Learning & Inference in Graphical Models
Monday, March 30, 2011

Instructors: Raquel Urtasun (rurtasun@ttic.edu)
Tamir Hazan (tamir@ttic.edu)

- Monday, Wednesday, Friday 1:30-2:20

http://ttic.uchicago.edu/~rurtasun/courses/GraphicalModels/graphical_models.html

Bayesian Networks

- Motivation: Medical Diagnostic

Probability p(d,x1,...,xn) over symptoms and diseases:

x1=0 if patient has no fever. x1=1 if patient has fever.

x2=0 if patient does not cough. x2=1 otherwise.

d is a disease: flu, ear infection, lung infection, ...

- Given the symptoms, what is the probability of a disease?

Bayesian Networks

- Given the symptoms, what is the probability of a disease?

  p(flu | cough, fever) >? p(no flu | cough, fever)

- Problems:

1) There are exponentially many entries in the probability distribution (at least 2^n possibilities). Each entry typed need to be compared to other entries.

2) One need to marginalize out the diseases. Summing over exponentially many elements.

Independence

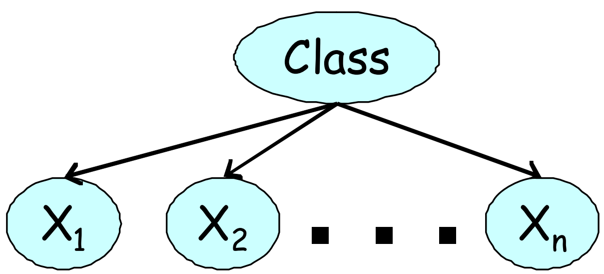
- If x1,...,xn are independent then

$$p(x_1, ..., x_n) = p(x_1) \cdots p(x_n)$$

- 2^n entries can be described by 2n numbers

Naive Bayes

- d= is a disease name
- x1,...,xn are the patient symptoms.
- Assume $X_1 \perp \cdots \perp X_n | D$
- $p(d, x_1, ..., x_n) = p(d) p(x_1|d) \cdots p(x_n|d)$



- The experts need to type D+2nD entries.

Naive Bayes

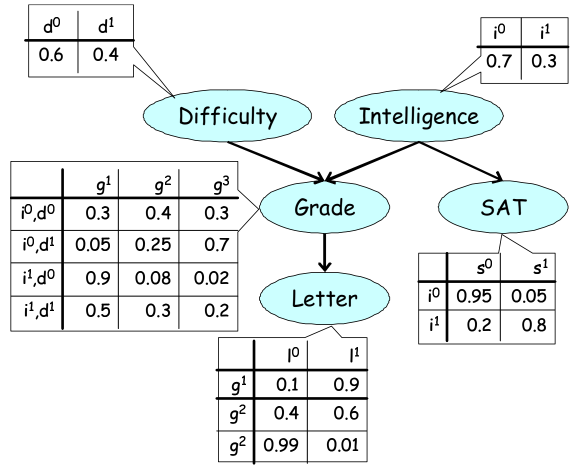
- Prediction is efficient:

$$\frac{p(d=flu|x_1,...,x_n)}{p(d=ear|x_1,...,x_n)} = \frac{p(d=flu)}{p(d=ear)} \prod_{i=1}^{n} \frac{p(x_i|d=flu)}{p(x_i|d=ear)}$$

- Statistical assumptions are too restrictive.

Bayesian Networks

- A way to describe a joint probability



$$p(I, D, G, S, L) = p(I)P(D)p(G|I, D)p(S|I)p(L|G)$$

- Chain Rule:

$$p(I, D, G, S, L) = p(I)P(D|I)p(G|I, D)p(S|I, D, G)p(L|I, D, G, S)$$

- Implicit independence statements!

Bayesian Networks

$$p(I, D, G, S, L) = p(I)P(D)p(G|I, D)p(S|I)p(L|G)$$

- Chain Rule:

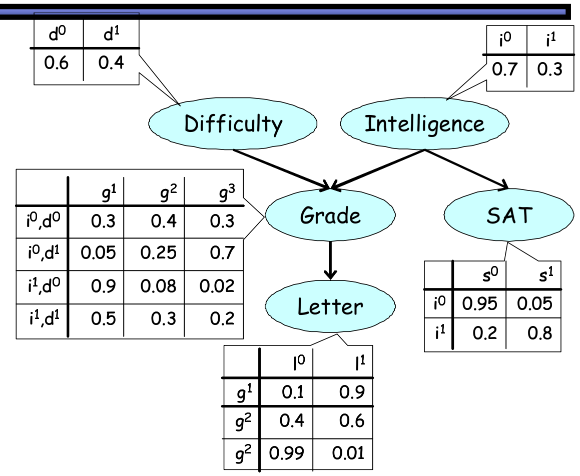
$$p(I, D, G, S, L) = p(I)P(D|I)p(G|I, D)p(S|I, D, G)p(L|I, D, G, S)$$

- Independence statements:

$$D \perp I, \quad S \perp \{D, G\}|I, \quad L \perp \{I, D, S\}|G$$

Bayesian Networks

$$p(I, D, G, S, L) = p(I)P(D)p(G|I, D)p(S|I)p(L|G)$$



- Independence statements:

$$D \perp I, \quad S \perp \{D, G\} | I, \quad L \perp \{I, D, S\} | G$$

- Claim: A variable is independent from its non-descendants given its parents.
- Result: There are more independence in the graph, e.g. $S \perp \{D, G, L\} | I$

Bayesian Networks

$p(I, D, G, S, L) = p(I)P(D)p(G|I, D)p(S|I)p(L|G)$

- proof of: $S \perp \{D, G, L\} | I$

$$p(D, G, S, L|I) \quad =? \;\; p(S|I)p(D, G, L|I)$$

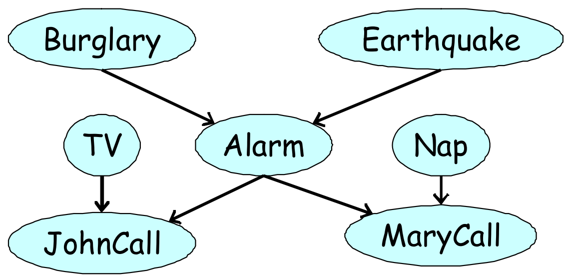
$$p(D, G, L|I) = \frac{p(I, D, G, L)}{p(I)} = \frac{\sum_S p(I, D, G, S, L)}{p(I)}$$

$$p(S|I) = \frac{p(I, S)}{p(I)} = \frac{\sum_{G,D,L} p(I, D, G, S, L)}{p(I)}$$

Bayesian Networks

- For general

directed graphs:



$$p(x_1, ..., x_n) = \prod_{i=1}^{n} p(x_i | x_{parents(i)})$$

- Bayesian network encodes conditional independencies

$$x_i \perp x_{non-descendants(i)} | x_{parents(i)}$$

Independency Maps

- given a distribution p(x1,...,xn) , we denote by I(p) its independency map, i.e. all statements of the form
$$X_I \perp X_J | X_K$$
for $I, J, K \subset \{1, ..., n\}$

- The directed graph gives some of the independencies of the distribution, through separation in directed graphs.