Bayesian Classifiers

Introduction to Data Mining, 2nd Edition
by
Tan, Steinbach, Karpatne, Kumar

Bayes Classifier

- A probabilistic framework for solving classification problems
- Conditional Probability:
  \[ P(Y \mid X) = \frac{P(X,Y)}{P(X)} \]
  \[ P(X \mid Y) = \frac{P(X,Y)}{P(Y)} \]
- Bayes theorem:
  \[ P(Y \mid X) = \frac{P(X \mid Y)P(Y)}{P(X)} \]
Example of Bayes Theorem

- Given:
  - A doctor knows that meningitis causes stiff neck 50% of the time
  - Prior probability of any patient having meningitis is $1/50,000$
  - Prior probability of any patient having stiff neck is $1/20$

- If a patient has stiff neck, what’s the probability he/she has meningitis?

$$P(M | S) = \frac{P(S | M)P(M)}{P(S)} = \frac{0.5 \times 1/50000}{1/20} = 0.0002$$

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?
Example Data

Given a Test Record:

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

- Can we estimate \(P(\text{Evade = Yes} \mid X)\) and \(P(\text{Evade = No} \mid X)\)?

In the following we will replace

- \(\text{Evade} = \text{Yes}\) by Yes,
- \(\text{Evade} = \text{No}\) by No

<table>
<thead>
<tr>
<th>T/d</th>
<th>Refund</th>
<th>Marital Status</th>
<th>Taxable Income</th>
<th>Evade</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Yes</td>
<td>Single</td>
<td>125K</td>
<td>No</td>
</tr>
<tr>
<td>2</td>
<td>No</td>
<td>Married</td>
<td>100K</td>
<td>No</td>
</tr>
<tr>
<td>3</td>
<td>No</td>
<td>Single</td>
<td>70K</td>
<td>No</td>
</tr>
<tr>
<td>4</td>
<td>Yes</td>
<td>Married</td>
<td>120K</td>
<td>No</td>
</tr>
<tr>
<td>5</td>
<td>No</td>
<td>Divorced</td>
<td>90K</td>
<td>Yes</td>
</tr>
<tr>
<td>6</td>
<td>No</td>
<td>Married</td>
<td>90K</td>
<td>No</td>
</tr>
<tr>
<td>7</td>
<td>Yes</td>
<td>Divorced</td>
<td>220K</td>
<td>No</td>
</tr>
<tr>
<td>8</td>
<td>No</td>
<td>Single</td>
<td>80K</td>
<td>Yes</td>
</tr>
<tr>
<td>9</td>
<td>No</td>
<td>Married</td>
<td>75K</td>
<td>No</td>
</tr>
<tr>
<td>10</td>
<td>No</td>
<td>Single</td>
<td>90K</td>
<td>Yes</td>
</tr>
</tbody>
</table>

Using Bayes Theorem for Classification

- Approach:
  - compute posterior probability \(P(Y \mid X_1, X_2, \ldots, X_d)\) using the Bayes theorem

\[
P(Y \mid X_1, X_2, \ldots, X_n) = \frac{P(X_1, X_2, \ldots, X_d \mid Y) P(Y)}{P(X_1, X_2, \ldots, X_d)}
\]

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

- How to estimate \(P(X_1, X_2, \ldots, X_d \mid Y)\)?
### Example Data

Given a Test Record:

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

<table>
<thead>
<tr>
<th>Tid</th>
<th>Refund</th>
<th>Marital Status</th>
<th>Taxable Income</th>
<th>Evade</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Yes</td>
<td>Single</td>
<td>125K</td>
<td>No</td>
</tr>
<tr>
<td>2</td>
<td>No</td>
<td>Married</td>
<td>100K</td>
<td>No</td>
</tr>
<tr>
<td>3</td>
<td>No</td>
<td>Single</td>
<td>70K</td>
<td>No</td>
</tr>
<tr>
<td>4</td>
<td>Yes</td>
<td>Married</td>
<td>120K</td>
<td>No</td>
</tr>
<tr>
<td>5</td>
<td>No</td>
<td>Divorced</td>
<td>85K</td>
<td>Yes</td>
</tr>
<tr>
<td>6</td>
<td>No</td>
<td>Married</td>
<td>60K</td>
<td>No</td>
</tr>
<tr>
<td>7</td>
<td>Yes</td>
<td>Divorced</td>
<td>220K</td>
<td>No</td>
</tr>
<tr>
<td>8</td>
<td>No</td>
<td>Single</td>
<td>85K</td>
<td>Yes</td>
</tr>
<tr>
<td>9</td>
<td>No</td>
<td>Married</td>
<td>75K</td>
<td>No</td>
</tr>
<tr>
<td>10</td>
<td>No</td>
<td>Single</td>
<td>90K</td>
<td>Yes</td>
</tr>
</tbody>
</table>

#### Using Bayes Theorem:

\[
\begin{align*}
\Box \quad P(\text{Yes} \mid X) &= \frac{P(X \mid \text{Yes})P(\text{Yes})}{P(X)} \\
\Box \quad P(\text{No} \mid X) &= \frac{P(X \mid \text{No})P(\text{No})}{P(X)}
\end{align*}
\]

\[ \Box \quad \text{How to estimate } P(X \mid \text{Yes}) \text{ and } P(X \mid \text{No})? \]

### Naïve Bayes Classifier

- Assume independence among attributes \( X_i \) when class is given:
  \[
  P(X_1, X_2, ..., X_d \mid Y_j) = P(X_1 \mid Y_j)P(X_2 \mid Y_j)\ldots P(X_d \mid Y_j)
  \]
  - 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.
### Conditional Independence

- $X$ and $Y$ are conditionally independent given $Z$ if $P(X|YZ) = P(X|Z)$

- Example: Arm length and reading skills
  - Young child has shorter arm length and limited reading skills, compared to adults
  - If age is fixed, no apparent relationship between arm length and reading skills
  - Arm length and reading skills are conditionally independent given age

### Naïve Bayes on Example Data

**Given a Test Record:**

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

- $P(X | Yes) = P(\text{Refund} = \text{No} | Yes) \times P(\text{Divorced} | Yes) \times P(\text{Income} = 120K | Yes)$

- $P(X | No) = P(\text{Refund} = \text{No} | No) \times P(\text{Divorced} | No) \times P(\text{Income} = 120K | No)$
### Estimate Probabilities from Data

#### Class: $P(Y) = \frac{N_{c}}{N}$
- e.g., $P(\text{No}) = \frac{7}{10}$, $P(\text{Yes}) = \frac{3}{10}$

#### For categorical attributes:

$$P(X_i | Y_k) = \frac{|X_{ik}|}{N_{ck}}$$
- where $|X_{ik}|$ is number of instances having attribute value $X_i$ and belonging to class $Y_k$
- Examples:
  - $P(\text{Status} = \text{Married} | \text{No}) = \frac{4}{7}$
  - $P(\text{Refund} = \text{Yes} | \text{Yes}) = 0$

<table>
<thead>
<tr>
<th>Tid</th>
<th>Refund</th>
<th>Marital Status</th>
<th>Taxable Income</th>
<th>Evade</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Yes</td>
<td>Single</td>
<td>125K</td>
<td>No</td>
</tr>
<tr>
<td>2</td>
<td>No</td>
<td>Married</td>
<td>100K</td>
<td>No</td>
</tr>
<tr>
<td>3</td>
<td>No</td>
<td>Single</td>
<td>70K</td>
<td>No</td>
</tr>
<tr>
<td>4</td>
<td>Yes</td>
<td>Married</td>
<td>120K</td>
<td>No</td>
</tr>
<tr>
<td>5</td>
<td>No</td>
<td>Divorced</td>
<td>95K</td>
<td>Yes</td>
</tr>
<tr>
<td>6</td>
<td>No</td>
<td>Married</td>
<td>60K</td>
<td>No</td>
</tr>
<tr>
<td>7</td>
<td>Yes</td>
<td>Divorced</td>
<td>220K</td>
<td>No</td>
</tr>
<tr>
<td>8</td>
<td>No</td>
<td>Single</td>
<td>85K</td>
<td>Yes</td>
</tr>
<tr>
<td>9</td>
<td>No</td>
<td>Married</td>
<td>75K</td>
<td>No</td>
</tr>
<tr>
<td>10</td>
<td>No</td>
<td>Single</td>
<td>90K</td>
<td>Yes</td>
</tr>
</tbody>
</table>

#### 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

Normal distribution:

\[ P(X_i | Y_j) = \frac{1}{\sqrt{2\pi\sigma_j^2}} e^{-\frac{(X_i - \mu_j)^2}{2\sigma_j^2}} \]

- One for each \((X_i, Y_j)\) pair

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

\[
P(\text{Income} = 120 | \text{No}) = \frac{1}{\sqrt{2\pi(54.54)}} e^{-\frac{(120-110)^2}{2(2975)}} = 0.0072
\]

Example of Naïve Bayes Classifier

Given a Test Record:

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

Naïve Bayes Classifier:

\[
P(\text{Refund} = \text{Yes} | \text{No}) = 3/7
\]
\[
P(\text{Refund} = \text{No} | \text{No}) = 4/7
\]
\[
P(\text{Refund} = \text{Yes} | \text{Yes}) = 0
\]
\[
P(\text{Refund} = \text{No} | \text{Yes}) = 1
\]
\[
P(\text{Marital Status} = \text{Single} | \text{No}) = 2/7
\]
\[
P(\text{Marital Status} = \text{Divorced} | \text{No}) = 1/7
\]
\[
P(\text{Marital Status} = \text{Married} | \text{No}) = 4/7
\]
\[
P(\text{Marital Status} = \text{Single} | \text{Yes}) = 2/3
\]
\[
P(\text{Marital Status} = \text{Divorced} | \text{Yes}) = 1/3
\]
\[
P(\text{Marital Status} = \text{Married} | \text{Yes}) = 0
\]

For Taxable Income:

- If class = No: sample mean = 110
  - sample variance = 2975
- If class = Yes: sample mean = 90
  - sample variance = 25

\[
\text{Since } P(X|\text{No})P(\text{No}) > P(X|\text{Yes})P(\text{Yes})
\]

Therefore \(P(\text{No}|X) > P(\text{Yes}|X)\)

\(\Rightarrow\) Class = No
Example of Naïve Bayes Classifier

Given a Test Record:

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

Naïve Bayes Classifier:

- \( P(\text{Yes}) = 3/10 \)
  - \( P(\text{No}) = 7/10 \)

\[ P(\text{Refund} = \text{Yes} \mid \text{No}) = 3/7 \]
\[ P(\text{Refund} = \text{No} \mid \text{No}) = 4/7 \]
\[ P(\text{Refund} = \text{Yes} \mid \text{Yes}) = 0 \]
\[ P(\text{Refund} = \text{No} \mid \text{Yes}) = 1 \]
\[ P(\text{Marital Status} = \text{Single} \mid \text{No}) = 2/7 \]
\[ P(\text{Marital Status} = \text{Divorced} \mid \text{No}) = 1/7 \]
\[ P(\text{Marital Status} = \text{Married} \mid \text{No}) = 4/7 \]
\[ P(\text{Marital Status} = \text{Single} \mid \text{Yes}) = 2/3 \]
\[ P(\text{Marital Status} = \text{Divorced} \mid \text{Yes}) = 1/3 \]
\[ P(\text{Marital Status} = \text{Married} \mid \text{Yes}) = 0 \]

For Taxable Income:

- \( P(\text{Yes} \mid \text{Divorced}) = 1/3 \times 3/10 / P(\text{Divorced}) \)
  - \( P(\text{No} \mid \text{Divorced}) = 1/7 \times 7/10 / P(\text{Divorced}) \)

Naïve Bayes Classifier:

- \( P(\text{Yes} \mid \text{Refund} = \text{No}, \text{Divorced}) = 1 \times 1/3 \times 3/10 / P(\text{Divorced}, \text{Refund} = \text{No}) \)
  - \( P(\text{No} \mid \text{Refund} = \text{No}, \text{Divorced}) = 4/7 \times 1/7 \times 7/10 / P(\text{Divorced}, \text{Refund} = \text{No}) \)

Issues with Naïve Bayes Classifier

Naïve Bayes Classifier:

- \( P(\text{Yes}) = 3/10 \)
  - \( P(\text{No}) = 7/10 \)

\[ P(\text{Refund} = \text{Yes} \mid \text{No}) = 3/7 \]
\[ P(\text{Refund} = \text{No} \mid \text{No}) = 4/7 \]
\[ P(\text{Refund} = \text{Yes} \mid \text{Yes}) = 0 \]
\[ P(\text{Refund} = \text{No} \mid \text{Yes}) = 1 \]
\[ P(\text{Marital Status} = \text{Single} \mid \text{No}) = 2/7 \]
\[ P(\text{Marital Status} = \text{Divorced} \mid \text{No}) = 1/7 \]
\[ P(\text{Marital Status} = \text{Married} \mid \text{No}) = 4/7 \]
\[ P(\text{Marital Status} = \text{Single} \mid \text{Yes}) = 2/3 \]
\[ P(\text{Marital Status} = \text{Divorced} \mid \text{Yes}) = 1/3 \]
\[ P(\text{Marital Status} = \text{Married} \mid \text{Yes}) = 0 \]

For Taxable Income:

- \( P(\text{Yes} \mid \text{Married}) = 0 \times 3/10 / P(\text{Married}) \)
  - \( P(\text{No} \mid \text{Married}) = 4/7 \times 7/10 / P(\text{Married}) \)

02/03/2018
Introduction to Data Mining
### Issues with Naïve Bayes Classifier

**Naïve Bayes Classifier:**

\[
P(\text{Refund} = \text{Yes} \mid \text{No}) = \frac{2}{6} \\
P(\text{Refund} = \text{No} \mid \text{No}) = \frac{4}{6} \\
P(\text{Refund} = \text{Yes} \mid \text{Yes}) = 0 \\
P(\text{Refund} = \text{No} \mid \text{Yes}) = 1 \\
P(\text{Marital Status} = \text{Single} \mid \text{No}) = \frac{2}{6} \\
P(\text{Marital Status} = \text{Divorced} \mid \text{No}) = 0 \\
P(\text{Marital Status} = \text{Married} \mid \text{No}) = \frac{4}{6} \\
P(\text{Marital Status} = \text{Single} \mid \text{Yes}) = \frac{2}{3} \\
P(\text{Marital Status} = \text{Divorced} \mid \text{Yes}) = \frac{1}{3} \\
P(\text{Marital Status} = \text{Married} \mid \text{Yes}) = \frac{0}{3}
\]

For Taxable Income:

- If class = No: sample mean = 91
- sample variance = 685
- If class = No: sample mean = 90
- sample variance = 25

### Consider the table with Tid = 7 deleted

<table>
<thead>
<tr>
<th>Tid</th>
<th>Refund</th>
<th>Marital Status</th>
<th>Taxable Income</th>
<th>Evade</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Yes</td>
<td>Single</td>
<td>125K</td>
<td>No</td>
</tr>
<tr>
<td>2</td>
<td>No</td>
<td>Married</td>
<td>100K</td>
<td>No</td>
</tr>
<tr>
<td>3</td>
<td>No</td>
<td>Single</td>
<td>70K</td>
<td>No</td>
</tr>
<tr>
<td>4</td>
<td>Yes</td>
<td>Married</td>
<td>120K</td>
<td>No</td>
</tr>
<tr>
<td>5</td>
<td>No</td>
<td>Divorced</td>
<td>95K</td>
<td>Yes</td>
</tr>
<tr>
<td>6</td>
<td>No</td>
<td>Married</td>
<td>60K</td>
<td>No</td>
</tr>
<tr>
<td>7</td>
<td>Yes</td>
<td>Divorced</td>
<td>220K</td>
<td>No</td>
</tr>
<tr>
<td>8</td>
<td>No</td>
<td>Single</td>
<td>85K</td>
<td>Yes</td>
</tr>
<tr>
<td>9</td>
<td>No</td>
<td>Married</td>
<td>75K</td>
<td>No</td>
</tr>
<tr>
<td>10</td>
<td>No</td>
<td>Single</td>
<td>90K</td>
<td>Yes</td>
</tr>
</tbody>
</table>

Given \( X = (\text{Refund} = \text{Yes}, \text{Divorced}, 120K) \)

\[
P(X \mid \text{No}) = \frac{2}{6} \times 0 \times 0.0083 = 0 \\
P(X \mid \text{Yes}) = 0 \times \frac{1}{3} \times 1.2 \times 10^{-9} = 0
\]

**Naïve Bayes will not be able to classify \( X \) as \text{Yes} or \text{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:

\[
\text{Original: } P(A_i \mid C) = \frac{N_{ic}}{N_c} \\
\text{Laplace: } P(A_i \mid C) = \frac{N_{ic} + 1}{N_c + c} \\
\text{m-estimate: } P(A_i \mid C) = \frac{N_{ic} + mp}{N_c + m}
\]

- \( c \): number of classes
- \( p \): prior probability of the class
- \( m \): parameter
- \( N_c \): number of instances in the class
- \( N_{ic} \): number of instances having attribute value \( A_i \) in class \( c \)
Example of Naïve Bayes Classifier

<table>
<thead>
<tr>
<th>Name</th>
<th>Give Birth</th>
<th>Can Fly</th>
<th>Live in Water</th>
<th>Have Legs</th>
<th>Class</th>
</tr>
</thead>
<tbody>
<tr>
<td>human</td>
<td>yes</td>
<td>no</td>
<td>no</td>
<td>yes</td>
<td>mammals</td>
</tr>
<tr>
<td>python</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>non-mammals</td>
</tr>
<tr>
<td>salmon</td>
<td>no</td>
<td>yes</td>
<td>no</td>
<td>no</td>
<td>non-mammals</td>
</tr>
<tr>
<td>frog</td>
<td>no</td>
<td>no</td>
<td>sometimes</td>
<td>yes</td>
<td>non-mammals</td>
</tr>
<tr>
<td>crocodile</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>yes</td>
<td>non-mammals</td>
</tr>
<tr>
<td>bat</td>
<td>yes</td>
<td>yes</td>
<td>no</td>
<td>no</td>
<td>mammals</td>
</tr>
<tr>
<td>penguin</td>
<td>no</td>
<td>yes</td>
<td>no</td>
<td>yes</td>
<td>non-mammals</td>
</tr>
<tr>
<td>cat</td>
<td>yes</td>
<td>no</td>
<td>no</td>
<td>yes</td>
<td>mammals</td>
</tr>
<tr>
<td>leopon shark</td>
<td>yes</td>
<td>no</td>
<td>yes</td>
<td>no</td>
<td>non-mammals</td>
</tr>
<tr>
<td>turtle</td>
<td>no</td>
<td>yes</td>
<td>sometimes</td>
<td>yes</td>
<td>non-mammals</td>
</tr>
<tr>
<td>penguin</td>
<td>no</td>
<td>no</td>
<td>sometimes</td>
<td>yes</td>
<td>non-mammals</td>
</tr>
<tr>
<td>porcupine</td>
<td>yes</td>
<td>no</td>
<td>no</td>
<td>yes</td>
<td>mammals</td>
</tr>
<tr>
<td>seal</td>
<td>no</td>
<td>no</td>
<td>yes</td>
<td>no</td>
<td>non-mammals</td>
</tr>
<tr>
<td>salamander</td>
<td>no</td>
<td>no</td>
<td>sometimes</td>
<td>yes</td>
<td>non-mammals</td>
</tr>
<tr>
<td>gila monster</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>yes</td>
<td>mammals</td>
</tr>
<tr>
<td>platypus</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>yes</td>
<td>mammals</td>
</tr>
<tr>
<td>owl</td>
<td>no</td>
<td>yes</td>
<td>yes</td>
<td>no</td>
<td>non-mammals</td>
</tr>
<tr>
<td>dolphin</td>
<td>yes</td>
<td>no</td>
<td>yes</td>
<td>no</td>
<td>non-mammals</td>
</tr>
</tbody>
</table>

A: attributes
M: mammals
N: non-mammals

$$P(A|M) = \frac{6}{7} \times \frac{6}{7} \times \frac{2}{7} = 0.06$$

$$P(A|N) = \frac{1}{13} \times \frac{10}{13} \times \frac{3}{13} \times \frac{4}{13} = 0.0042$$

$$P(A|M)P(M) = \frac{0.06 \times 7}{20} = 0.021$$

$$P(A|N)P(N) = \frac{0.004 \times 13}{20} = 0.0027$$

$$P(A|M)P(M) > P(A|N)P(N)$$

⇒ Mammals

Naïve Bayes (Summary)

- Robust to isolated noise points
- Handle missing values by ignoring the instance during probability estimate calculations
- Robust to irrelevant attributes
- Independence assumption may not hold for some attributes
  - 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

Naïve Bayes

- How does Naïve Bayes perform on the following dataset?

Naïve Bayes can construct oblique decision boundaries
Naïve Bayes

- How does Naïve Bayes perform on the following dataset?

<table>
<thead>
<tr>
<th>Y = 1</th>
<th>1</th>
<th>1</th>
<th>1</th>
<th>0</th>
</tr>
</thead>
<tbody>
<tr>
<td>Y = 2</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Y = 3</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Y = 4</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

X = 1    X = 2    X = 3    X = 4

Conditional independence of attributes is violated

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
    - Arc corresponds to dependence relationship between a pair of variables
  - A probability table associating each node to its immediate parent
Conditional Independence

A node in a Bayesian network is conditionally independent of all of its nondescendants, if its parents are known.

Naïve Bayes assumption:
**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**

<table>
<thead>
<tr>
<th>Exercise</th>
<th>Diet</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes</td>
<td>0.7</td>
</tr>
<tr>
<td>No</td>
<td>0.3</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Diet</th>
</tr>
</thead>
<tbody>
<tr>
<td>Healthy</td>
</tr>
<tr>
<td>Unhealthy</td>
</tr>
</tbody>
</table>

| Exercise=Yes | 0.7 | Exercise=No | 0.3 |
| Diet=Healthy | 0.25 | Diet=Unhealthy | 0.75 |

<table>
<thead>
<tr>
<th>Exercise</th>
<th>Diet</th>
<th>Heart Disease</th>
</tr>
</thead>
<tbody>
<tr>
<td>Healthy</td>
<td>Healthy</td>
<td>0.25</td>
</tr>
<tr>
<td>Healthy</td>
<td>Unhealthy</td>
<td>0.45</td>
</tr>
<tr>
<td>Unhealthy</td>
<td>Healthy</td>
<td>0.75</td>
</tr>
<tr>
<td>Unhealthy</td>
<td>Unhealthy</td>
<td>0.25</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>HD=Yes</th>
<th>HD=No</th>
</tr>
</thead>
<tbody>
<tr>
<td>CP=Yes</td>
<td>0.8</td>
</tr>
<tr>
<td>BP=High</td>
<td>0.95</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>HD=Yes</th>
<th>HD=No</th>
</tr>
</thead>
<tbody>
<tr>
<td>BP=High</td>
<td>0.8</td>
</tr>
</tbody>
</table>
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.