Probabilistic Reasoning

Unit # 1

Course Outline

- Knowledge Representation
- Modeling and Reasoning with Bayesian Networks
- Handling Knowledge Acquisition Issues
- Belief Updating in Singly and Multiply Connected Networks
- Dynamic Bayesian Networks and their Variants
- Parameter and Structure Learning
- Papers Reading (a lot of them!)

Useful Information

- Course Wiki
  - http://cse655fall2014.wikispaces.com
- Software Tools
  - **GeNIe** (http://genie.sis.pitt.edu/)
  - Netica (http://www.norsys.com/netica.html)
  - Hugin (http://www.hugin.com/)
  - Bayesian PowerConstructor (http://webdocs.cs.ualberta.ca/~jcheng/bnpc.htm)
  - **IBAYes** (http://ailab.iba.edu.pk/ibayes.html)

Reference Books

- **K.B. Korb and A.E. Nicholson, Bayesian Artificial Intelligence, Chapman & Hall/CRC, 2010.**
- R.E. Neapolitan, Learning Bayesian Networks, Prentice Hall, 2003.
- U.B. Kjaerulff and A.L. Madsen, Bayesian Networks and Influence Diagrams: A Guide to Construction and Analysis, Springer, 2007.
- A. Mittal, Bayesian Network Technologies: Applications and Graphical Models, IGI Publishing, 2007.
- F.V. Jensen and T.D. Nielsen, Bayesian Networks and Decision Graphs, Springer, 2007.
- O. Pourret, P. Naïm, and B. Marcot, Bayesian Networks: A Practical Guide to Applications, Wiley, 2008.
- **R.E. Neapolitan and X. Jiang, Probabilistic Methods for Bioinformatics: With an Introduction to Bayesian Networks, Morgan Kaufmann Publishers, 2009.**

Marks Distribution (Tentative)

• 2 Midterms (20 marks each)	40
• Final	40
• Term Paper	10
• Assignment/Presentation	10

Uncertainty

- Three distinct forms of uncertainty which an intelligent system need to cope with:
  - Ignorance
  - Physical Randomness
  - Vagueness

Bayesianism

- Bayesianism is the philosophy that asserts that in order to understand human opinion as it ought to be, constrained by ignorance and uncertainty, the probability calculus is the single most important tool for representing appropriate strengths of belief.

Conditional Independence

- Two events A and B are independent if knowing that A has happened does not say anything about B happening.

$$P(A\ B) = P(A)\ P(B)$$
$$P(A \mid B) = P(A)$$

- Two events A and B are conditionally independent given a third event C precisely if the occurrence or non-occurrence of A and B are independent events in their conditional probability distribution given C.

$$P(A\ B \mid C) = P(A \mid C)\ P(B \mid C)$$
$$P(A \mid B\ C) = P(A \mid C)$$

Bayes Theorem

- $P(A \mid B) = \frac{P(B \mid A)\,P(A)}{P(B)}$

  $= \frac{P(B \mid A)\,P(A)}{P(B|A)P(A)+P(B|\neg A)P(\neg A)}$
- P(A) is the prior probability and P(A | B) is the posterior probability.
- Suppose events $A_1, A_2, \ldots\ldots, A_k$ are mutually exclusive and exhaustive; i.e., exactly one of the events must occur. Then for any event B:

  $P(A_i \mid B) = \frac{P(B \mid A_i)\,P(A_i)}{\sum P(B \mid A_i)\,P(A_i)}$

Example I

- According to American Lung Association, 7% of the population has lung cancer. Of these people having lung disease, 90% are smokers; and of those not having lung disease, 25.3% are smokers.
- Determine the probability that a randomly selected smoker has lung cancer.

Example I Solution

- Let L = Lung Cancer, S = Smoker
- Given that
  - P(L) = 0.07
  - P(S | L) = 0.90 P(~S | L) = 0.10
  - P(S | ~L) = 0.253 P(~S | ~L) = 0.747
- Find probability, P(L | S)

$$P(L \mid S) = \frac{P(S \cap L)}{P(S)} = \frac{P(S \mid L)P(L)}{P(S|L)P(L) + P(S|{\sim}L)P({\sim}L)}$$

$$P(L|S) = \frac{0.9 \ x \ 0.07}{0.9 \ x \ 0.07 + 0.253 \ x \ 0.93}$$

Example II

- Assume that about 1 in 1000 individuals in a given organization have committed a security violation.
- Assume that the **sensitivity** of a routine screening polygraph is about 85%. That is, the probability that the polygraph report will indicate a concern is about 85% if the individual has committed a security violation.
- Assume the **specificity** of the polygraph is about 80%. That is, if the individual has not committed a security violation, there is about an 80% chance that the polygraph report will not indicate a concern.
- What is the posterior probability that an individual whose polygraph report indicates a concern has committed a security violation?

Example II Solution

- Let
  - S = Security Violation Committed,
  - T = Test Positive
- Given that
  - P(S) = 0.001
  - P(T | S) = 0.85 P(~T | S) = 0.15
  - P(T | ~S) = 0.20 P(~T | ~S) = 0.80
- Find probability, P(S | T)

$$P(S|T) = \frac{P(T|S)P(S)}{P(T|S)P(S) + P(T|\sim S)P(\sim S)}$$

$$P(T|S) = \frac{0.85\ x\ 0.001}{0.85\ x\ 0.001 + 0.20\ x\ 0.999}$$

Example III

- Suppose that Bob can decide to go to work by one of three modes of transportation, car, bus, or commuter train. Because of high traffic, if he decides to go by car, there is a 50% chance he will be late. If he goes by bus, which has special reserved lanes but is sometimes overcrowded, the probability of being late is only 20%. The commuter train is almost never late, with a probability of only 1%, but is more expensive than the bus.
- Suppose that Bob is late one day, and his boss wishes to estimate the probability that he drove to work that day by car. Since he does not know which mode of transportation Bob usually uses, he gives a prior probability of 1/3 to each of the three possibilities. What is the boss' estimate of the probability that Bob drove to work?

Chain Rule

- Given variables $X_1, X_2, \ldots.., X_n$, the chain rule is

$$P(x_1, x_2, \ldots.., x_n) = P(x_n \mid x_{n-1}, \ldots.., x_2, x_1) \ldots\ldots\ldots P(x_2 \mid x_1) P(x_1)$$

Example (Source Neapolitan, 2009)

- Let S = Sex, H = Height and W = Wage

Case	Sex	Height (inches)	Wage ($)
1	female	64	30,000
2	female	64	30,000
3	female	64	40,000
4	female	64	40,000
5	female	68	30,000
6	female	68	40,000
7	male	64	40,000
8	male	64	50,000
9	male	68	40,000
10	male	68	50,000
11	male	70	40,000
12	male	70	50,000

Example (Cont'd)

s	$P(s)$
female	$1/2$
male	$1/2$

h	$P(h)$
64	$1/2$
68	$1/3$
70	$1/6$

w	$P(w)$
30,000	$1/4$
40,000	$1/2$
50,000	$1/4$

The joint distribution of S and H is as follows:

s	h	$P(s,h)$
female	64	$1/3$
female	68	$1/6$
female	70	0
male	64	$1/6$
male	68	$1/6$
male	70	$1/6$

Example (Cont'd) Marginal and Joint Distribution

s \ h	64	68	70	Distribution of S
female	$1/3$	$1/6$	0	$1/2$
male	$1/6$	$1/6$	$1/6$	$1/2$
Distribution of H	$1/2$	$1/3$	$1/6$	

Example (Cont'd)
Chain Rule

$$P(s,h,w) = P(w|h,s)P(h|s)P(s).$$

There are eight combinations of values of the three random variables. The table that follows shows that the equality holds for two of the combination.

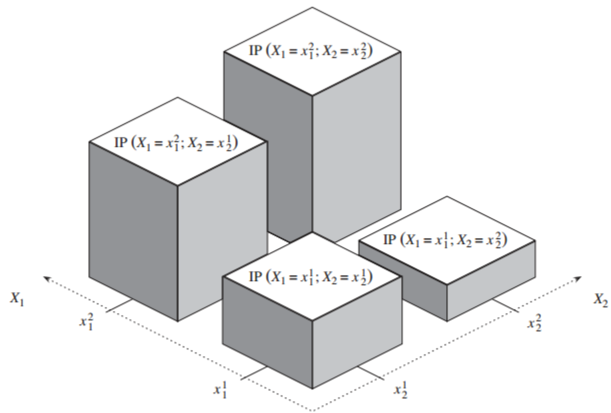
s	h	w	$P(s,h,w)$	$P(w\|h,s)P(h\|s)P(s)$
female	64	30,000	$\frac{1}{6}$	$\left(\frac{1}{2}\right)\left(\frac{2}{3}\right)\left(\frac{1}{2}\right) = \frac{1}{6}$
female	64	40,000	$\frac{1}{12}$	$\left(\frac{1}{2}\right)\left(\frac{1}{3}\right)\left(\frac{1}{2}\right) = \frac{1}{12}$

It is left as an exercise to show that the equality holds for the other six combinations.

Exercise 1
(Source: Neapolitan, 2009)

- Are H and W independent?
- Are H and W conditionally independent given S?

Joint Probability Distribution of Two Variables (Source: Pourret et al)



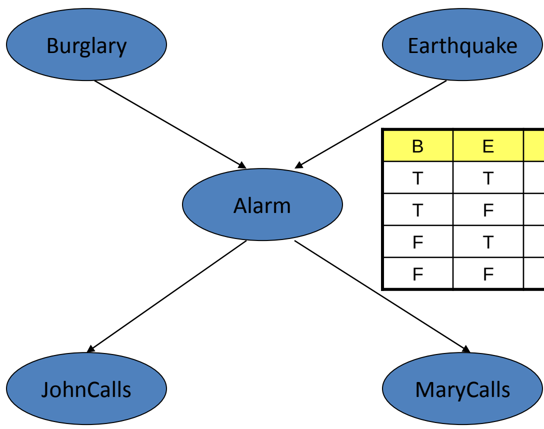
Exercise 2 (Source: Neapolitan, 2009)

- A forgetful nurse is supposed to give Mr. Nguyen a pill each day. The probability that the nurse will forget to give the pill on a given day is 0.3. If Mr. Nguyen receives the pill, the probability he will die is 0.1. If he does not receive the pill, the probability he will die is 0.8. Mr. Nguyen died today.
- Use Bayes' Theorem to compute the probability that the nurse forgot to give him the pill.

Exercise 3
(Source: Neapolitan, 2009)

- An oil well might be drilled on Professor Neapolitan's farm in Texas. Based on what has happened on similar farms, we judge the probability of oil being present to be 0.5, the probability of only natural gas being present to be 0.2, and the probability of neither being present to be 0.3.
- If oil is present, a geological test will give a positive result with probability 0.9; if only natural gas is present, it will give a positive result with probability 0.3; and if neither is present, the test will be positive with probability 0.1.
- Suppose the test comes back positive. Use Bayes' Theorem to compute the probability that oil is present.

Earthquake Example (Pearl)



P(B)
0.1

P(E)
0.02

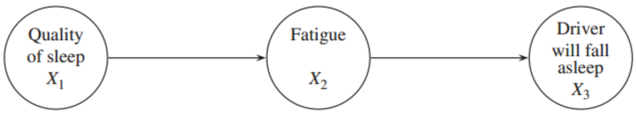
B	E	P(A|B,E)
T	T	0.95
T	F	0.94
F	T	0.29
F	F	0.001

A	P(J|A)
T	0.9
F	0.05

A	P(M|A)
T	0.7
F	0.01

Example (Source: Pourret et al.)

- A lorry driver is due to make a 600-mile trip. To analyze the risk of his falling asleep while driving, let us consider whether (1) he sleeps 'well' (more than seven hours) on the night before and (2) he feels tired at the beginning of the trip.



Bayesian Networks

- Two problems with using full joint distributions for probabilistic models:
  - Unless there are only a few variables, the joint is WAY too big to represent explicitly
  - Hard to estimate anything empirically about more than a few variables at a time

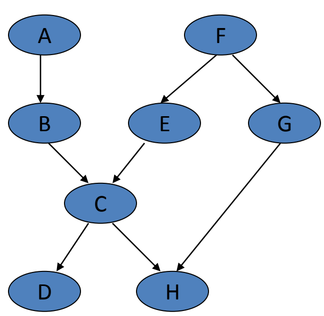
- Bayesian Networks are a technique for describing complex joint distributions using a bunch of simple, local distributions
  - We describe how variables locally interact
  - Local interactions chain together to give global, indirect interactions

Bayesian Networks

- A BN is a Directed Acyclic Graph (DAG) in which:
  - A set of random variables makes up the nodes in the network.
  - A set of directed links or arrows connects pairs of nodes.
  - Each node has a conditional probability table that quantifies the effects the parents have on the node.
- The intuitive meaning of an arrow from a parent to a child is that the parent directly influences the child.
- The direction of this influence is often taken to represent casual influence.
- These influences are quantified by conditional probabilities.

Bayesian Networks (Cont'd)

- A problem domain is modeled by a list of variables $X_1, X_2, \ldots, X_n$.
- Knowledge about the problem domain is represented by a joint probability $P(X_1, X_2, \ldots, X_n)$.
- General probability distribution of 8 variables with 2 states each has 2^8 = 256 possible values and 2^8 -1 probabilities need to be specified.
- Assumes that each node is conditionally independent of all its non-descendants given its parents.
- Product of all conditional probabilities is the joint probability of all variables.
  - $P(X_1, X_2, \ldots, X_n = \prod P(X_i \mid parents(X_i))$



18 probabilities are required to specify the joint distribution

GeNIe Demo