SC4000/CZ4041/CE4041: Machine Learning

Lesson 4: Bayesian Belief Networks

Kelly KE

School of Computer Science and Engineering, NTU, Singapore

Acknowledgements: some figures are adapted from the lecture notes of the books "Introduction to Data Mining" (Chap. 5).

Bayesian Classifiers: Recall

- Estimate $P(y|\boldsymbol{x})$ via Bayes rule:

$$P(y=c|\boldsymbol{x}) = \frac{P(\boldsymbol{x}, y=c)}{P(\boldsymbol{x})} = \frac{P(\boldsymbol{x}|y=c)P(y=c)}{P(\boldsymbol{x})}$$

- Make predictions based on maximum posterior:

$$y^* = c^* \text{ if } c^* = \arg\max_c P(y=c|\boldsymbol{x})$$

the 0/1 loss

$$\Updownarrow$$

$$y^* = c^* \text{ if } c^* = \arg\max_c \boxed{P(\boldsymbol{x}|y=c)}\boxed{P(y=c)}$$

Easy to estimate

Still difficult to estimate. $\boldsymbol{x}$ contains many input variables. Some are discrete, and others are continuous.

Naïve Bayes Classifier

- To make estimation of $P(\boldsymbol{x}|y)$ from training data tractable, Naïve Bayes classifiers assume features are <u>conditionally independent</u> given the class label:

$$P(\boldsymbol{x}|y=c)=\prod_{i=1}^{d}P(x_i|y=c)$$

where $\boldsymbol{x}=[x_1,x_2,\dots,x_d]$

- The conditional independence assumption may not hold in practice.
  - Correlated features can degrade the performance.

An Example

- Suppose the probability of a person having a specific disease D is 50%.
- There are two medical tests, T_1 and T_2, of binary values (positive or negative). The outcomes of T_2 are perfectly positively correlated with T_1 if a person has the disease, but are independent of T_1 if the person does not have the disease.

When a person has the disease, the outcomes of T_1 and T_2 are both positive (or negative).

- If a person has the disease, the probabilities of tests T_1 and T_2 being negative are 40%, respectively.
- If a person does not have the disease, the probabilities of T_1 and T_2 being negative are 60% and 65%, respectively.

An Example (cont.)

- If the two tests T_1 and T_2 are both negative for a particular patient, diagnose whether the patient has the disease?

Variables Definition

- Let X_1 denote the outcome of T_1
  - $X_1 = 1$: positive
  - $X_1 = 0$: negative
- Let X_2 denote the outcome of T_2
  - $X_2 = 1$: positive
  - $X_2 = 0$: negative
- Let Y denote whether a person has the disease D
  - $Y = 1$: yes
  - $Y = 0$: no

Probabilities

- Suppose that the probability of a person having a specific disease D is 50%.

$$P(Y=0)=50\% \text{ and } P(Y=1)=50\%$$

- The outcomes of T_2 are perfectly positively correlated with T_1 if a person has the disease.
  - Given that a person has the disease, if the outcome of T_1 is positive then the outcome of T_2 is always positive, and if the outcome of T_1 is negative then the outcome of T_2 is always negative.

$$P(X_1=1,X_2=1|Y=1)=\frac{|(X_1=1)\wedge(X_2=1)\wedge(Y=1)|}{|Y=1|}$$

$$=\frac{|(X_1=1)\wedge(Y=1)|}{|Y=1|}=P(X_1=1|Y=1)$$

$$P(X_1=0,X_2=0|Y=1)=P(X_1=0|Y=1)$$

Probabilities (cont.)

- The outcomes of T_2 are perfectly positively correlated with T_1 if a person has the disease, but are independent of T_1 if the person does not have the disease.
  - Given that a person does not have the disease, the outcomes of T_1 and T_2 are independent.

$$P(X_1 = 1, X_2 = 1 | Y = 0) = P(X_1 = 1 | Y = 0) P(X_2 = 1 | Y = 0)$$

$$P(X_1 = 1, X_2 = 0 | Y = 0) = P(X_1 = 1 | Y = 0) P(X_2 = 0 | Y = 0)$$

$$P(X_1 = 0, X_2 = 1 | Y = 0) = P(X_1 = 0 | Y = 0) P(X_2 = 1 | Y = 0)$$

$$P(X_1 = 0, X_2 = 0 | Y = 0) = P(X_1 = 0 | Y = 0) P(X_2 = 0 | Y = 0)$$

  - The conditional independence assumption in Naïve Bayes Classifiers holds when a person does not have the disease.

Probabilities (cont.)

- If a person has the disease, the probabilities of tests T_1 and T_2 being negative are 40%, respectively.

$$P(X_1 = 0|Y = 1) = 0.4 \Longrightarrow P(X_1 = 1|Y = 1) = 0.6$$

$$P(X_2 = 0|Y = 1) = 0.4 \Longrightarrow P(X_2 = 1|Y = 1) = 0.6$$

- If a person does not have the disease, the probabilities of T_1 and T_2 being negative are 60% and 65%, respectively.

$$P(X_1 = 0|Y = 0) = 0.6 \Longrightarrow P(X_1 = 1|Y = 0) = 0.4$$

$$P(X_2 = 0|Y = 0) = 0.65 \Longrightarrow P(X_2 = 1|Y = 0) = 0.35$$

- Given a patient with $X_1 = 0,$ and $X_2 = 0$, to estimate

$$P(Y = 0|X_1 = 0, X_2 = 0) \text{ and } P(Y = 1|X_1 = 0, X_2 = 0)$$

Using a Naïve Bayes Classifier

$$P(Y=0|X_1=0,X_2=0) = \frac{\boxed{P(X_1=0,X_2=0|Y=0)}P(Y=0)}{P(X_1=0,X_2=0)}$$ Bayes rule

Using Naïve Bayes assumption

$$= \frac{\boxed{P(X_1=0|Y=0)P(X_2=0|Y=0)}P(Y=0)}{P(X_1=0,X_2=0)}$$

$$= \frac{0.6 \times 0.65 \times 0.5}{P(X_1=0,X_2=0)} = \frac{0.195}{P(X_1=0,X_2=0)}$$

Prediction: $Y = 0$

$$P(Y=1|X_1=0,X_2=0) = \frac{\boxed{P(X_1=0,X_2=0|Y=1)}P(Y=1)}{P(X_1=0,X_2=0)}$$ Bayes rule

Using Naïve Bayes assumption

$$= \frac{\boxed{P(X_1=0|Y=1)P(X_2=0|Y=1)}P(Y=1)}{P(X_1=0,X_2=0)}$$

$$= \frac{0.4 \times 0.4 \times 0.5}{P(X_1=0,X_2=0)} = \frac{0.08}{P(X_1=0,X_2=0)}$$

Features are Correlated When $Y = 1$

- However, because X_1 and X_2 are perfectly positively correlated when $Y = 1$,

$$P(X_1 = 0, X_2 = 0|Y = 1) = P(X_1 = 0\ |Y = 1) = 0.4$$

$$P(Y = 1|X_1 = 0, X_2 = 0) = \frac{\boxed{P(X_1 = 0, X_2 = 0|Y = 1)}P(Y = 1)}{P(X_1 = 0, X_2 = 0)}$$

Perfectly correlated

$$= \frac{\boxed{P(X_1 = 0\ |Y = 1)}P(Y = 1)}{P(X_1 = 0, X_2 = 0)} = \frac{0.4 \times 0.5}{P(X_1 = 0, X_2 = 0)}$$

- X_1 and X_2 are independent if the person does not have the disease.

$$P(Y = 0|X_1 = 0, X_2 = 0) = \frac{P(X_1 = 0|Y = 0)P(X_2 = 0|Y = 0)P(Y = 0)}{P(X_1 = 0, X_2 = 0)}$$

$$= \frac{0.195}{P(X_1 = 0, X_2 = 0)}$$

Prediction: $Y = 1$

Bayesian Belief Networks

- A more general approach to modeling the independence and conditional independence among $\boldsymbol{x}$ and y, s.t. the computation of $\boxed{P(\boldsymbol{x}, y)} = P(\boldsymbol{x}|y)\mathrm{P}(y)$ is tractable.
  - Suppose all features are discrete (if there are both continuous and discrete features, the estimation is much more difficult).
- Representation: a Bayesian network provides a graphical representation of the probabilistic relationships among a set of random variables including features and output class.
- Two key elements:
  - A directed acyclic graph (DAG) encoding the dependence relationships among a set of variables
  - A probability table associating each node to its immediate parent nodes

A DAG Example

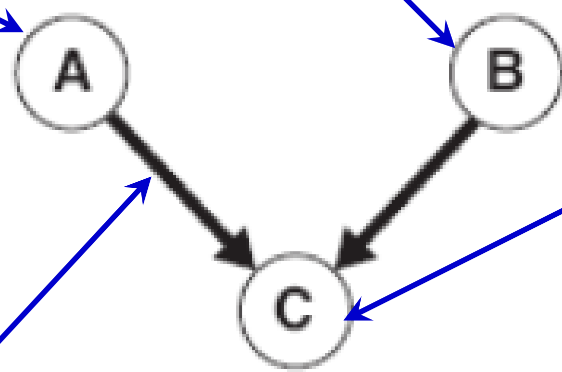
$P(A|B) = P(A)$

- Consider three random variables A, B and C, where A and B are independent variables and each has a direct influence on a third variable, C

$P(A,B) = \boxed{P(A)}P(B)$

If there is a directed arc from B to C, then B is the **parent** of C, and C is the **child** of B

node ➔ variable

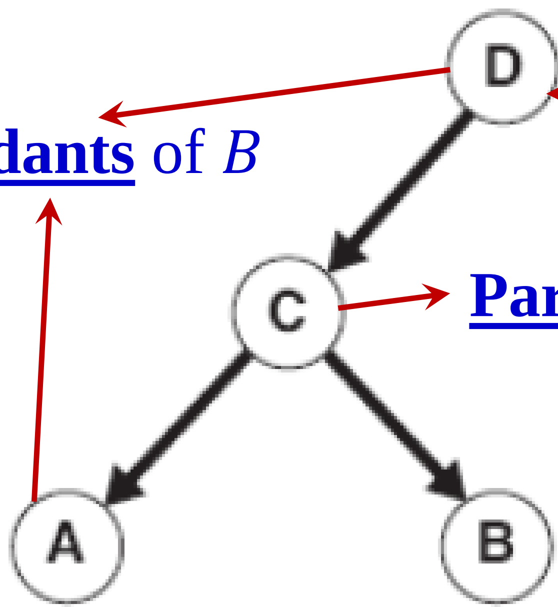


Directed arc ➔ dependence relationship (C depends on A and B)

Another DAG Example

If there is a directed path from *D* to *B*, then *D* is an **ancestor** of *B*, and *B* is a **descendant** of *D*

Non-descendants of *B*



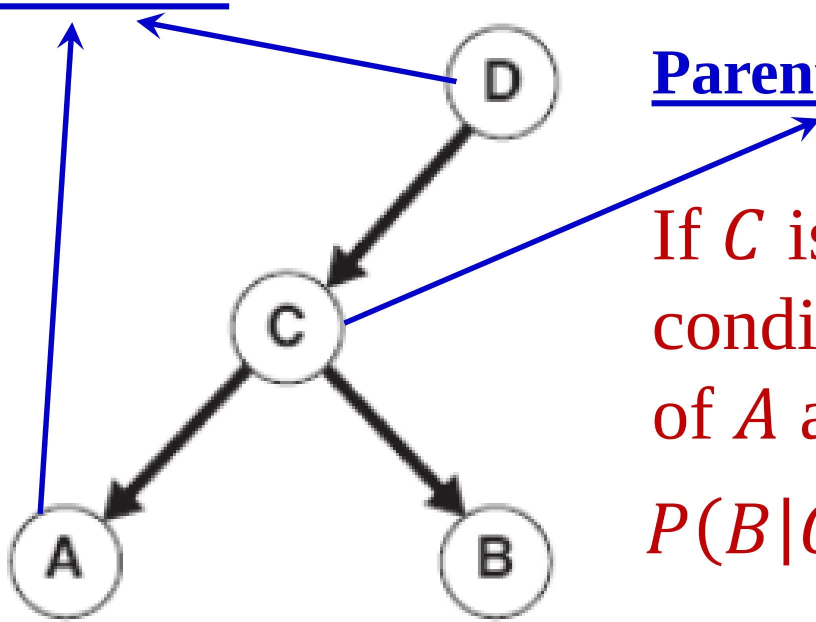
Parent of *B*

DAG: Conditional Independence

- Property (**conditional independence**): a node in a Bayesian network is conditionally independent of its non-descendants, if its parents are known.

Non-descendants of B

Parent of B



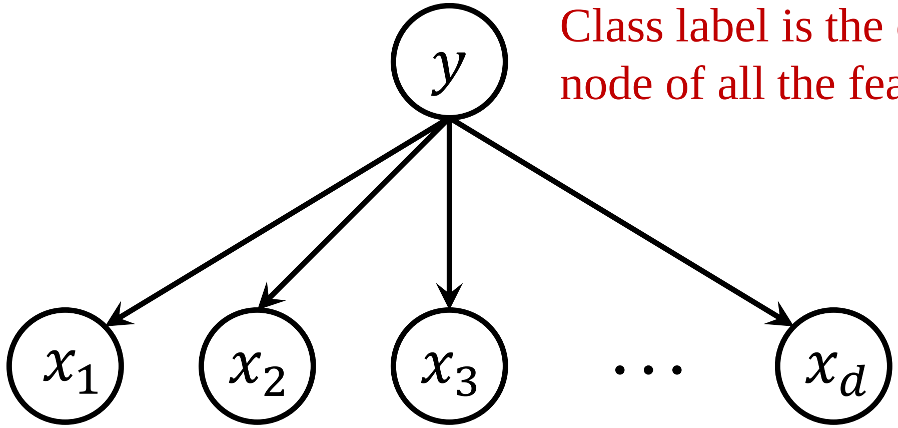
If C is observed, then B is conditionally independent of A and D

$$P(B|C,A,D) = P(B|C)$$

$$P(B,A,D|C) = P(B|C)P(D|C)\,P(A|C)$$

A Special Case: Naïve Bayes

- A Naïve Bayes classifier can be represented using a special DAG:



Class label is the only parent node of all the features

Once a class label is given, all the features are conditionally independent to each other, because each feature is a non-descendant to any other features

$$P(x_1, x_2, .., x_d | y = c) = \prod_{i=1}^{d} P(x_i | y = c)$$

BBN Representation

- Besides the conditional independence conditions imposed by the network topology, each node is also associated with a probability table.
  - If a node X does not have any parents, then the table contains only the prior probability $P(X)$
  - If a node X has only one parent, Z, then the table contains the conditional probability $P(X|Z)$
  - If a node X has multiple parents, $\{Z_1, Z_2, \ldots, Z_k\}$, then the table contains the conditional probability $P(X|Z_1, Z_2, \ldots, Z_k)$

BBN Representation: Example

A Bayesian network for modeling patients with heart disease or heartburn problems

$P(\text{E=No}) = 1 - P(\text{E=Yes}) = 0.3$

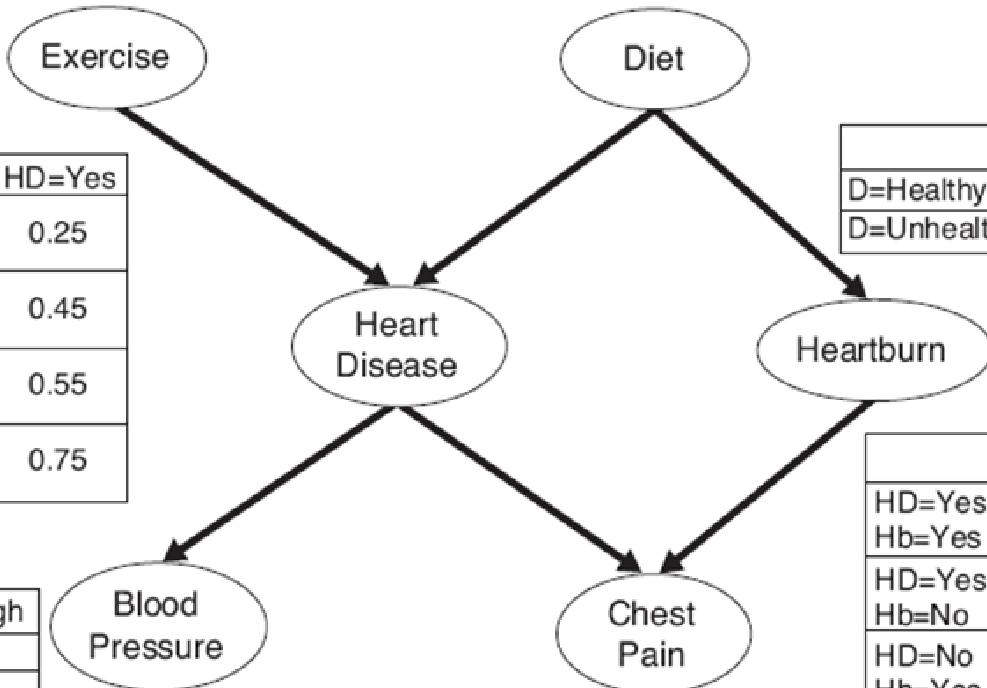
$P(\text{E=Yes}) = 0.7$

E=Yes
0.7

D=Healthy
0.25

$P(\text{HD=No}|\text{E=Yes, D=Healthy})$
$= 1 - P(\text{HD=Yes}|\text{E=Yes, D=Healthy})$
$= 0.75$

$P(\text{HD=Yes}|\text{E=Yes, D=Healthy})$
$= 0.25$



	HD=Yes
E=Yes D=Healthy	0.25
E=Yes D=Unhealthy	0.45
E=No D=Healthy	0.55
E=No D=Unhealthy	0.75

	Hb=Yes
D=Healthy	0.2
D=Unhealthy	0.85

	BP=High
HD=Yes	0.85
HD=No	0.2

	CP=Yes
HD=Yes Hb=Yes	0.8
HD=Yes Hb=No	0.6
HD=No Hb=Yes	0.4
HD=No Hb=No	0.1

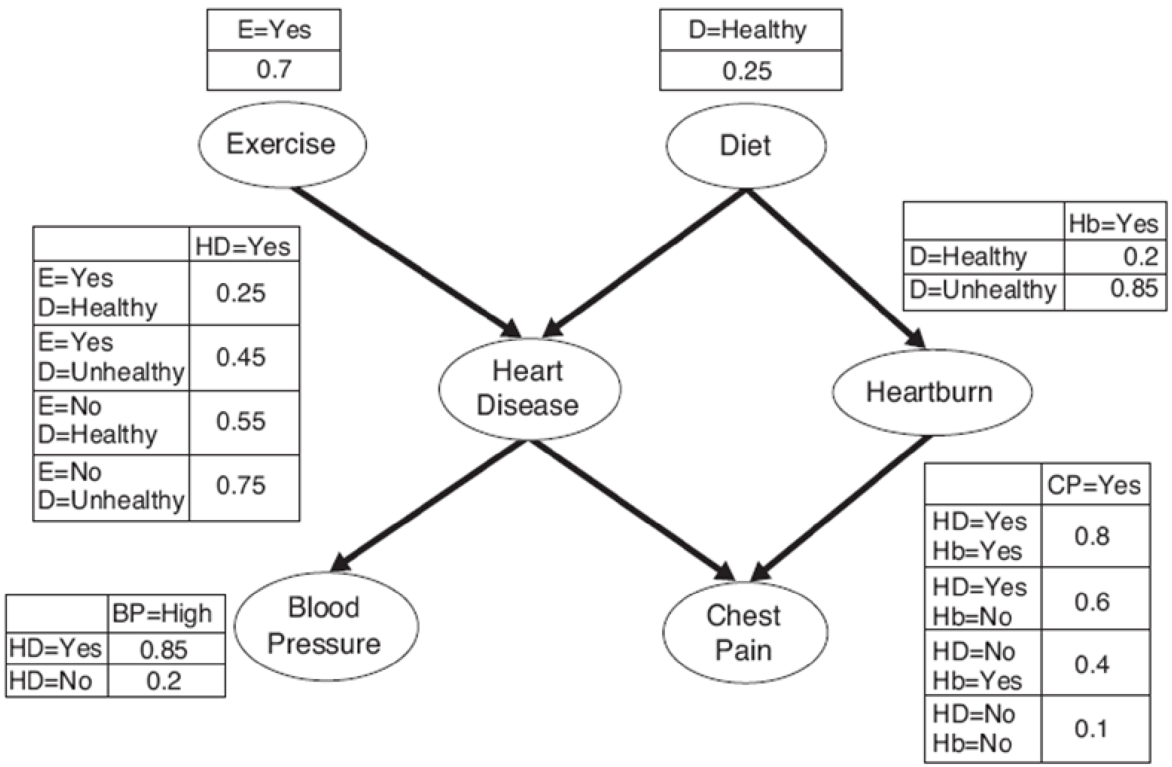
BBN Model Building

- Two steps in the training phase:
  - Creating the structure of the network, i.e., DAG
    - Network topology can be obtained by encoding the subjective knowledge of domain experts
    - Or can be learned from data (structure learning) --- still an open problem
  - Estimating the probability values in the table associated with each node
    - Counting based on the definition of the corresponding probabilities
- Note: in this module we only focus on how to use a BBN to make predictions (or inference)

Suppose Heart Disease is our target variable to make prediction on (i.e., output), the other variables are input features, whose values can be observed or missing.

Inference: Example 1

- Without any additional information, to determine whether the person is likely to have heart disease.



$P(\text{HD=No})$ *vs.* $P(\text{HD=Yes})$

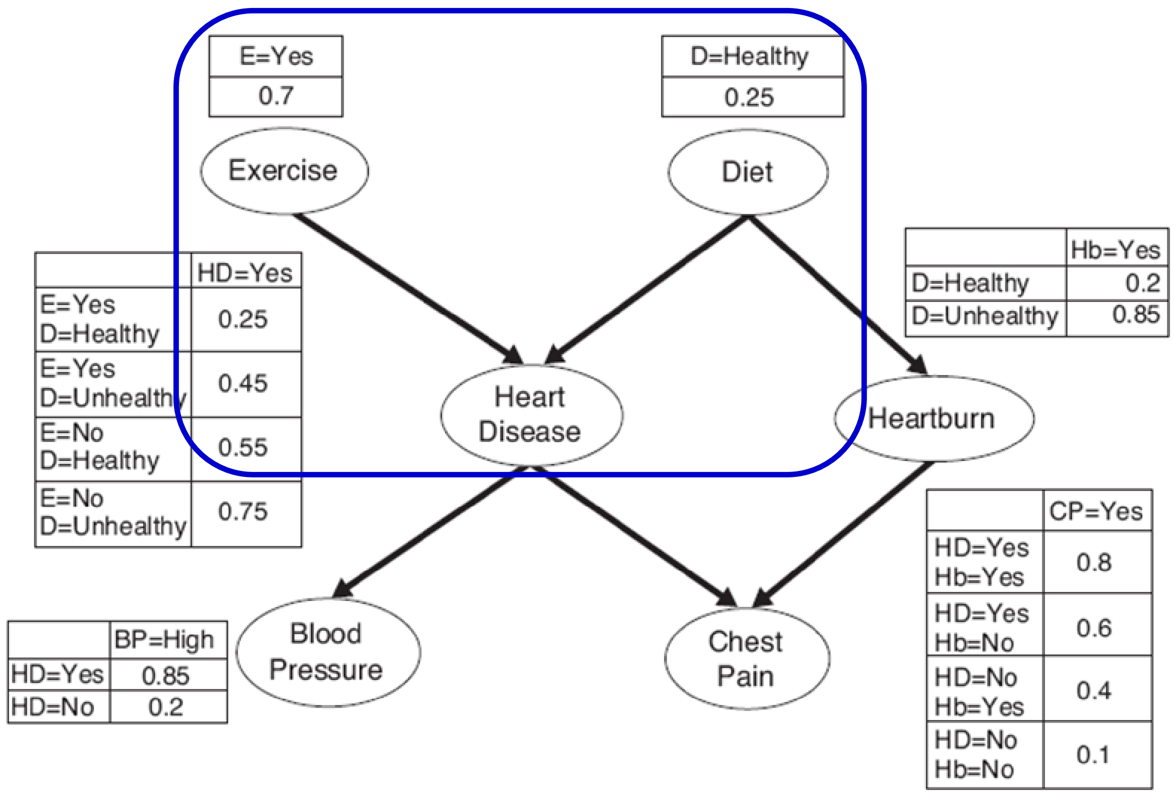
Inference: Example 1 (cont.)

$P(\text{HD=Yes})$ **?**

$$P(\text{HD=Yes}) = \sum_{\alpha}\sum_{\beta} P(\text{HD=Yes}, \text{E}=\alpha, \text{D}=\beta)$$

Sum Rule

$$\alpha = \{\text{Yes}, \text{No}\} \quad \beta = \{\text{Healthy}, \text{Unhealthy}\}$$



E=Yes
0.7

D=Healthy
0.25

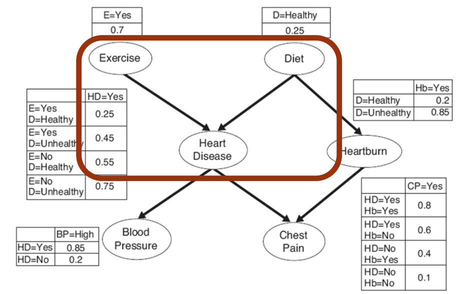
	HD=Yes
E=Yes D=Healthy	0.25
E=Yes D=Unhealthy	0.45
E=No D=Healthy	0.55
E=No D=Unhealthy	0.75

	Hb=Yes
D=Healthy	0.2
D=Unhealthy	0.85

	BP=High
HD=Yes	0.85
HD=No	0.2

	CP=Yes
HD=Yes Hb=Yes	0.8
HD=Yes Hb=No	0.6
HD=No Hb=Yes	0.4
HD=No Hb=No	0.1

Inference: Example 1 (cont.)



$$P(\text{HD=Yes}) = \sum_{\alpha}\sum_{\beta} \boxed{P(\text{HD=Yes}, \text{E}=\alpha, \text{D}=\beta)}$$

Product Rule $$= \sum_{\alpha}\sum_{\beta} \boxed{P(\text{HD=Yes}|\text{E}=\alpha, \text{D}=\beta)\boxed{P(\text{E}=\alpha, \text{D}=\beta)}}$$

$$= \sum_{\alpha}\sum_{\beta} P(\text{HD=Yes}|\text{E}=\alpha, \text{D}=\beta)\boxed{P(\text{E}=\alpha)P(\text{D}=\beta)}$$

Independence

$$\alpha = \{\text{Yes}, \text{No}\} \qquad \beta = \{\text{Healthy}, \text{Unhealthy}\}$$

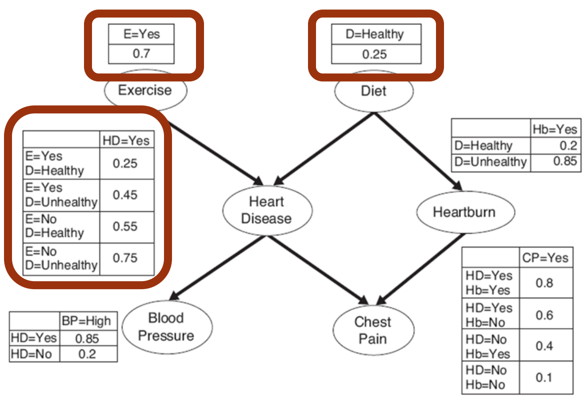
$$P(\text{HD=Yes}) = \sum_{\alpha}\sum_{\beta} P(\text{HD=Yes}|\text{E}=\alpha, \text{D}=\beta)P(\text{E}=\alpha)P(\text{D}=\beta)$$

$$= P(\text{HD=Yes}|\text{E=Yes, D=Healthy})P(\text{E=Yes})P(\text{D=Healthy})$$

$$+ P(\text{HD=Yes}|\text{E=Yes, D=Unhealthy})P(\text{E=Yes})P(\text{D=Unhealthy})$$

$$+ P(\text{HD=Yes}|\text{E=No, D=Healthy})P(\text{E=No})P(\text{D=Healthy})$$

$$+ P(\text{HD=Yes}|\text{E=No, D=Unhealthy})P(\text{E=No})P(\text{D=Unhealthy})$$



Look up probability tables

$$= 0.25 \times 0.7 \times 0.25 + 0.45 \times 0.7 \times 0.75 + 0.55 \times 0.3 \times 0.25 + 0.75 \times 0.3 \times 0.75$$
$$= 0.49$$

Inference: Example 1 (cont.)

$P(\mathrm{HD{=}Yes}) = 0.49$

$P(\mathrm{HD{=}No}) = 1 - P(\mathrm{HD{=}Yes}) = 0.51$

- Therefore, the person has a slightly higher chance of not getting the heart disease.

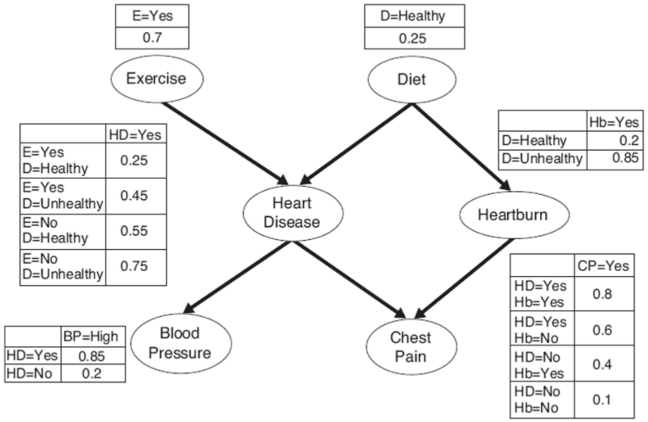
Inference: Example 2

- If the person has high blood pressure, chest pain and heartburn, but does regular exercise and eats a healthy diet, to diagnose whether the patient has heart disease:

$$P(\text{HD=Yes}|\text{BP=High, Hb=Yes, CP=Yes, D=Healthy, E=Yes})$$

vs.

$$P(\text{HD=No}|\text{BP=High, Hb=Yes, CP=Yes, D=Healthy, E=Yes})$$



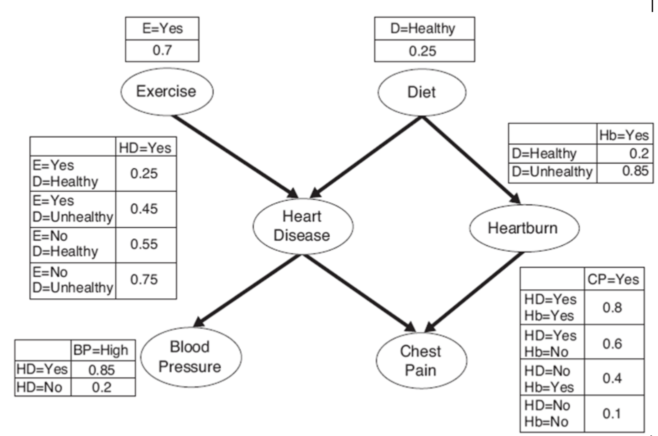
$$P(\text{HD=Yes}|\text{BP=High, Hb=Yes, CP=Yes, D=Healthy, E=Yes})$$

$$= \frac{\boxed{P(\text{HD=Yes, BP=High, Hb=Yes, CP=Yes, D=Healthy, E=Yes})}}{P(\text{BP=High, Hb=Yes, CP=Yes, D=Healthy, E=Yes})}$$

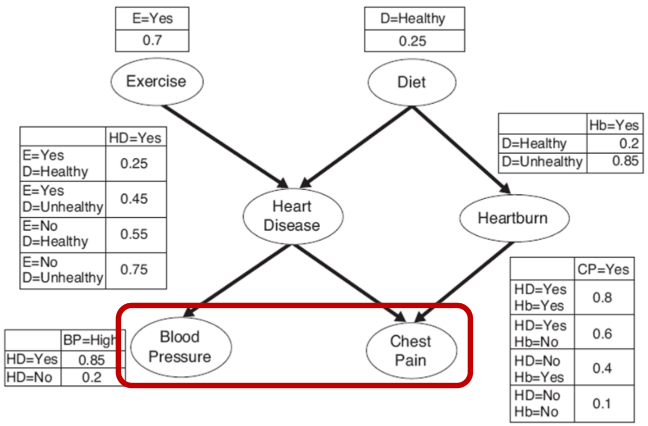
How to estimate the joint probability in the numerator?

Keep in mind that the goal is to rewrite the joint probability into an equivalent form, such that for all the probabilities in the rewritten equivalent form, their values can be found from the tables in the BBN.

There are many ways to rewrite the above joint probability, e.g.,
$P(\text{D, E} \mid \text{HD, BP, Hb, CP})P(\text{HD, BP, Hb, CP})$, or
$P(\text{D, BP, CP} \mid \text{E, Hb, HD})P(\text{E, Hb, HD})$, many others



Which one is useful for the joint probability simplification?



From the BBN, we found that the variables BP and CP are child nodes of HD and Hb, but not a parent node for any other variables. That means we could not find any conditional probabilities in the tables, which involves BP and CP in the condition, like $P(\mathrm{D, E} \mid \mathrm{HD, BP, Hb, CP})$. In other words, if in the rewritten equivalent form, there is a probability where BP or (and) CP is (are) in the condition, we still have to further rewrite it such that BP and CP are not in the condition of any conditional probability term.

The above analysis motivates us to transform the joint probability (numerator) using the following form based on the product rule:

$$P(\mathrm{HD{=}Yes, BP{=}High, Hb{=}Yes, CP{=}Yes, D{=}Healthy, E{=}Yes})$$

$$= P(\mathrm{BP{=}High, CP{=}Yes} \mid \mathrm{HD{=}Yes, Hb{=}Yes, E{=}Yes, D{=}Healthy})P(\mathrm{HD{=}Yes, Hb{=}Yes, E{=}Yes, D{=}Healthy})$$

Not in the condition of the conditional probability

Denote by $\mathbf{U} = \{\mathrm{BP, CP}\}$ and $\mathbf{V} = \{\mathrm{HD, Hb, E, D}\}$

$$P(\mathbf{U}, \mathbf{V}) = P(\mathbf{U} \mid \mathbf{V})P(\mathbf{V})$$

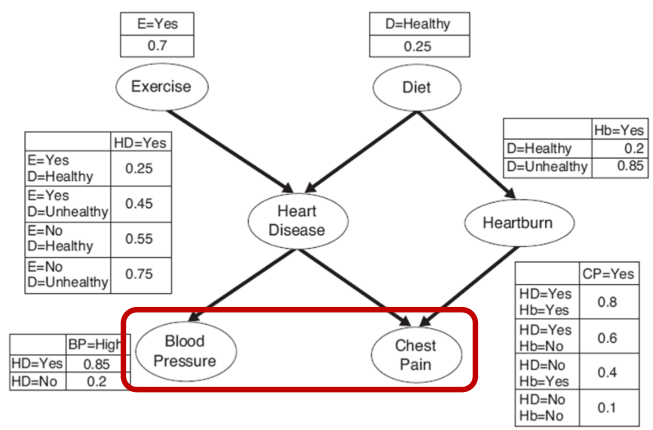
$$P(\text{HD=Yes}|\text{BP=High, Hb=Yes, CP=Yes, D=Healthy, E=Yes})$$

$$= \frac{\boxed{P(\text{HD=Yes, BP=High, Hb=Yes, CP=Yes, D=Healthy, E=Yes})}}{P(\text{BP=High, Hb=Yes, CP=Yes, D=Healthy, E=Yes})}$$

Denote by $\mathbf{U} = \{\text{BP, CP}\}$ and $\mathbf{V} = \{\text{HD, Hb, E, D}\}$

$$P(\mathbf{U}, \mathbf{V}) = P(\mathbf{U}|\mathbf{V})P(\mathbf{V})$$

$$\boxed{\underbrace{\overbrace{P(\text{BP=High, CP=Yes | HD=Yes, Hb=Yes, E=Yes, D=Healthy})}^{1}}_{\text{Slide 29}}\underbrace{\overbrace{\text{P(HD=Yes, Hb=Yes, E=Yes, D=Healthy)}}^{2}}_{\text{Slide 30}}}$$



Recall: if **A** and **B** are conditionally independent given **C**, we have

$$P(\mathbf{A}|\mathbf{B},\mathbf{C}) = P(\mathbf{A}|\mathbf{C}) \text{ or } P(\mathbf{A},\mathbf{B}|\mathbf{C}) = P(\mathbf{A}|\mathbf{C})P(\mathbf{B}|\mathbf{C})$$

1

$$P(\text{BP=High, CP=Yes} \mid \boxed{\text{HD=Yes, Hb=Yes,}}\ \text{E=Yes, D=Healthy})$$

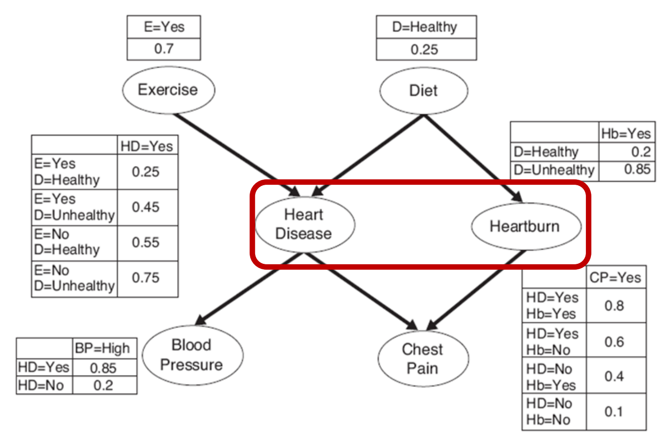
- HD and Hb are parents of BP and CP
- E and D are non-descendants of BP and CP
- Therefore, BP and CP are conditionally independent of E and D given HD and Hb

$$P(\text{BP=High, CP=Yes} \mid \text{HD=Yes, Hb=Yes})$$

- HD and Hb are parents of CP, and BP is a non-descendant of CP
- Therefore, BP and CP are conditionally independent given HD and Hb

$$\boxed{P(\text{BP=High} \mid \text{HD=Yes, Hb=Yes})}\,P(\text{CP=Yes} \mid \text{HD=Yes, Hb=Yes})$$

- HD is parent of BP
- Hb is a non-descendant of BP
- Therefore, BP and Hb are conditionally independent given HD



$$P(\text{BP=High} \mid \text{HD=Yes})\,P(\text{CP=Yes} \mid \text{HD=Yes, Hb=Yes})$$

$$= 0.85 \times 0.8 = 0.68$$

2

P(HD=Yes, Hb=Yes, E=Yes, D=Healthy)

Denote by $\mathbf{U} = \{\text{HD}, \text{Hb}\}$ and $\mathbf{V} = \{\text{E}, \text{D}\}$, $P(\mathbf{U}, \mathbf{V}) = P(\mathbf{U}|\mathbf{V})P(\mathbf{V})$

$$= P(\text{HD=Yes, Hb=Yes} \mid \text{E=Yes, D=Healthy})\,\text{P(E=Yes, D=Healthy)}$$

Given E and D, HD and Hb are conditionally independent

Independence

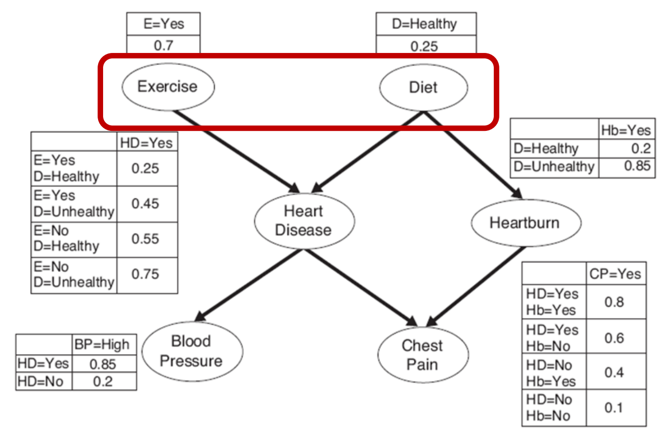
$$P(\text{HD=Yes} \mid \text{E=Yes, D=Healthy})\,P(\text{Hb=Yes} \mid \text{E=Yes, D=Healthy})$$

Given D, Hb and E are conditionally independent

$$P(\text{HD=Yes} \mid \text{E=Yes, D=Healthy})\,P(\text{Hb=Yes} \mid \text{D=Healthy})$$

$$= P(\text{HD=Yes} \mid \text{E=Yes, D=Healthy})\,P(\text{Hb=Yes} \mid \text{D=Healthy})\,\text{P(E=Yes)P(D=Healthy)}$$

$$= 0.25 \times 0.2 \times 0.7 \times 0.25 = 0.00875$$



$$P(\text{HD=Yes} \mid \text{BP=High, Hb=Yes, CP=Yes, D=Healthy, E=Yes})$$

$$= \frac{P(\text{BP=High, CP=Yes} \mid \text{HD=Yes, Hb=Yes, E=Yes, D=Healthy})\text{P(HD=Yes, Hb=Yes, E=Yes, D=Healthy)}}{P(\text{BP=High, Hb=Yes, CP=Yes, D=Healthy, E=Yes})}$$

$$= \frac{0.68 \times 0.00875}{P(\text{BP=High, Hb=Yes, CP=Yes, D=Healthy, E=Yes})} = \frac{0.00595}{P(\text{BP=High, Hb=Yes, CP=Yes, D=Healthy, E=Yes})}$$

$$P(\text{HD=No} \mid \text{BP=High, Hb=Yes, CP=Yes, D=Healthy, E=Yes})$$

$$= \frac{P(\text{BP=High, CP=Yes} \mid \textbf{HD=No}\text{, Hb=Yes, E=Yes, D=Healthy})\text{P(}\textbf{HD=No}\text{, Hb=Yes, E=Yes, D=Healthy)}}{P(\text{BP=High, Hb=Yes, CP=Yes, D=Healthy, E=Yes})}$$

$$= \frac{0.08 \times 0.02625}{P(\text{BP=High, Hb=Yes, CP=Yes, D=Healthy, E=Yes})} = \frac{0.0021}{P(\text{BP=High, Hb=Yes, CP=Yes, D=Healthy, E=Yes})}$$

$$\boxed{P(\text{HD=Yes}|\text{BP=High, Hb=Yes, CP=Yes, D=Healthy, E=Yes}) > P(\text{HD=No}|\text{BP=High, Hb=Yes, CP=Yes, D=Healthy, E=Yes})}$$

The person has a higher chance of getting the heart disease.

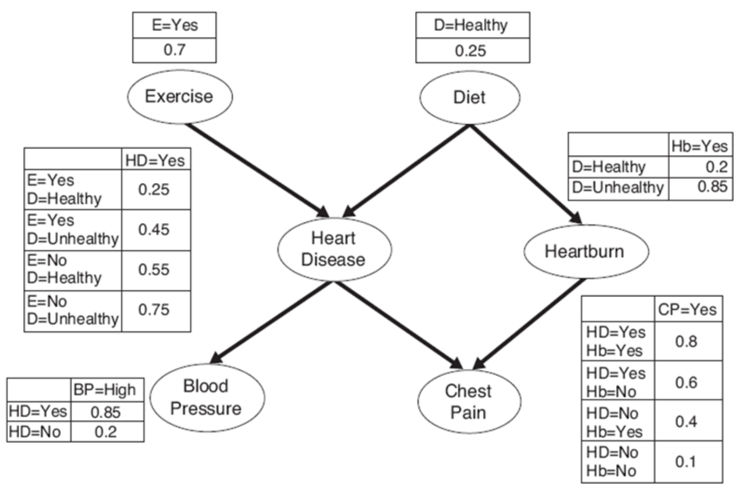
Inference: Example 3

- If the person has high blood pressure, but exercises regularly and eats a healthy diet, to diagnose about heart disease (estimate the probabilities).

$$P(\mathrm{HD{=}Yes} | \mathrm{BP{=}High}, \mathrm{D{=}Healthy}, \mathrm{E{=}Yes})$$

vs.

$$P(\mathrm{HD{=}No} | \mathrm{BP{=}High}, \mathrm{D{=}Healthy}, \mathrm{E{=}Yes})$$



Tutorial

Hint: Using BBNs for Inference

- Given a BBN, and an inference (prediction) task:
  1. Translate the problem into a probabilistic language, i.e., what probabilities to be estimated?
  2. If the probabilities to be estimated cannot be obtained from the probability tables of the BBN, then
     A. Identify a subgraph which captures the dependence between input variables (features) and the output variable (class)
     B. Based on the network topology, apply product rule, sum rule and the properties of conditional independence and independence to induce equivalent forms of the probabilities until all probabilities can be found from the probability tables

Bayesian Belief Networks: Summary

- BBNs use a (directed) graphical model to model dependence among variables
  - Other directed graphical models: Hidden Markov Models, Dynamic Bayesian Networks, etc.
  - Undirected graphical models: e.g., Markov Random Fields, Conditional Random Fields, etc.
- Network structure construction is difficult
  - Use domain knowledge – not complete, may not accurate
  - Learn structure from data – computationally expensive, greedy algorithm → not optimal

Thank you!