Bayesian Belief Networks

Prof. Navneet Goyal
BITS, Pilani

Bayesian Belief Networks

- Naïve BC assumes Class Conditional Independence
- This assumption simplifies computations
- When this assumption holds true, Naïve BC is most accurate compared to all other classifiers
- In real problems, dependencies do exist between variables
- 2 methods to overcome this limitation of NBC
  - Bayesian networks, that combine Bayesian reasoning with causal relationships between attributes
  - Decision trees, that reason on one attribute at the time, considering 'most important' attributes first

Bayesian Belief Networks (BBN)

- Instead of requiring all attributes to be conditionally independent given the class, we specify which pair of attributes are conditionally independent.
- Flexible way to modeling class conditional probabilities $P(X|Y)$
- BBN provides a graphical representation of the probabilistic relationships among a set of RVs
  - Directed Acyclic Graph (DAG)
  - Probability table

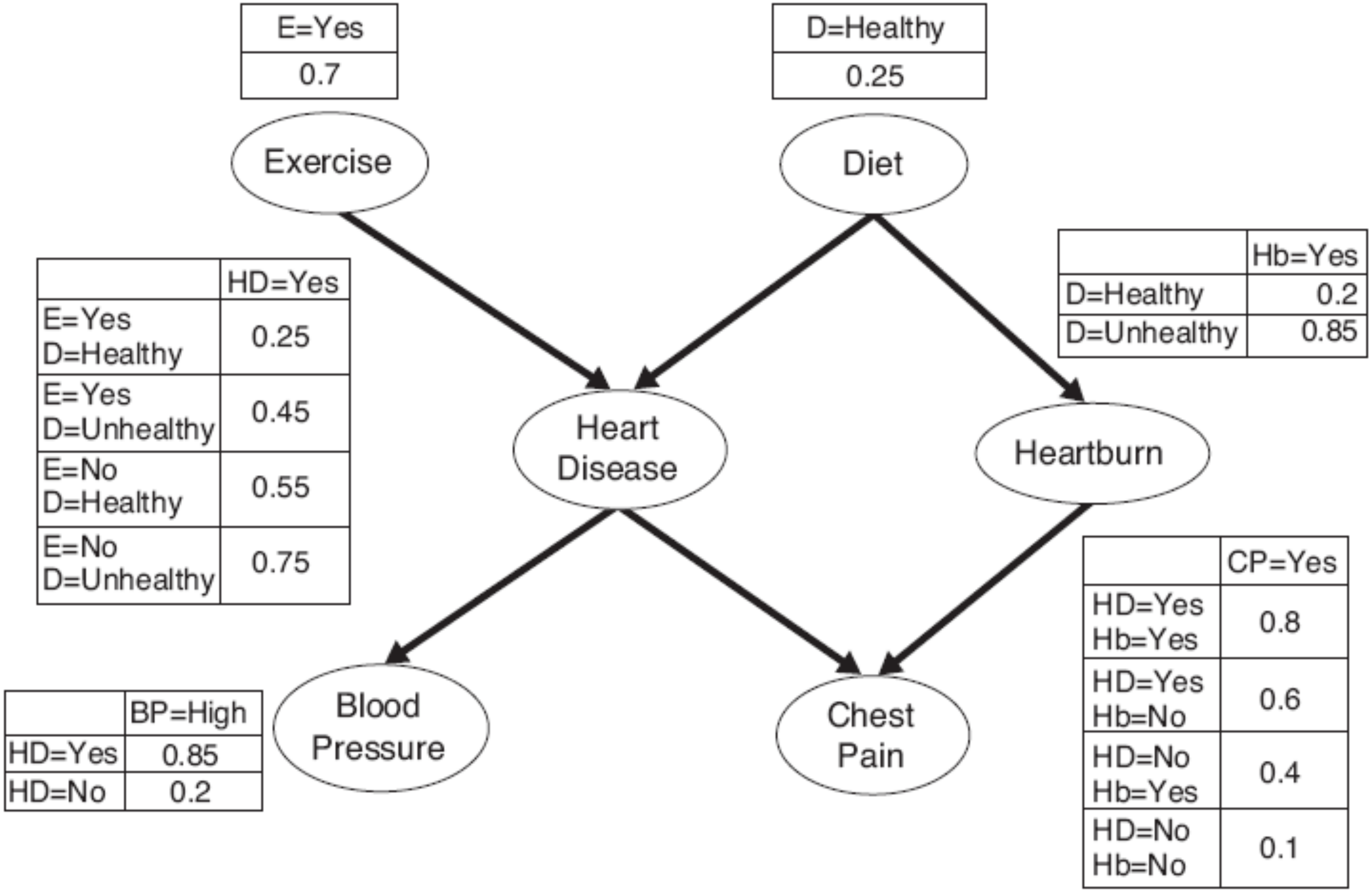


Figure 5.13. A Bayesian belief network for detecting heart disease and heartburn in patients.

- Fig. taken form Tan & kumar – Intro. to DM book

Conditional Independence

- Let X, Y, & Z denote three set of random variables. The variables in X are said to be conditionally independent of Y, given Z if
  - $P(X|Y,Z) = P(X|Z)$
- Rel. bet. a person’s arm length and his/her reading skills!!
- One might observe that people with longer arms tend to have higher levels of reading skills
- How do you explain this rel.?

Conditional Independence

- Can be explained through a confounding factor, AGE
- A young child tends to have short arms and lacks the reading skills of an adult
- If the age of a person is fixed, then the observed rel. between arm length and reading skills disappears
- We can this conclude that arm length and reading skills are conditionally independent when the age variable is fixed

P(reading skills| long arms,age) = P(reading skills|age)

Conditional Independence

$$
\begin{aligned}
P(X,Y|Z) &= P(X,Y,Z)/P(Z) \\
&= P(X,Y,Z)/P(Y,Z) \times P(Y,Z)/P(Z) \\
&= P(X|Y,Z) \times P(Y|Z) \\
&= P(X|Z) \times P(Y|Z)
\end{aligned}
$$

This explains the Naïve Bayesian:

$P(X|Ci) = P(x1, x2, x3,\ldots,xn|C) = \prod P(x_k|C)$

Bayesian Belief Networks

- Belief Networks
- Bayesian Networks
- Probabilistic Networks

Bayesian Belief Networks

♦ Conditional Independence (CI) assumption made by NBC may be too rigid

♦ Specially for classification problems in which attributes are somewhat correlated

♦ We need a more flexible approach for modeling the class conditional probabilities

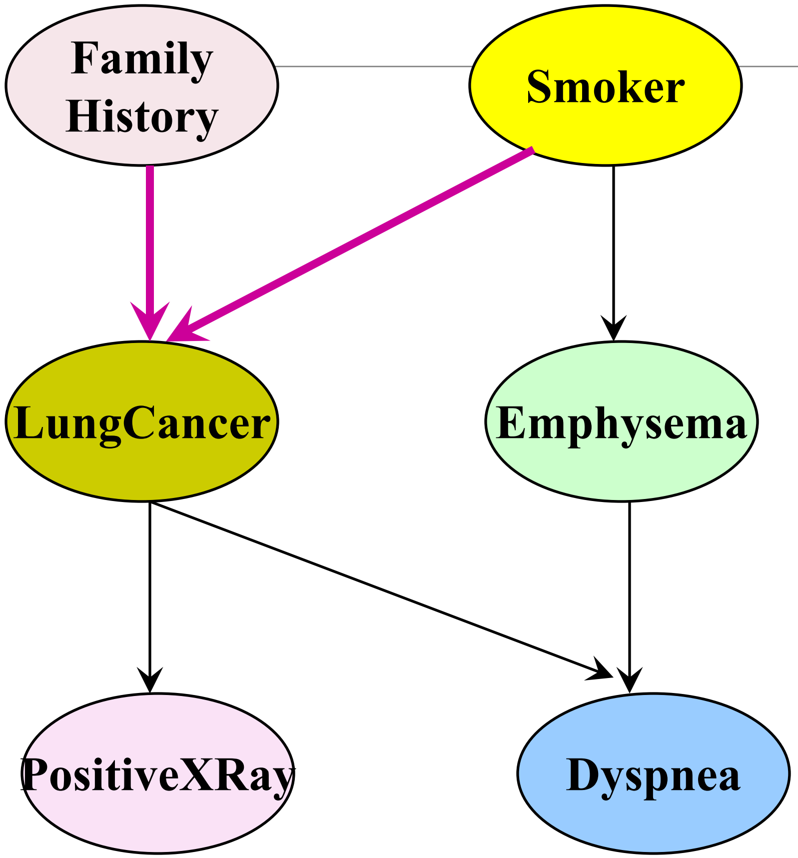
$$P(X/Ci) = P(x1, x2, x3, \ldots, xn|C)$$

♦ instead of requiring that all the attributes be CI given the class, BBN allows us to specify which pair of attributes are CI

Bayesian Belief Networks

- BBN is an expert system that captures all existing knowledge

Bayesian Belief Networks



Bayesian Belief Networks

	(FH, S)	(FH, ~S)	(~FH, S)	(~FH, ~S)
LC	0.8	0.5	0.7	0.1
~LC	0.2	0.5	0.3	0.9

The conditional probability table for the variable LungCancer

Example Taken from Han et. al. book on Data Mining, 3e

Bayesian Belief Networks

- 6 boolean variables
- Arcs allow representation of causal knowledge
- Having lung cancer is influenced by family history and smoking (risk factors)
- Positive Xray is ind. of whether the patient has a FH or if he/she is a smoker given that we know that the patient has lung cancer
- Once we know the outcome of Lung Cancer, FH & Smoker do not provide any additional info. about Positive Xray

Bayesian Belief Networks

- Lung Cancer is CI of Emphysema, given its parents, FH & Smoker
- BBN has a Conditional Probability Table (CPT) for each variable in the DAG
- CPT for a variable Y specifies the conditional distribution P(Y|parents(Y))

	(FH, S)	(FH, ~S)	(~FH, S)	(~FH, ~S)
LC	0.8	0.5	0.7	0.1
~LC	0.2	0.5	0.3	0.9

CPT for LungCancer

P(LC=Y|FH=Y,S=Y) = 0.8

P(LC=N|FH=N,S=N) = 0.9

Bayesian Belief Networks

- A node within the network can selected as an ‘output’ node, representing a class label attribute
- Rather than returning a single class label, the classification process can return a probability distribution that gives the probability of each class
- Probability of evidence queries
  - what is the prob. that an individual will have LC given that it has both +ve Xray and Dyspnea?
- Probability of explanation queries
  - which group of population is most likely to have both +ve Xray and Dyspnea?

Bayesian Belief Networks

- node A is a parent or immediate predecessor of node X, and node X is a descendant of node A, if ∃ a directed arc from node A to node X
- The intrinsic relationship among the variables in a Bayesian network is as follows:
  - Each variable in a Bayesian network is conditionally independent of its non-descendants in the network given its parents

$$p(X_1 = x_1, X_2 = x_2, \ldots, X_m = x_m) = \prod_{i=1}^{m} p(X_i = x_i | parents(X_i))$$

Note that child node probability depends only on its parents

BBN: Clothing Purchase Example

- A clothes retailer operates two outlets:
  New York and Los Angeles

- Both producing sales throughout the four seasons

- The retailer is interested in probabilities concerning three articles of clothing in particular: warm coats, business shirts, and Bermuda shorts.

- Questions of interest include the fabric weight of the article of clothing (light, medium, or heavy) and the color of the article (bright, neutral, or dark).

BBN: Clothing Purchase Example

- Variables involved:
  - Season
  - Location
  - Clothing purchase
  - Fabric weight
  - Color

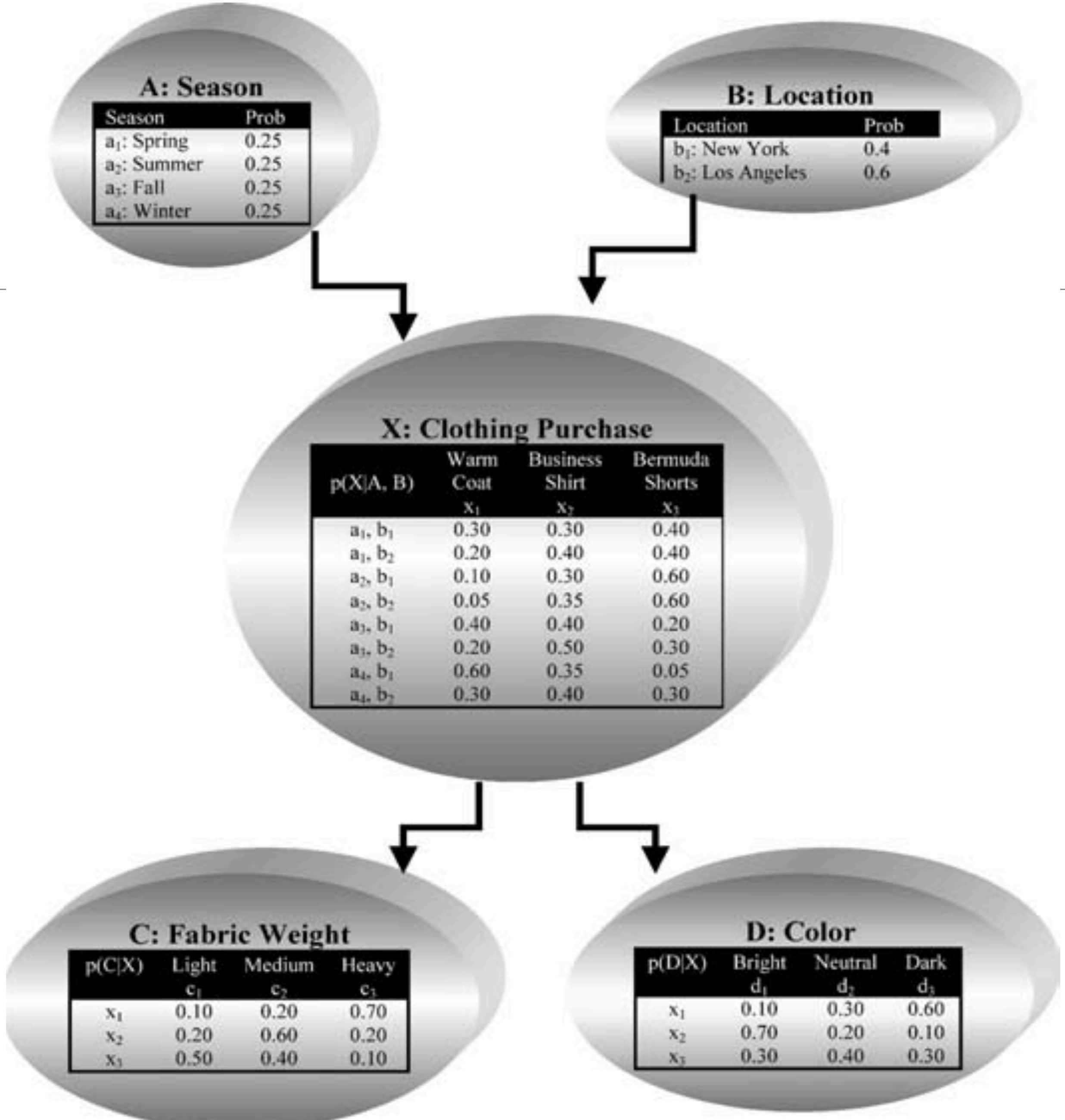
- Model this problem using BBN

BBN: Clothing Purchase Example

- What is the dependence relationship among these variables?
- For example, does the season of the year depend on the color of the clothes purchased?
  - Certainly not, since a customer's purchase of some bright clothing doesn't mean that spring is here, for example, although the customer may wish it so.
- The season of the year does not depend on any of the other variables
- Similarly, location does not depend on the other variables, and is therefore placed at the top of the network.
- Since the fabric weight and the color of the clothing are not known until the article is purchased, the node for the variable clothing purchase is inserted next into the network, with arcs to each of the fabric weight and color nodes.

BBN: Clothing Purchase Example

- The second consideration for constructing a Bayesian network is to specify all of the entries in the probability tables for each node
- The probabilities in the season node table indicate that clothing sales for this retail establishment are uniform throughout the four seasons.
- The probabilities in the location node probability table show that 60% of sales are generated from Los Angeles store and 40% from New York store.
- Note that these two tables need not supply conditional probabilities, since the nodes are at the top of the network


A: Season
Season Prob
a1: Spring 0.25
a2: Summer 0.25
a3: Fall 0.25
a4: Winter 0.25
B: Location
Location Prob
b1: New York 0.4
b2: Los Angeles 0.6
X: Clothing Purchase
p(X|A, B) Warm Coat x1 Business Shirt x2 Bermuda Shorts x3
a1, b1 0.30 0.30 0.40
a1, b2 0.20 0.40 0.40
a2, b1 0.10 0.30 0.60
a2, b2 0.05 0.35 0.60
a3, b1 0.40 0.40 0.20
a3, b2 0.20 0.50 0.30
a4, b1 0.60 0.35 0.05
a4, b2 0.30 0.40 0.30
C: Fabric Weight
p(C|X) Light c1 Medium c2 Heavy c3
x1 0.10 0.20 0.70
x2 0.20 0.60 0.20
x3 0.50 0.40 0.10
D: Color
p(D|X) Bright d1 Neutral d2 Dark d3
x1 0.10 0.30 0.60
x2 0.70 0.20 0.10
x3 0.30 0.40 0.30

BBN: Clothing Purchase Example

Conditional Independence

- Color is conditionally independent of season given clothing purchased.
- Color is conditionally independent of fabric weight given clothing purchased.
- Fabric weight is conditionally independent of color given clothing purchased.
- Fabric weight is conditionally independent of location given clothing purchased.
- Fabric weight is conditionally independent of season given clothing purchased.
- Note that we could say that season is conditionally independent of location given its parents.
  - But since season has no parents in the Bayes net, this means that season and location are (unconditionally) independent.

BBN: Clothing Purchase Example

Finding Probabilities using BBN

- Find the probability that light-fabric neutral-colored Bermuda shorts were purchased in New York in the winter

$$
\begin{aligned}
p(A &= a_4, B = b_2, C = c_1, D = d_2, X = x_3) \\
&= p(A = a_4)p(B = b_2)p(X = x_3|A = a_4 \cap B = b_2)p(C = c_1|X = x_3) \\
&\quad p(D = d_2|X = x_3)p(A = a_4, B = b_1, C = c_1, D = d_2, X = x_3) \\
&= p(A = a_4)p(B = b_1)p(X = x_3|A = a_4 \cap B = b_1)p(C = c_1|X = x_3) \\
&\quad p(D = d_2|X = x_3) \\
&= p(season = winter)p(location = New\ York) \\
&\quad \cdot p(clothing = shorts \mid season = winter \text{ and } location = New\ York) \\
&\quad \cdot p(fabric = light \mid clothing = shorts)p(color = neutral \mid clothing = shorts) \\
&= 0.25(0.4)(0.05)(0.50)(0.40) = 0.001
\end{aligned}
$$