Bayes nets

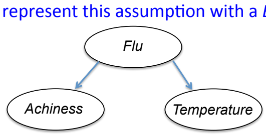
CS B553
Spring 2013

Announcements

- Readings and lecture notes online on OnCourse
  - Under the "Wiki" tab
- Assignment 1 online
  - Due next Thursday
  - Looking for a partner? Stay after class or post in the OnCourse Forum.

An example

- Say we want to decide whether someone has the flu (F) based on their temperature (T) and achiness (A)
- A, T, and F are clearly **not** independent
- But a weaker assumption of conditional independence may be appropriate, $A \perp T | F$
  - Says that A and T are independent *for a given value of F*
  - We can represent this assumption with a *Bayesian network*:

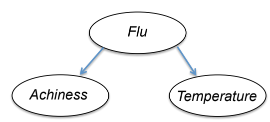


Another example

- Now we can factor P(A,T,F) as:

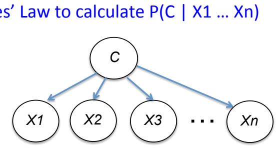
$$P(A,T,F) = P(A|F)P(T|F)P(F)$$

- To decide whether someone has the flu given observed symptoms, we'll want to compute P(F | A, T)
  - How to compute this?



Naïve Bayes model

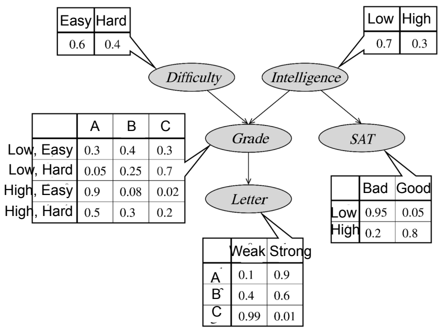
- Assuming conditional independence among observed variables is called *naïve Bayes*
  - Class label *C* we want to infer
  - Set of observable variables X1, X2, … Xn
  - Assume that observable variables are independent conditioned on the class label *C*
  - Estimate prior distribution P(C) and conditional distributions P(X1|C), …, P(Xn | C) from training data
  - Use Bayes' Law to calculate P(C | X1 … Xn)



Another example

- Suppose we want to model students in CS B553, using several random variables:
  - Intelligence (I)
  - GPA (G)
  - SAT score (S)
  - Difficulty of courses taken (D)
  - Strength of letter of recommendation (L)
- Intuitively, arrows in the BN represent direct dependencies between variables
  - Assuming these dependences, how does the joint distribution P(I,G,S,D,L) factor?

Conditional probability distributions

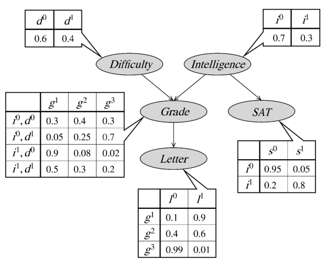


Types of reasoning in Bayes nets

- Causal reasoning
  - Probabilities “flow downwards” towards leaves of tree
- Evidential reasoning
  - Probabilities “flow upwards” away from leaves
- Intercausal reasoning
  - Knowledge about one cause affects your belief of another cause
  - Probabilities “flow” downwards and then upwards

Independencies in this example

- Is $G \perp S$?
- Is $G \perp S | I$?
- Is $L \perp I | G$?
- Is $D \perp I$?
- Is $D \perp I | G$?
- Is $G \perp L | D, I$?



Bayesian networks

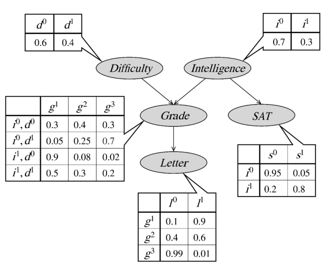
- A Bayesian network is defined by a pair (G,P), where:
  - G is a dag (directed acyclic graph), with nodes corresponding to variables {X1, X2, … Xn} and edges to direct dependencies
  - P is a probability distribution that satisfies independence assumptions induced by G
- The dag G encodes the conditional independence assumptions:

$$X_i \perp \mathrm{Nd}(X_i) \mid \mathrm{Pa}(X_i)$$

  - where Nd(Xi) is the set of non-descendants of Xi, and Pa(Xi) is the set of parents of Xi

Independencies in this example

- Is $G \perp S$?
- Is $G \perp S | I$?
- Is $L \perp I | G$?
- Is $D \perp I$?
- Is $D \perp I | G$?
- Is $G \perp L | D, I$?



$$X_i \perp \mathrm{Nd}(X_i) \mid \mathrm{Pa}(X_i)$$

(But other independencies may also hold!)

Factorization of Bayes nets

- Given a Bayes net (G,P) over variables {X1, X2 … Xn}, the joint probability distribution factors as,

$$P(X_1, \ldots, X_n) = \prod_{i=1}^{n} P(X_i \mid \mathrm{Pa}(X_i))$$

  - Proof sketch?

Independencies in Bayes nets

- We already have a set of some conditional independence relationships, given by:

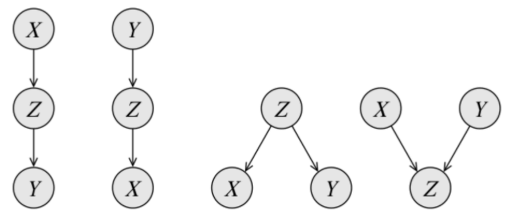
$$X_i \perp \mathrm{Nd}(X_i) \mid \mathrm{Pa}(X_i)$$

  – These are just the relationships directly defined by G; there are often others

- We'd like to derive all conditional independence relationships
  – (Why?)

For three nodes, Four cases

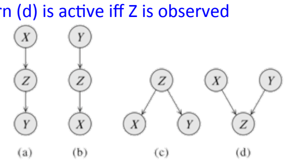
- Is $X \perp Y \mid Z$ in each case? **Is $X \perp Y$ in each case?**



(a)	(b)	(c)	(d)
Yes	Yes	Yes	No
No	No	No	Yes

Active trails

- Influence can "flow" from X to Y along *active trails*
  – Patterns (a), (b), and (c) are active iff Z is not observed
  – Pattern (d) is active iff Z is observed

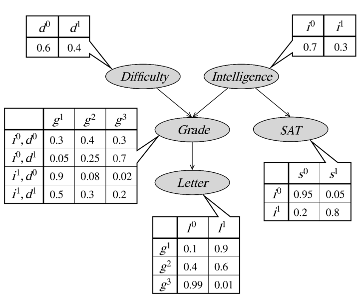


- In general, a trail is active given a set of nodes **Z** if
  – In any "v-structure" (type (d)) above, the middle node or one of its descendants is in **Z**
  – No other node along the trail is in **Z**

d-separatedness

- Sets of nodes X and Y are *d-separated* with respect to nodes Z if there is no active trail between X and Y.
- If X and Y are d-separated given Z in G, then we're guaranteed that $X \perp Y \mid Z$ in P
- If X and Y are not d-separated given Z in G, then there exists some distribution for which X and Y are dependent
- **Stronger result: For almost all possible distributions, d-separation is equivalent to conditional independence.**

Independencies in this example



d^0	d^1
0.6	0.4

i^0	i^1
0.7	0.3

	g^1	g^2	g^3
i^0, d^0	0.3	0.4	0.3
i^0, d^1	0.05	0.25	0.7
i^1, d^0	0.9	0.08	0.02
i^1, d^1	0.5	0.3	0.2

	s^0	s^1
i^0	0.95	0.05
i^1	0.2	0.8

	l^0	l^1
g^1	0.1	0.9
g^2	0.4	0.6
g^3	0.99	0.01

Sample application: Solving constraint problems

- Say someone gives us a Boolean expression like,

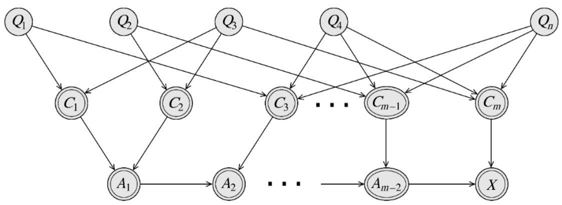
$$(Q_1 \text{ OR } Q_2 \text{ OR } Q_3) \text{ AND } (Q_4 \text{ OR } Q_5 \text{ OR } Q_6) \text{ AND } \ldots$$

  – Where Q's may repeat and may include negation
  – We'd like to decide if there exists an assignment of boolean values to the Q's such that the expression is true

- This problem can be solved using a Bayes nets
  – Use conditional probabilities to "program" AND and OR operations

CSP with a Bayes net

- Q nodes represent variables in CSP
- C nodes have CPDs that implement "OR" operations
- A nodes have CPDs that implement "AND" operations
- CSP can be satisfied iff P(X) > 0



Bad news!

- We just showed a reduction from 3-SAT to Bayes net inference!
  - I.e., if we could solve Bayes net inference efficiently, then we can solve 3-SAT efficiently
  - But 3-SAT is NP-hard
  - So Bayes net inference is NP hard also…