Probabilistic Reasoning

Unit # 2

Graphical Models

- Models are descriptions of how (a portion of) the world works
- Models are always simplifications
  - May not account for every variable
  - May not account for all interactions between variables

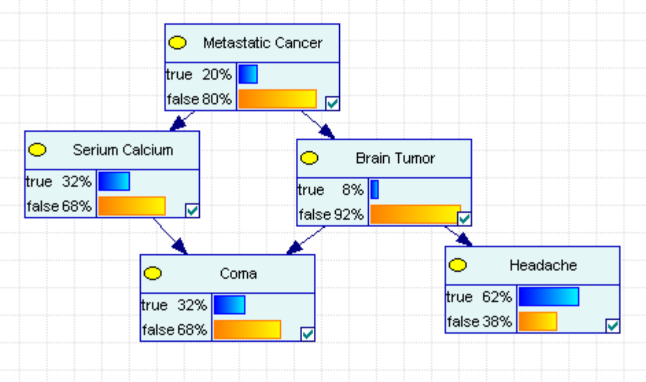
Inference Problem

- The canonical inference problem: find the posterior probability distribution for some variable(s) given direct or virtual evidence about other variable(s).
- Diagnostic Reasoning
  - Reasoning from symptoms to cause
  - This reasoning occurs in the opposite direction to the network arcs
- Predictive Reasoning
  - Reasoning from new information about causes to new beliefs about effects
  - Follows the directions of the network arcs.

Cancer Example (Pearl)

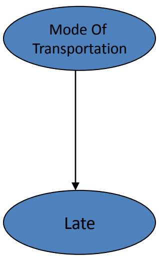
- Metastatic cancer is a possible cause of a brain tumor and is also an explanation for increased total serum calcium. In turn, either of these could explain a patient falling into a coma. Severe headache is also possibly associated with a brain tumor.

Cancer Example (Cont'd)



Revised "Bob late to work" Example

- Mode of Transportation: {Car, Not Car (Public Transport)}
- Late: {True, False}
- P(C) = 1/3, P(¬C) = 2/3
- P(L | C) = 0.4, P(L | ¬C) = 0.2



- **P(L) = ?**
- **P(C | L), P(¬C | L)**

Revised “Bob late to work” Example

- One approach is to compute the joint distribution through
  - $P(X_1, X_2, \ldots, X_n = \prod P(X_i \mid \text{parents}(X_i))$
  - $P(L, C) = P(L \mid C)\, P(C) = 0.4 \times 0.33$
  - $P(L, \neg C) = P(L \mid \neg C)\, P(\neg C) = 0.2 \times 0.67$
  - $P(\neg L, C) = P(\neg L \mid C)\, P(C) = 0.6 \times 0.33$
  - $P(\neg L, \neg C) = P(\neg L \mid \neg C)\, P(\neg C) = 0.8 \times 0.67$
- Now compute P(L), P(C | L)
  - $P(L) = P(L, C) + P(L, \neg C)$
  - $P(C \mid L) = \dfrac{P(C, L)}{P(L)}$

Revised “Bob late to work” Example

Or we could use Bayes Theorm based propagation.

$$P(C \mid L) = \frac{P(L \mid C)\, P(C)}{P(L)} = \frac{0.4 \times 0.33}{P(L)}$$

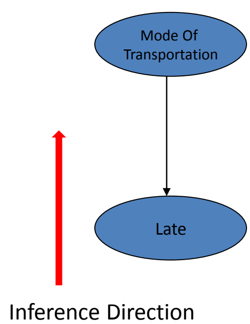
where

$P(L) = P(L \wedge C) + P(L \wedge \neg C)$

$= P(L \mid C)P(C) + P(L \mid \neg C)\, P(\neg C)$

$= 0.4 \times 0.33 + 0.2 \times 0.67 = ?$

Both approaches (on the previous and the current slides) should give the same result

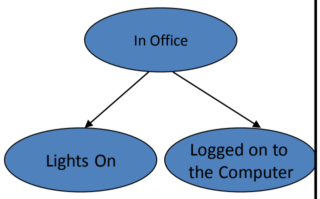


Lecturer's Life Example (Korb & Nicholson)

- Dr. Ann Nicholson spends 60% of her work time in her office. The rest of her work time is spent elsewhere. When Ann is in her office, half the time her light is off (when she is trying to hide from students and get some real work done).
- When she is not in her office, she leaves her light on only 5% of the time.
- 80% of the time she is in her office, Ann is logged onto the computer.
- Because she sometimes logs onto the computer from home, 10% fo the time she is not in her office, she is still logged onto the computer.
- Suppose a student checks Dr. Nicholson's login status and sees that she is logged on. What effect does this have on the student's belief that Dr. Nicholson's light is on?

Lecturer's Life Example (Cont'd)

- In Office (O): {T, F}
- Lights On (L): {T, F}
- Logged on to the Computer (C) : {T, F}
- P(O) = 0.6
- $P(L \mid O) = 0.5, P(L \mid \neg O) = 0.05$
- $P(C \mid O) = 0.8, P(C \mid \neg O) = 0.1$

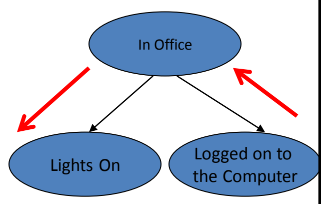


- **$P(L \mid C) = ?$**

Lecturer's Life Example (Cont'd)

- One possibility is to compute the joint distribution and then compute the conditional probability
  - $P(X_1, X_2, \ldots, X_n = \prod P(X_i \mid parents(X_i))$
  - $P(O, L, C) = P(L \mid O)\, P(C \mid O)\, P(O)$
  - $P(O, L, \neg C) = P(L \mid O)\, P(\neg C \mid O)\, P(O)$
  - $P(O, \neg L, C) = P(\neg L \mid O)\, P(C \mid O)\, P(O)$
  - $P(O, \neg L, \neg C) = P(\neg L \mid O)\, P(\neg C \mid O)\, P(O)$
  - $P(\neg O, L, C) = P(L \mid \neg O)\, P(C \mid \neg O)\, P(\neg O)$
  - $P(\neg O, L, \neg C) = P(L \mid \neg O)\, P(\neg C \mid \neg O)\, P(\neg O)$
  - $P(\neg O, \neg L, C) = P(\neg L \mid O)\, P(C \mid \neg O)\, P(\neg O)$
  - $P(\neg O, \neg L, \neg C) = P(\neg L \mid \neg O)\, P(\neg C \mid \neg O)\, P(O)$
- After computing the joint distribution, we can compute P(L | C)

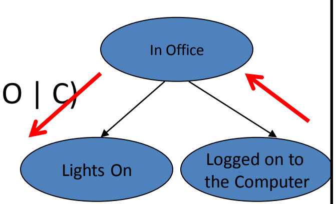
Inference Direction



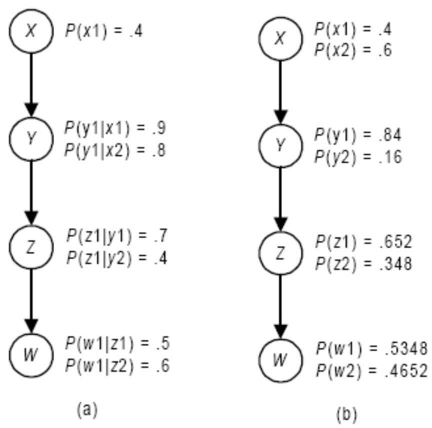
Lecturer's Life Example (Cont'd)

- The other approach is to do inference using probability propagation based on Bayes Theorem
- First compute the posterior probability P(O | C) {referred to as P(O*)}
  - $P(O^*) = \dfrac{P(C \mid O)\, P(O)}{P(C)}$
  - And then P(L | C) {referred to as P(L*)}
  - $P(L^*) = P(L \mid O)P(O^*) + P(L \mid \neg O)P(\neg O^*)$

Inference Direction



Inference in Singly Connected Networks

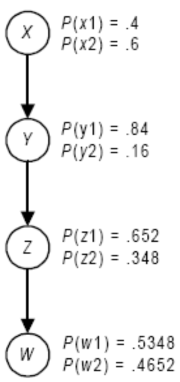


P(y1) = P(y1 | x1) P(x1)
+ P(y1 | x2) P(x2)
P(y2) = 1 – P(y1)

P(z1) = P(z1 | y1) P(y1)
+ P(z1 | y2) P(y2)
P(z2) = 1 – P(z1)

P(w1) = ???

Inference in Singly Connected Networks



Suppose we get evidence that w1 is true, i.e., P(w1) = 1.

Now compute the posterior probabilities:
P*(z1), P*(y1), P*(x1)

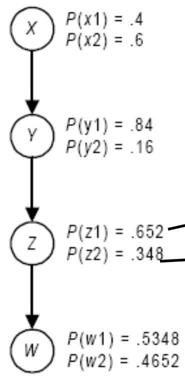
P*(z1) = P(z1 | w1) P(w1) ← P(w1)=1
+ P(z1 | w2) P(w2) ← P(w2)=0

Computing P(z1 | w1) using Bayes theorem:
P(z1 | w1) = P(w1 | z1) P(z1) / P(w1)
P(z1 | w1) = 0.5 x 0.652 / 0.5348 = 0.61

=>
P*(z1) = 0.61 * 1 + 0 = 0.61

Computation of P*(y1)
on the next slide

Inference in Singly Connected Networks



Now update the probability of Y.
P*(y1) = P(y1 | z1) P(z1)
+ P(y1 | z2) P(z2)

P(z1)=0.61
P(z2)=0.39
} Computed on the previous slide

Computing P(y1 | z1) and P(y1 | z2) using Bayes theorem:
P(y1 | z1) = P(z1 | y1) P(y1) / P(z1)
P(y1 | z1) = 0.7 x 0.84 / 0.652 = 0.90

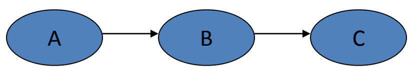
P(y1 | z2) = P(z2 | y1) P(y1) / P(z2)
P(y1 | z2) = 0.3 x 0.84 / 0.348 = 0.72

Finally plug the values in the equation at the top
P*(y1) = P(y1 | z1) P(z1) + P(y1 | z2) P(z2)
= 0.90 x 0.61 + 0.72 X 0.39
= 0.83
Confirm the results using Genie.

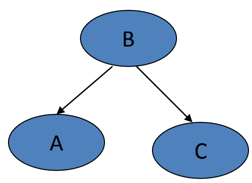
Finally compute P*(x1)

Conditional Dependence/Independence

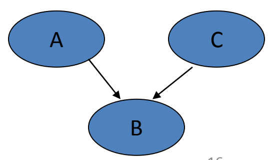
- **Causal Chains (Serial Connection)**
  - smoking causes cancer which causes dyspnoea
  - if we have evidence on B then A and C become independent



- **Common Causes (Diverging Connection)**
  - cancer is a common cause of the two symptoms, a positive XRay result and dyspnoea
  - If we have evidence on B then A and C become independent

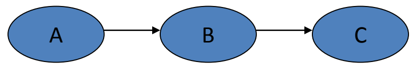


- **Common Effects (Converging Connection)**
  - Cancer is a common effect of pollution and smoking
  - If we have evidence on B then A and C become dependent



Serial Connection Example

- Consider the following BN



P(A) = 0.4

P(B|A) = 0.7
P(B|¬A) = 0.2

P(C|B) = 0.05
P(C|¬B) = 0.8

- Using the Bayesian Chain Rule, compute the joint distribution (i.e., eight values: P(A, B, C), P(A, B, ¬C), ………., P(¬A, ¬B, C) and P(¬A, ¬B, ¬C)
- Now compare
  - P(A | B, C) with P(A | B)
  - P(C | A, B) with P(C | B)

Practice Questions

- Develop a simple diverging connection style BN (as shown in the earlier slide).
- After computing the joint probability distribution, verify that
  - P(A | B, C) = P(A | B)
  - P(C | A, B) = P(C | B)

- Develop a simple converging connection style BN. After computing the joint distribution, verify that
  - P(A | C) = P(A) but P(A | B, C) != P(A | B)
  - P(C | A) = P(C) but P(C | A, B) != P(C | B)