RECITATION NOTES - 10708
09/18/08
DHRUV BATRA

→ BN Semantics / Representation

→ Defn. of BN (general BN)

structure { → set of random variables $\{X_1, \ldots, X_n\} \triangleq \mathcal{X}$

→ DAG over $\mathcal{X}$

parameters { → CPTs $P(X_i \mid Pa(X_i))$

→ Joint factorizes $P(\mathcal{X}) = \prod_{i=1}^{n} P(X_i \mid Pa(X_i))$

why?

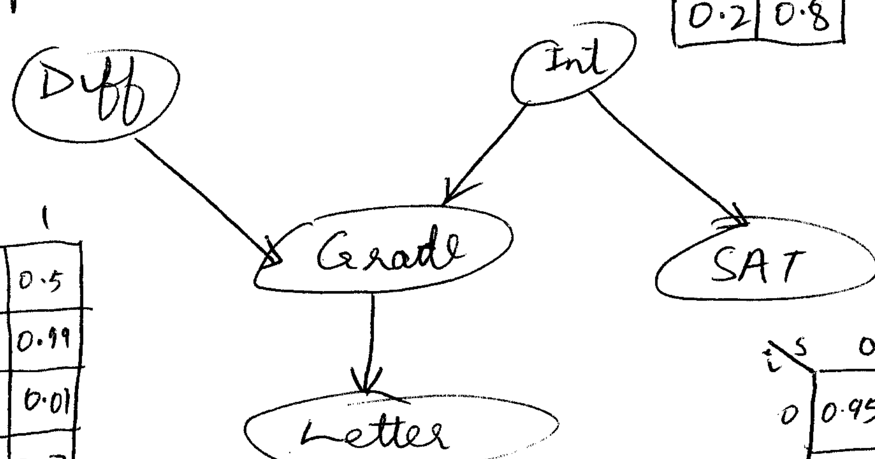
→ <u>Local Markov Assumptions</u>

$(X_i \perp NonDesc \mid Pa(X_i))$

→ Example

0	1
0.1	0.9

0	1
0.2	0.8



di \ g	0	1
00	0.5	0.5
01	0.01	0.99
10	0.99	0.01
11	0.3	0.7

i \ s	0	1
0	0.95	0.05
1	0.2	0.8

g \ l	0	1
0	0.9	0.1
1	0.2	0.8

→ ~~Queries~~ Queries

→ Inference / MAP

→ MPE (Most probable explaination)

<u>Inference</u>/

Ask for prob. of subset given observations

① $\underbrace{P(L \mid G)}$ (simplest query) Why?

Lookup in CPT

Ask for prob of a complete assignment

② $P(d=1, i=1, s=0, g=1, l=1)$

$= P(D=1)\, P(I=1)\, P(G=1 \mid D=1, G=1) \cdot P(S=0 \mid I=1)$
$\quad P(l=1 \mid G=1)$

$= 0.9 \times 0.8 \times 0.7 \times 0.2 \times 0.8$

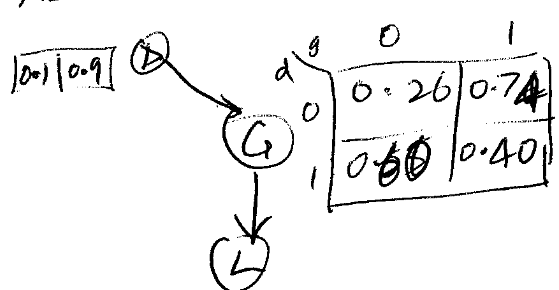
$= 63 \times 64 \times 2 \times 10^{-5} = (4096 - 64) \cdot 2 \times 10^{-5}$

$= 4032 \cdot 2 \times 10^{-5}$

$= 8064 \times 10^{-5} \triangleq 0.081$

---

③ $P(D \mid G=1)$ [Assume partial graph]



$$P(D=1 \mid G=1) = \frac{P(G=1 \mid D=1)\, P(D=1)}{P(G=1)}$$

$$= \frac{P(G=1 \mid D=1)\, P(D=1)}{\sum_{d=0}^{1} P(G=1 \mid D=d)\, P(D=d)}$$

$$= \frac{0.9 \times 0.4}{(0.1 \times 0.74) + (0.9 \times 0.4)} \approx 0.83$$

Do you expect this to change when rest of variables are added in?
why? (hint: v-structure)

MAP

① ~~what is~~ ~~Ma~~ ~~arg~~
MAP $\triangleq$ arg max $P(Q \mid E)$

e.g what is MAP value of $D \mid G=1$

$\boxed{\text{ans } 1}$

Discussion: Sometimes we're more interested in MAP than prob.

MPE

Most probable explaination of all variables other than observation

1. No observation
"Most likely student"

$\arg\max P(D, I, G, L, S)$

Naively 2^5

~~2~~

Discussion MPE & MAP explainations might be different!

→ Graphs & Distributions

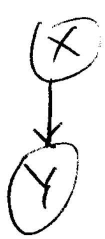
→ Defns: $I_\ell(G)$ Local Markov Assumptions

→ $I(P)$ (true) independence assertions from data/model world

→ [Defn]: I-map
G is Imap if $I_\ell(G) \subseteq I(P)$

Example

Is this an I-map for any/all $P(X, Y)$?



Is this?

→ ~~Defn~~ What does it mean for a distribution P to "factorize" over BN G

[Defn]: P factorizes over G if

$$P(X) = \prod_{i=1}^{n} P(X_i \mid Pa(X_i))$$

Refined Defn of BN as (G, P) where P factorizes over G

~~2 Fund~~ Equivalence of I-maps & P's (that factorize)

Given: G over $\mathcal{X} = \{x_1, \dots, x_n\}$, P distribution over $\mathcal{X}$

→ ~~Def~~ Claim 1:

G is an I-map for P $\Longrightarrow$ P factorizes according to G

ie $I_\ell(G) \subseteq I(P)$

→ Proof: Topological ordering + Chain Rule

Th 3.2.6 KF

Implication: If we find an I-map, we can (potentially) save some parameters

→ A trivial I-map always exists, but that ~~doe~~ doesn't help

→ Claim 2:

P factorizes according to G $\implies$ G is an I-map of P

Proof: HW1

Implication: G can help us read-off independance / dependance relations beyond $I_\ell(G)$