OntarioTech Science

Kourosh Davoudi
kourosh@uoit.ca

Classification: Naïve Bayes

CSCI 4150U: Data Mining

Learning Outcome

- What is the Nearest Neighbor Classifier?
  - Learn the ideas
  - Know the issues
- What is the Naïve Bayes classifier
  - Learn the main ideas
  - Explain are the issues and considerations
- What is Bayesian Belief Network?
- What are the Support Vector Machines?
  - Understand the main ideas
- What are ensemble approaches?
  - Learn the ideas and different approaches

Bayes Classifier

- A probabilistic framework for solving classification problems is based on the Bayesian Theorem
- Conditional Probability:

$$P(Y \mid X) = \frac{P(X,Y)}{P(X)}$$

$$P(X \mid Y) = \frac{P(X,Y)}{P(Y)}$$

- Bayes theorem:

$$P(Y \mid X) = \frac{P(X \mid Y)P(Y)}{P(X)}$$

Using Bayes Theorem for Classification

- Consider each attribute and class label as random variables
- Given a record with attributes $(X_1, X_2, \ldots, X_d)$
  - Goal is to predict class Y
  - Specifically, we want to find the value of Y that maximizes

$$P(Y | X_1, X_2, \ldots, X_d)$$

- Can we estimate $P(Y | X_1, X_2, \ldots, X_d)$ directly from data?

Tid	Refund	Marital Status	Taxable Income	Evade
1	Yes	Single	125K	No
2	No	Married	100K	No
3	No	Single	70K	No
4	Yes	Married	120K	No
5	No	Divorced	95K	Yes
6	No	Married	60K	No
7	Yes	Divorced	220K	No
8	No	Single	85K	Yes
9	No	Married	75K	No
10	No	Single	90K	Yes

Example Data

Tid	Refund	Marital Status	Taxable Income	Evade
1	Yes	Single	125K	No
2	No	Married	100K	No
3	No	Single	70K	No
4	Yes	Married	120K	No
5	No	Divorced	95K	Yes
6	No	Married	60K	No
7	Yes	Divorced	220K	No
8	No	Single	85K	Yes
9	No	Married	75K	No
10	No	Single	90K	Yes

Given a Test Record:

$$X = (\text{Refund} = \text{No}, \text{Divorced}, \text{Income} = 120\text{K})$$

Can we estimate these probability?

$$P(Evade = Yes \mid X)$$

$$P(Evade = No \mid X)$$

Using Bayes Theorem for Classification

- Approach:
  - Compute posterior probability $P(Y \mid X1, X2, \ldots, X_d)$ using the Bayes theorem

$$\boxed{P(Y \mid X_1 X_2 \ldots X_n) = \frac{P(X_1 X_2 \ldots X_d \mid Y) P(Y)}{P(X_1 X_2 \ldots X_d)}}$$

  - Maximum a-posteriori: Choose Y that maximizes

$$P(Y \mid X_1, X_2, \ldots, X_d)$$

  - Equivalent to choosing value of Y that maximizes

$$\mathrm{P}(X_1, X_2, \ldots, X_d \mid Y)\, \mathrm{P}(Y)$$

How to estimate $\mathrm{P}(X_1, X_2, \ldots, X_d \mid Y)\, \mathrm{P}(Y)$?

Example Data

Tid	Refund	Marital Status	Taxable Income	Evade
1	Yes	Single	125K	No
2	No	Married	100K	No
3	No	Single	70K	No
4	Yes	Married	120K	No
5	No	Divorced	95K	Yes
6	No	Married	60K	No
7	Yes	Divorced	220K	No
8	No	Single	85K	Yes
9	No	Married	75K	No
10	No	Single	90K	Yes

Given a Test Record:

$$X = (\text{Refund} = \text{No}, \text{Divorced}, \text{Income} = 120\text{K})$$

Using Bayes Theorem:

$$P(\text{Yes} \mid X) = \frac{P(X \mid \text{Yes})P(\text{Yes})}{P(X)}$$

$$P(\text{No} \mid X) = \frac{P(X \mid \text{No})P(\text{No})}{P(X)}$$

So, the problem is reduced to estimation of:

$$P(X \mid \text{Yes})$$

$$P(X \mid \text{No})$$

How to estimate $P(X_1, X_2, \ldots, X_d \mid Y_j)$?

Naïve Bayes Classifier

- Assume independence among attributes X_i when class is given:

$$P(X_1, X_2, \ldots, X_d \mid Y_j) = P(X_1 \mid Y_j)\, P(X_2 \mid Y_j) \ldots P(X_d \mid Y_j)$$

- Now we can estimate $P(X_i \mid Y_j)$ for all X_i and Y_j combinations from the training data

- New point is classified to Y_j if $P(Y_j) \prod P(X_i \mid Y_j)$ is maximal.

Naïve Bayes on Example Data

Given a Test Record: $X = (\text{Refund} = \text{No}, \text{Divorced}, \text{Income} = 120\text{K})$

Tid	Refund	Marital Status	Taxable Income	Evade
1	Yes	Single	125K	No
2	No	Married	100K	No
3	No	Single	70K	No
4	Yes	Married	120K	No
5	No	Divorced	95K	Yes
6	No	Married	60K	No
7	Yes	Divorced	220K	No
8	No	Single	85K	Yes
9	No	Married	75K	No
10	No	Single	90K	Yes

P(X | Yes) =

P(Refund = No | Yes) x

P(Divorced | Yes) x

P(Income = 120K | Yes)

P(X | No) =

P(Refund = No | No) x

P(Divorced | No) x

P(Income = 120K | No)

Estimate Probabilities from Data

- $P(y)$ = fraction of instances of class y
  - Examples:

    $P(No) = 7/10,$
    $P(Yes) = 3/10$

- For categorical attributes:

$$P(X_i = c \mid y) = n_c / n$$

where n_c is number of instances having attribute value $X_i = c$ and belonging to class y and n is the number of instances of class y

  - Examples:

    $P(Status=Married|No) = 4/7$
    $P(Refund=Yes|Yes)=0$

Tid	Refund	Marital Status	Taxable Income	Evade
1	Yes	Single	125K	No
2	No	Married	100K	No
3	No	Single	70K	No
4	Yes	Married	120K	No
5	No	Divorced	95K	Yes
6	No	Married	60K	No
7	Yes	Divorced	220K	No
8	No	Single	85K	Yes
9	No	Married	75K	No
10	No	Single	90K	Yes

What is the $P(Refund = Yes|No)$?

Tid	Refund	Marital Status	Taxable Income	Evade
1	Yes	Single	125K	No
2	No	Married	100K	No
3	No	Single	70K	No
4	Yes	Married	120K	No
5	No	Divorced	95K	Yes
6	No	Married	60K	No
7	Yes	Divorced	220K	No
8	No	Single	85K	Yes
9	No	Married	75K	No
10	No	Single	90K	Yes

A. 2/7
B. 3/7
C. 3/5
D. 2/5

Estimate Probabilities from Data

- For continuous attributes:
  - Discretization: Partition the range into bins:
    - Replace continuous value with bin value
      - Attribute changed from continuous to ordinal
  - Probability density estimation:
    - Assume attribute follows a normal distribution
    - Use data to estimate parameters of distribution (e.g., mean and standard deviation)
    - Once probability distribution is known, use it to estimate the conditional probability $P(X_i|Y)$

Estimate Probabilities from Data

Tid	Refund	Marital Status	Taxable Income	Evade
1	Yes	Single	125K	No
2	No	Married	100K	No
3	No	Single	70K	No
4	Yes	Married	120K	No
5	No	Divorced	95K	Yes
6	No	Married	60K	No
7	Yes	Divorced	220K	No
8	No	Single	85K	Yes
9	No	Married	75K	No
10	No	Single	90K	Yes

- Normal distribution:

$$P(X_i \mid Y_j) = \frac{1}{\sqrt{2\pi\sigma_{ij}^2}} e^{-\frac{(X_i-\mu_{ij})^2}{2\sigma_{ij}^2}}$$

One for each (X_i, Y_i) pair

- For (Income, Class=No):
  - If Class=No
    - sample mean = 110
    - sample variance = 2975

$$P(Income = 120 \mid No) = \frac{1}{\sqrt{2\pi}(54.54)} e^{-\frac{(120-110)^2}{2(2975)}} = 0.0072$$

Example of Naïve Bayes Classifier

Given a Test Record: $X = (\text{Refund} = \text{No}, \text{Divorced}, \text{Income} = 120\text{K})$

Naïve Bayes Classifier:

P(No) = 7/10,
P(Yes) = 3/10

P(Refund = Yes | No) = 3/7
P(Refund = No | No) = 4/7
P(Refund = Yes | Yes) = 0
P(Refund = No | Yes) = 1
P(Marital Status = Single | No) = 2/7
P(Marital Status = Divorced | No) = 1/7
P(Marital Status = Married | No) = 4/7
P(Marital Status = Single | Yes) = 2/3
P(Marital Status = Divorced | Yes) = 1/3
P(Marital Status = Married | Yes) = 0

For Taxable Income:
If class = No: sample mean = 110
sample variance = 2975
If class = Yes: sample mean = 90
sample variance = 25

- $$\begin{aligned} P(X \mid No) &= P(Refund{=}No \mid No) \\ &\quad \times P(Divorced \mid No) \\ &\quad \times P(Income{=}120K \mid No) \\ &= 4/7 \times 1/7 \times 0.0072 = 0.0006 \end{aligned}$$

- $$\begin{aligned} P(X \mid Yes) &= P(Refund{=}No \mid Yes) \\ &\quad \times P(Divorced \mid Yes) \\ &\quad \times P(Income{=}120K \mid Yes) \\ &= 1 \times 1/3 \times 1.2 \times 10^{-9} = 4 \times 10^{-10} \end{aligned}$$

Since $P(X|No)P(No) > P(X|Yes)P(Yes)$

Therefore $P(No|X) > P(Yes|X)$ => Class = No

Activity: Naïve Bayes Classifier

Name	Give Birth	Can Fly	Live in Water	Have Legs	Class
human	yes	no	no	yes	mammals
python	no	no	no	no	non-mammals
salmon	no	no	yes	no	non-mammals
whale	yes	no	yes	no	mammals
frog	no	no	sometimes	yes	non-mammals
komodo	no	no	no	yes	non-mammals
bat	yes	yes	no	yes	mammals
pigeon	no	yes	no	yes	non-mammals
cat	yes	no	no	yes	mammals
leopard shark	yes	no	yes	no	non-mammals
turtle	no	no	sometimes	yes	non-mammals
penguin	no	no	sometimes	yes	non-mammals
porcupine	yes	no	no	yes	mammals
eel	no	no	yes	no	non-mammals
salamander	no	no	sometimes	yes	non-mammals
gila monster	no	no	no	yes	non-mammals
platypus	no	no	no	yes	mammals
owl	no	yes	no	yes	non-mammals
dolphin	yes	no	yes	no	mammals
eagle	no	yes	no	yes	non-mammals

Give Birth	Can Fly	Live in Water	Have Legs	Class
yes	no	yes	no	?

Find the class using Naïve Bayes?

(On Piazza)

Issues with Naïve Bayes Classifier

Consider the table with Tid = 7 deleted

Tid	Refund	Marital Status	Taxable Income	Evade
1	Yes	Single	125K	No
2	No	Married	100K	No
3	No	Single	70K	No
4	Yes	Married	120K	No
5	No	Divorced	95K	Yes
6	No	Married	60K	No
8	No	Single	85K	Yes
9	No	Married	75K	No
10	No	Single	90K	Yes

Given $X = (Refund = Yes, Divorced, 120K)$

$$P(X \mid No) = 2/6 \times 0 \times 0.0083 = 0$$

$$P(X \mid Yes) = 0 \times 1/3 \times 1.2 \times 10^{-9} = 0$$

Naïve Bayes Classifier:

$P(Refund = Yes \mid No) = 2/6$
$P(Refund = No \mid No) = 4/6$
$P(Refund = Yes \mid Yes) = 0$
$P(Refund = No \mid Yes) = 1$
$P(Marital\ Status = Single \mid No) = 2/6$
$P(Marital\ Status = Divorced \mid No) = 0$
$P(Marital\ Status = Married \mid No) = 4/6$
$P(Marital\ Status = Single \mid Yes) = 2/3$
$P(Marital\ Status = Divorced \mid Yes) = 1/3$
$P(Marital\ Status = Married \mid Yes) = 0/3$

For Taxable Income:

If class = No: sample mean = 91
sample variance = 685

If class = No: sample mean = 90
sample variance = 25

Naïve Bayes will not be able to classify X as Yes or No!

Issues with Naïve Bayes Classifier

- If one of the conditional probabilities is zero, then the entire expression becomes zero
  - Need to use other estimates of conditional probabilities than simple fractions
  - Probability estimation:

$$Original\colon\ P(X_i = c|y) = \frac{n_c}{n}$$

$$Laplace\ Estimate\colon\ P(X_i = c|y) = \frac{n_c + 1}{n + v}$$

$$m - estimate\colon\ P(X_i = c|y) = \frac{n_c + mp}{n + m}$$

n: number of training instances belonging to class y

n_c: number of instances with $X_i = c$ and $Y = y$

v: total number of attribute values that X_i can take

p: initial estimate of $P(X_i = c|y)$

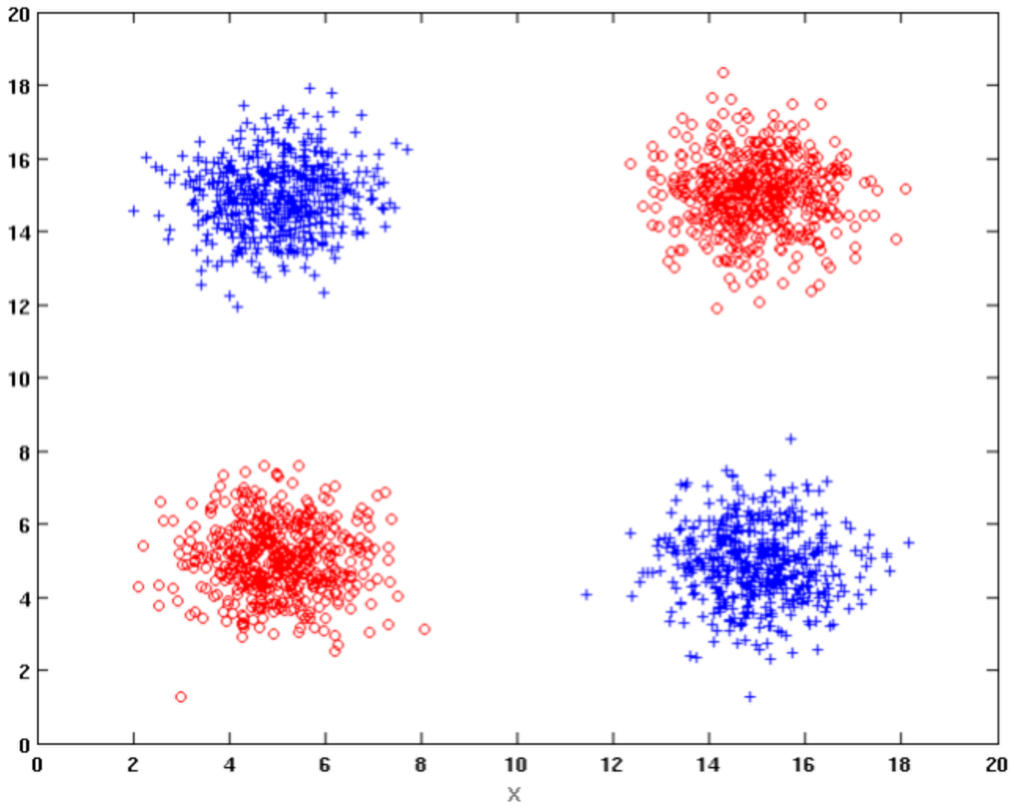
m: hyper-parameter for our confidence in p

Naïve Bayes (Summary)

- Robust to isolated noise points
- Handle missing values by ignoring the instance during probability estimate calculations
- Robust to irrelevant attributes
- Redundant and correlated attributes will violate class conditional assumption
    - Use other techniques such as Bayesian Belief Networks (BBN)

Naïve Bayes

- How does Naïve Bayes perform on the following dataset?



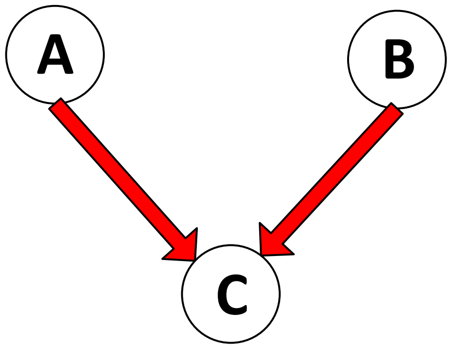
Conditional independence of attributes is violated

Learning Outcome

- What is the Nearest Neighbor Classifier?
  - Learn the ideas
  - Know the issues
- What is the Naïve Bayes classifier
  - Learn the main ideas
  - Explain how it works
  - What are the issues and considerations
- What is Bayesian Belief Network?
  - Understand the main ideas

Bayesian Belief Networks

- Provides graphical representation of probabilistic relationships among a set of random variables
- Consists of:
  - A directed acyclic graph (dag)
    - Node corresponds to a variable
    - Edges corresponds to dependence relationship between a pair of variables



  - A probability table associating each node to its immediate parent

Conditional Independence

- Assuming no conditional independence, what is the joint probability distribution $P(A, B, C, D)$?

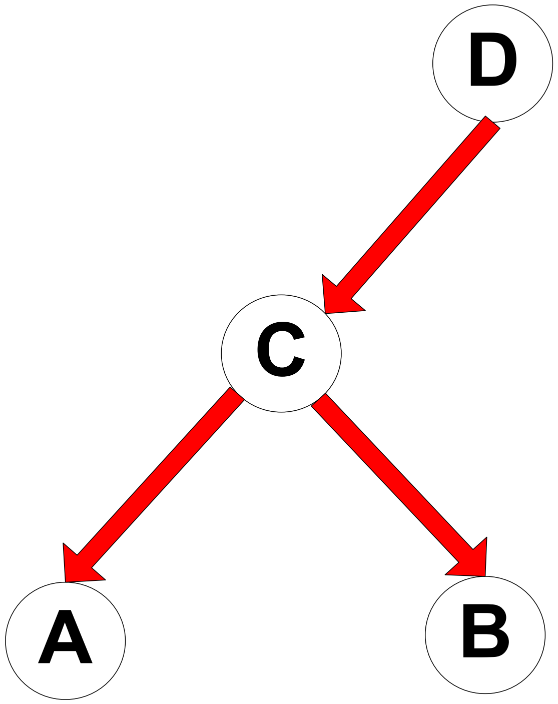
Conditional Independence

- Assuming no conditional independence, what is the joint probability distribution $P(A, B, C, D)$?

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

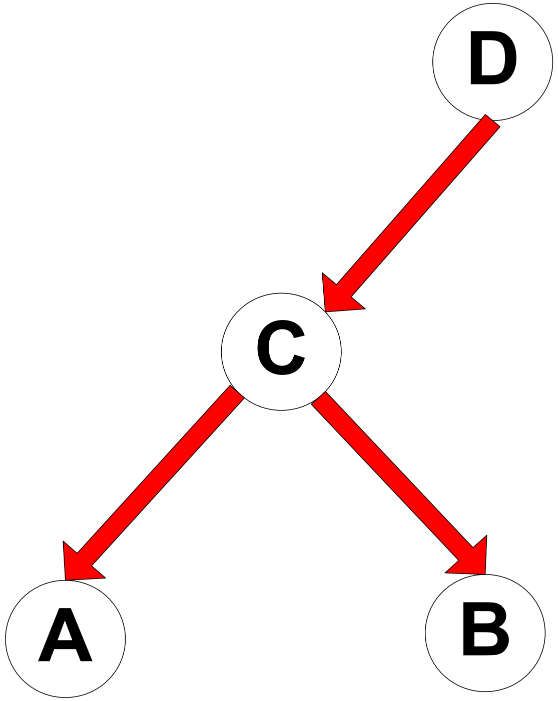
Conditional Independence

- Given the following Bayesian network, what is the joint probability distribution $P(A, B, C, D)$?



Conditional Independence

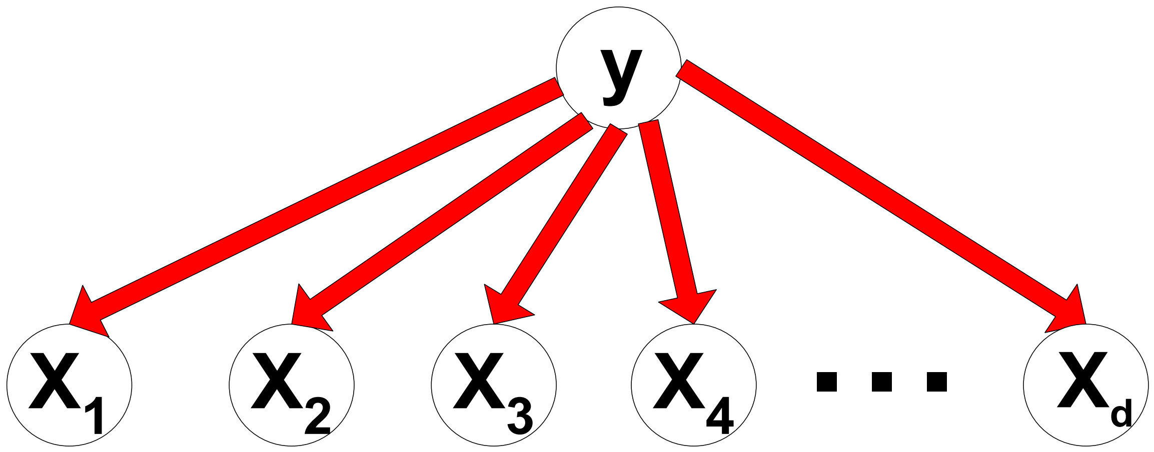
- Given the following Bayesian network, what is the joint probability distribution $P(A, B, C, D)$?



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

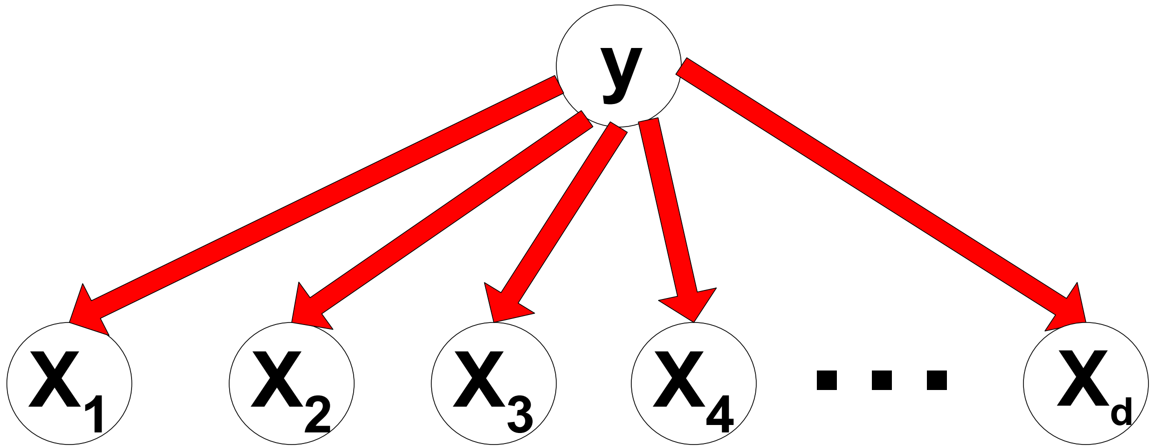
Conditional Independence

- Naïve Bayes assumption:



Conditional Independence

- Naïve Bayes assumption:



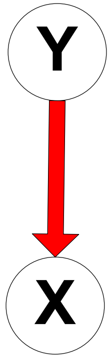
$$P(y|X) = \frac{P(y,X)}{P(X)} = \frac{P(y)*P(X_1|y)\ldots P(X_n|y)}{P(X)}$$

$$X = (X_1, X_2, \ldots, X_n)$$

Probability Tables

- If X does not have any parents, table contains prior probability $P(X)$

- If X has only one parent (Y), table contains conditional probability $P(X|Y)$

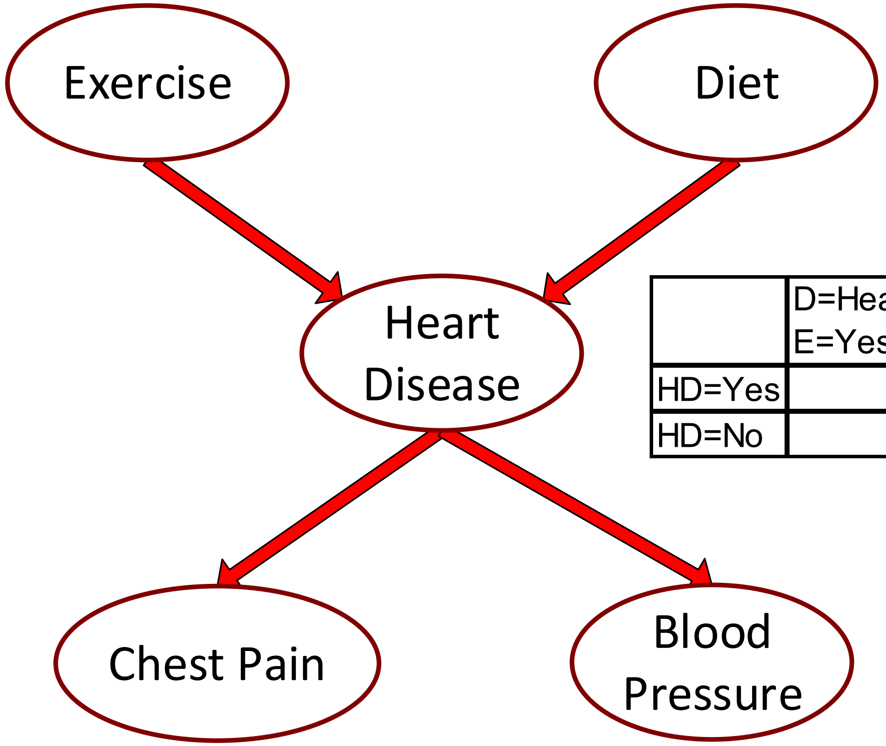


- If X has multiple parents ($Y_1, Y_2, \ldots, Y_k$), table contains conditional probability $P(X|Y_1, Y_2, \ldots, Y_k)$

Example of Bayesian Belief Network

Exercise=Yes	0.7
Exercise=No	0.3

Diet=Healthy	0.25
Diet=Unhealthy	0.75



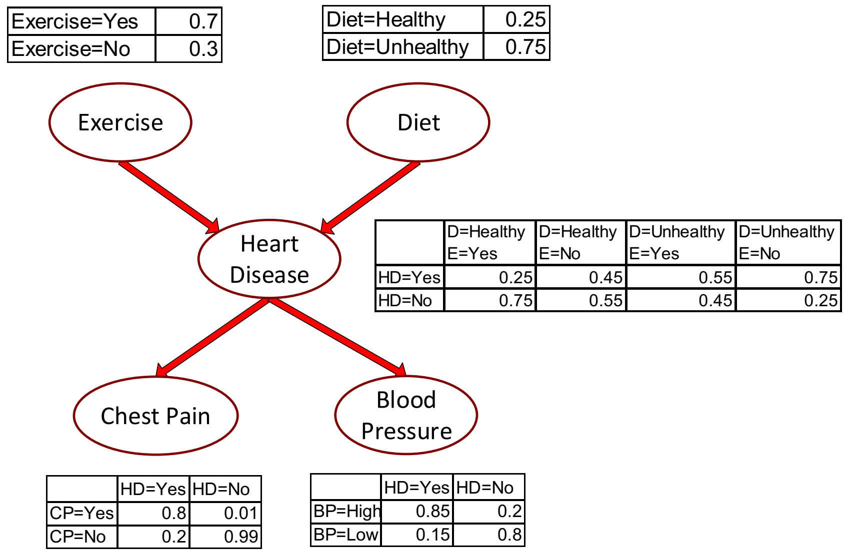
	D=Healthy E=Yes	D=Healthy E=No	D=Unhealthy E=Yes	D=Unhealthy E=No
HD=Yes	0.25	0.45	0.55	0.75
HD=No	0.75	0.55	0.45	0.25

	HD=Yes	HD=No
CP=Yes	0.8	0.01
CP=No	0.2	0.99

	HD=Yes	HD=No
BP=High	0.85	0.2
BP=Low	0.15	0.8

Example of Inferencing using BBN

- Given: $X = (E{=}No,\ D{=}Yes,\ CP{=}Yes,\ BP{=}High)$
  - Compute $P(HD|E,D,CP,BP)$?



Exercise=Yes	0.7
Exercise=No	0.3

Diet=Healthy	0.25
Diet=Unhealthy	0.75

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

	HD=Yes	HD=No
CP=Yes	0.8	0.01
CP=No	0.2	0.99

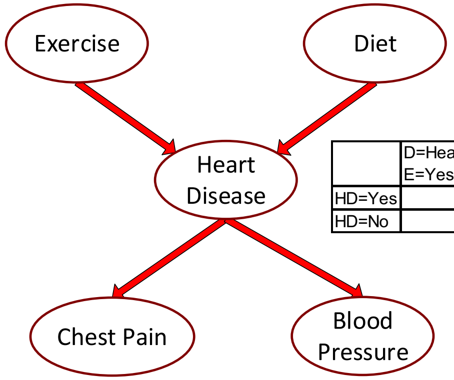
	HD=Yes	HD=No
BP=High	0.85	0.2
BP=Low	0.15	0.8

Example of Inferencing using BBN

- Given: $X = (E{=}No,\ D{=}Yes,\ CP{=}Yes,\ BP{=}High)$
  - Compute $P(HD|E,D,CP,BP)$?

Exercise=Yes	0.7
Exercise=No	0.3

Diet=Healthy	0.25
Diet=Unhealthy	0.75



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

	HD=Yes	HD=No
CP=Yes	0.8	0.01
CP=No	0.2	0.99

	HD=Yes	HD=No
BP=High	0.85	0.2
BP=Low	0.15	0.8

$$P(HD|E,D,CP,BP) = \frac{P(HD,\ E,D,CP,BP)}{P(E,D,CP,BP)} = \frac{P(HD|\ E,\ D) * P(CP|HD) * P(BP|HD) * P(E) * P(D)}{P(E,D,CP,BP)}$$

$$\propto P(HD|\ E,\ D) * P(CP|HD),\ P(BP|HD) * P(E) * P(D)$$

Example of Inferencing using BBN

- Given: $X = (E{=}No,\ D{=}Yes,\ CP{=}Yes,\ BP{=}High)$
  - Compute $P(HD|E,D,CP,BP)$?

- $P(HD{=}Yes|\ E{=}No,D{=}Yes) = 0.55$
  $P(CP{=}Yes|\ HD{=}Yes) = 0.8$
  $P(BP{=}High|\ HD{=}Yes) = 0.85$
  - $P(HD{=}Yes|E{=}No,D{=}Yes,CP{=}Yes,BP{=}High) \propto 0.55 \times 0.8 \times 0.85 = 0.374$

- $P(HD{=}No|\ E{=}No,D{=}Yes) = 0.45$
  $P(CP{=}Yes|\ HD{=}No) = 0.01$
  $P(BP{=}High|\ HD{=}No) = 0.2$
  - $P(HD{=}No|E{=}No,D{=}Yes,CP{=}Yes,BP{=}High) \propto 0.45 \times 0.01 \times 0.2 = 0.0009$

Classify X as Yes