \mid Z \in I(G)$

(2.3.1) $(X) \rightarrow (Z) \rightarrow (Y)$

Is $X \perp Y$?

No, counterexample

X uniform $\quad P(X) = \begin{bmatrix} 0.5 \\ 0.5 \end{bmatrix}$

$Z = X \quad P(Z \mid X) =$

$Y = Z$

Z \ X	0	1
0	1	
1		1

Now, in such a $P(X, Y, Z)$ X determines Z
so $X \perp Z$ false.

2.3.2 $X \to Z \to Y$

Is $X \perp Y \mid Z \in I(G)$

Yes!

$$P(Y=y \mid X=x, Z=z) = \frac{P(Y=y, X=x, Z=z)}{P(X=x, Z=z)}$$

$$= \frac{P(X=x, Y=y, Z=z)}{\sum_y P(X=x, Y=y, Z=z)}$$

$$= \frac{P(X=x)P(Z=z \mid X=x)\,P(Y=y \mid Z=z)}{\sum_y P(X=x)\,P(Z=z \mid X=x)P(Y=y \mid Z=z)}$$

$\Sigma(\,) = 1$

$$= P(Y=y \mid Z=z)$$

2.3.3 $X \to Z \leftarrow Y$

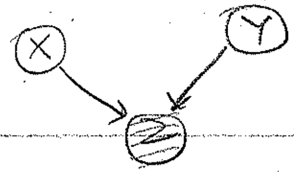
Is $X \perp Y \in I(G)$

Yes! $P(X=x, Y=y) = \sum_z P(X=x, Y=y, Z=z)$

$$= \sum_z P(X=x)\,P(Y=y)\,P(Z=z \mid X=x, Y=y)$$

$\Sigma(\,) = 1$

$$= P(X=x)\,P(Y=y)$$

2.3.4  Is $X \perp Y \mid Z \in I(G)$?

- No!

Counterexample $P(x) = P(y) = \begin{bmatrix} 0.5 \\ 0.5 \end{bmatrix}$ uniform

$$Z = X \underset{\text{XOR}}{\oplus} Z$$

So $P(z \mid x, y) =$

$Z \backslash XY$	00	01	10	11
0	1			1
1		1	1	

Now, knowing x & z uniquely determine y
& " y & z " " x

Specifically $P(X=0 \mid Y=0, Z=0) = \boxed{1}$

$$P(X=0 \mid Z=0) = \frac{P(X=0, Z=0)}{P(Z=0)}$$

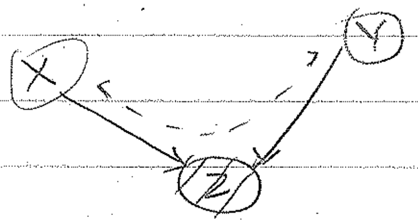
$$= \frac{\sum_y P(X=0, Y=y, Z=0)}{\sum_{x,y} P(X=x, Y=y, Z=0)}$$

$$= \frac{\sum_y P(X=0) P(Y=y) P(Z=0 \mid X=0, Y=y)}{\sum_{x,y} P(X=x) P(Y=y) P(Z=0 \mid x, y)}$$

$$= \frac{P(X=0) P(Y=0) \cdot 1}{P(X=0) P(Y=0) \cdot 1 + P(X=1) P(Y=1) \cdot 1}$$

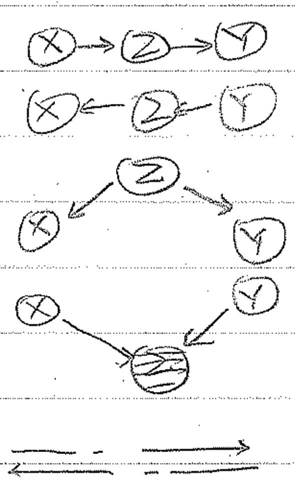
$$= \frac{0.5^2}{2 \times 0.5^2} = \boxed{\frac{1}{2}}$$

This phenomenon is called "Explaining Away"

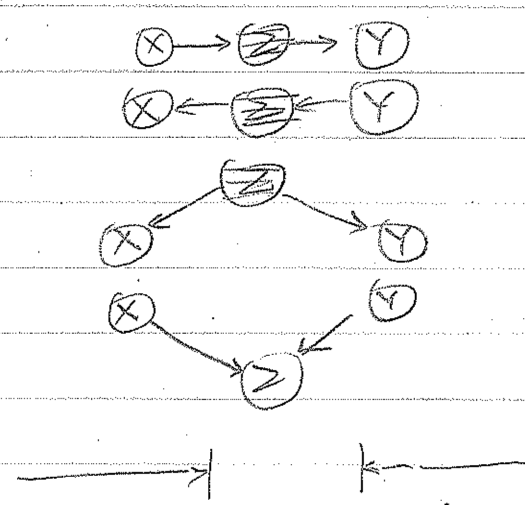


X & Y "cause" Z. If you know Z happened, knowing X explaining away Y.

③ Flow of Info / Bayes Ball Rule



Info flows



Info blocked