- Annoncementsa : just do the wiki demo
  Make sure you read them all
~~(add announcement: thanks to all the students who've been helping~~

Lec 22
4/21/15

M~~ore intuition on~~
~~Bay~~
More intuition on Bayes
Nets.

## Outline

- review of dependencies; info in BNs: [clicker II, III]
  and introduce running example
- why trees are nice (variable elimination → message passing from ~~parent~~ ~~children~~ ~~children~~ parent to ~~children~~ child ~~parent~~ ~~parent~~
  [by defn, the kids are independent given the parent] and from
  ~~a ch~~ a child to its single parent.
  - o.w., as in the "Job" node, ~~you can get blowup in~~ where "Job" is to be solved for, you get a blowup in factor sizes.
- tree organization: "above" vs. "at and below
- in HMMs, traditionally a slightly diff organization
  ("forward" includes "at & above").

which leads to trellis-graph-organized dynamic program.
<use square nodes, to avoid confusing w/ a graphical model>

- can also be used to compute Viterbi paths.

<can we use inverted display in order to show on the board?>

lecpt I : graphical model for variables relevant to a student taking a particular student.

What affects $\underline{G}$, a student's grade in the course?
In this model, course grade is affected by the student's intelligence & the difficulty of the course.

Difficulty comes from how coherent the lecturer is, but if we know how difficult $\underset{\underline{D}}{\uparrow}$ the course is, knowing C, coherence, doesn't give us any extra info about the grade.

Grades affect how good your letter of recommendation is;

~~Yo~~ Your letter of rec. & your SAT scores apparently, according to this model, affect ~~your~~ what job you get;

Your grade & your job affect your happiness.

clicker q II.

4=15% (A) is false: if I know your ~~grade~~ SAT score, I have ~~some info about your SAT score~~ a more accurate guess @ your ~~SAT score~~ grade than I did before.

12=46% (B) is true: if I know your intelligence, my accuracy @ guessing your grade isn't going to change if I'm also told your SAT score.

2=8% (C) is false. spse I know ~~your~~ you've got a good job. if I know you got a bad letter of recommendation, I would ~~update my~~ upgrade my guess that you got ~~a~~ a high SAT score.

8=31%
<so D is false>

this is review, don't write

(A)+(B): children are not indep, but are cond indep given parent.

(C) parents are not indep given children

clicker q III : motivates : why do we write things like $\underline{m_X(Y,Z)}$ instead of $P(Y,Z)$?
(or whatever)

~~X has~~
sum over X of products involving Y, Z, marginalizing X out.

4=16% (A) - <note: the "equals" ~~a~~ didn't print>
true.

3=12% (B) - ~~false~~. : $P(D) = \sum_c P(C,D) = \sum_c \underbrace{P(D=d|C=c)}_{\text{given}}\underbrace{P(C=c)}_{\text{given (cond on "no parent")}}$

but you have to compute the sum.

so we write $m_c(D)$, indicating some work needs to be done,

3=12% (C) true

try to reserve "$P(\cdot)$" for what's given by the BN tables.
inside summations

13=52% [... so the answer is (D)]
2 Es?

---

Exercise to see how structure affects our computations:
Let's try to compute $P(J) = \sum_{\cancel{C,D,I,\ldots}} P_{BN}(C, D, I, \ldots, H)$
all except J

clicker q IV : a leaf below J. <yes, we eliminated happiness...>

14=56% said disappears
4=16% said becomes subscript
7=28% said remains an argument

$\sum_{\text{all except } J, H} P \;\ldots\ldots\; \cancel{\sum_h P(H=h|G,J)}$
1.

If we don't know happiness, it doesn't ~~matter~~ affect our guess @ Job.

- cross off happiness.

let's try to simplify further
clicker q V an orphan w/ 1 child.

2=8% disappears
12=46% said b/c $m_c$
12=$\boxed{46\%}$ said remains an arg.
probably didn't make leap to $m_c$.
- cross off coherence.

→ no, b/c the point is to eliminate that variable from the expression.

$\sum_{\text{all except } J, H, C}$

$\underline{\sum_c P(C=c)P(D|C=c)}$
↓
change to $m_c(D)$.
requires some work for us to do.
STORE this for rainy day.

can similarly knock off orphan D
skip computation.

$$\sum_{\text{all except } J,H,C,D} \cdots \quad \underbrace{\sum_d \underbrace{m_c(D=d)}_{\text{rewrite as } m_c(D=d \mid I)} P(G \mid D=d, I)}_{\underbrace{\sum_d P(G, D=d \mid I)}_{\text{mathematically equals } P(G|I) \text{, given by table.}}}$$

in lecture, I skipped the elimination of D, which, naturally, gives ~~you~~ you a different result when you eliminate I. But you still get a multi-argument m.

: parent w/ multiple children affecting (eventually) J ?

$$\sum_{\text{all except } J,H,C,D,I} \quad \underbrace{\sum_i P(I=i)\, P(G|I)\, P(S|I)}$$

not given in table, define as $m_I(G,S)$

note: now we've got two interlocked variables G, S are "connected".

so the answer is (B).

~~most~~ I accidentally signaled that "test prep" suggests the answer is (C) since we hadn't used C before.

0 A
7: 37% B
12: 63% C

Tree structure: ~~this new~~ that never happens

(messages)
* only get info up from children or down from parent.

so, we don't get big combinations in the m's
computation is faster.

an HMM looks like this..

so ~~working upwards is great!~~
~~But, wait.~~ This doesn't look like what the bfwd alg does. ~~I guess b/c we are marginalizing over the observed~~?

make sure X, Y, Z are observed.

what about working downwards?

what is P(C)?

eliminate root? yuck...

$$\sum_A P(W|A)P(B|A)P(A)$$

did not do

1st eliminate W:

$$\sum_W \cancel{P(W|A)}$$

now eliminate root:

$$\underline{\sum_A P(B|A)P(A)}$$
$$\Downarrow$$
$$m_A(B)$$

now eliminate leaf X:

[illegible]

$$\sum_X \cancel{P(X|B)}$$

now eliminate chain-root B:

$$\underline{\sum_B P(C|B)m_A(B)}$$
$$\Downarrow$$
$$m_B(C)$$

elim leaf X, leaf Y.

$$m_B(C) = \sum_B P(C|B)\underline{m_A(B)}$$
$$\hookrightarrow \sum_A P(A,B)$$

↳ store this!

So, let's talk about computing $P(B$ ~~$\mid W = \ldots, X = x, Y =$~~ $, \overset{=x}{X}, \overset{=y}{Y}, Z=z)$

Bottom up:

$P(Z=z \mid C)$

P

~~lst~~ eliminate the "leaf" C

$$\underbrace{\sum_C P(C \mid B)\, P(Z=z \mid C)}_{m_C(B, Z=z)}$$

"bkwd": given 2nd state B, what's the prob of 3rd emission = z?

which we can write as
$\beta_2(b) = P(Z=z \mid B=b)$

~~eliminate~~
now we have to do the root

$$\underbrace{\sum_A P(B \mid A)\, P(X=x \mid A)\, P(A)}_{m_A(B, X=x)}$$

~~forward~~ P(2nd state is B ! 1st emission is X),
almost forward.

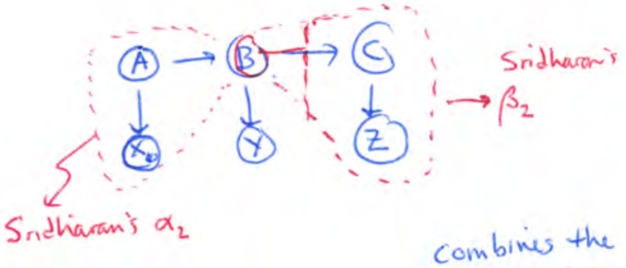
which we can write as a
"Sridharan α":
$\alpha_2(b) = P(X=x, B=b)$
(~~generically, P(X=x,~~

~~and~~ and we have one more leaf

$P(Y=y \mid B)$

prob the 2nd emission is "y" given 2nd state is B.

~~"forward" score, which we can write as~~



combines the $\alpha_2$, the $\beta_2$, P of transitioning ~~from B~~ to C given B,
prob of ~~B output~~ Y=y given B,
~~and you get the prob of the whole~~