Bayesian Classifiers

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:
  \[ P(Y \mid X) = \frac{P(X \mid Y)P(Y)}{P(X)} \]
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?

<table>
<thead>
<tr>
<th>Tit</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>

Using Bayes Theorem for Classification

• Approach:
  – compute posterior probability \(P(Y| X_1, X_2, \ldots, X_d)\) using the Bayes theorem

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

  – \textit{Maximum a-posteriori}: Choose \(Y\) that maximizes \(P(Y| X_1, X_2, \ldots, X_d)\)

  – Equivalent to choosing value of \(Y\) that maximizes \(P(X_1, X_2, \ldots, X_d|Y)P(Y)\)

• How to estimate \(P(X_1, X_2, \ldots, X_d | Y)\)?
Example Data

Given a Test Record:

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

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

In the following we will replace
- \( \text{Evade} = \text{Yes} \) by \( \text{Yes} \), and
- \( \text{Evade} = \text{No} \) by \( \text{No} \)

Using Bayes Theorem:

- \[ P(\text{Yes} | X) = \frac{P(X | \text{Yes}) P(\text{Yes})}{P(X)} \]
- \[ P(\text{No} | X) = \frac{P(X | \text{No}) P(\text{No})}{P(X)} \]

How to estimate \( P(X | \text{Yes}) \) and \( P(X | \text{No}) \)?
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 Classifier

• Assume independence among attributes \( X_i \) when class is given:
  \( P(X_1, X_2, \ldots, X_d | Y_j) = P(X_1| Y_j) \cdot P(X_2| Y_j) \cdots P(X_d| Y_j) \)

  – Now we can estimate \( P(X_i| Y_j) \) for all \( X_i \) and \( Y_j \) combinations from the training data

  – New point is classified to \( Y_j \) if \( P(Y_j) \cdot \prod P(X_i| Y_j) \) is maximal.
Naïve Bayes on Example Data

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

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

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

Estimate Probabilities from Data

- \( P(y) = \) fraction of instances of class \( y \)
  - e.g., \( P(\text{No}) = 7/10, \)
    \( P(\text{Yes}) = 3/10 \)

- For categorical attributes:
  \( P(X_i = c | y) = \frac{n_c}{n} \)
  - where \( |X_i = c| \) is number of instances having attribute value \( X_i = c \) and belonging to class \( y \)
  - Examples:
    \( P(\text{Status} = \text{Married} | \text{No}) = 4/7 \)
    \( P(\text{Refund} = \text{Yes} | \text{Yes}) = 0 \)
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 $P(X_i|Y)$

<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>

**Normal distribution**:

$$P(X_i|Y_j) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(X_i-\mu_j)^2}{2\sigma^2}}$$

– One for each $(X_i, Y_j)$ pair

• For (Income, Class=No):
  – If Class=No
    – Sample mean = 110
    – Sample variance = 2975

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

Given a Test Record:

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

Naïve Bayes Classifier:

\[
P(\text{Yes}) = \frac{3}{10} \\
P(\text{No}) = \frac{7}{10}
\]

If we only know that marital status is Divorced, then:

\[
P(\text{Yes} | \text{Divorced}) = \frac{1}{3} \times \frac{3}{10} / P(\text{Divorced}) \\
P(\text{No} | \text{Divorced}) = \frac{1}{7} \times \frac{7}{10} / P(\text{Divorced})
\]

If we also know that Refund = No, then

\[
P(\text{Yes} | \text{Refund = No, Divorced}) = \frac{1}{3} \times \frac{1}{3} \times \frac{3}{10} / P(\text{Divorced, Refund = No}) \\
P(\text{No} | \text{Refund = No, Divorced}) = \frac{1}{7} \times \frac{1}{7} \times \frac{7}{10} / P(\text{Divorced, Refund = No})
\]

If we also know that Taxable Income = 120, then

\[
P(\text{Yes} | \text{Refund = No, Divorced, Income = 120}) = 1.2 \times 10^{-6} \times \frac{1}{3} \times \frac{1}{3} \times \frac{3}{10} / P(\text{Divorced, Refund = No, Income = 120}) \\
P(\text{No} | \text{Refund = No, Divorced, Income = 120}) = 0.0072 \times \frac{4}{7} \times \frac{1}{7} \times \frac{7}{10} / P(\text{Divorced, Refund = No, Income = 120})
\]

Even in absence of information about any attributes, we can use Apriori Probabilities of Class Variable:

Naïve Bayes Classifier:

\[
P(\text{Yes}) = \frac{3}{10} \\
P(\text{No}) = \frac{7}{10}
\]
Issues with Naïve Bayes Classifier

**Given a Test Record:**

\[ X = (\text{Married}) \]

Naïve Bayes Classifier:

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

For Taxable Income:

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

Naïve Bayes Classifier:

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

For Taxable Income:

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

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

\[ P(X \mid \text{No}) = 2/6 \times 0 \times 0.0083 = 0 \]

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

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: \( P(X_i = c \mid y) = \frac{n_c}{n} \)

Laplace Estimate: \( P(X_i = c \mid y) = \frac{n_c + 1}{n + v} \)

m – estimate: \( P(X_i = c \mid y) = \frac{n_c + mp}{n + m} \)

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>no</td>
<td>no</td>
<td>no</td>
<td>non-mammals</td>
</tr>
<tr>
<td>whale</td>
<td>yes</td>
<td>yes</td>
<td>no</td>
<td>yes</td>
<td>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>komodo</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>no</td>
<td>no</td>
<td>yes</td>
<td>mammals</td>
</tr>
<tr>
<td>pigeon</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>yes</td>
<td>no</td>
<td>yes</td>
<td>mammals</td>
</tr>
<tr>
<td>leopard shark</td>
<td>yes</td>
<td>no</td>
<td>yes</td>
<td>yes</td>
<td>mammals</td>
</tr>
<tr>
<td>turtle</td>
<td>yes</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>yes</td>
<td>non-mammals</td>
</tr>
<tr>
<td>salamander</td>
<td>no</td>
<td>no</td>
<td>yes</td>
<td>yes</td>
<td>non-mammals</td>
</tr>
<tr>
<td>pika monster</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>yes</td>
<td>non-mammals</td>
</tr>
<tr>
<td>platypus</td>
<td>no</td>
<td>no</td>
<td>yes</td>
<td>no</td>
<td>mammals</td>
</tr>
<tr>
<td>penguin</td>
<td>no</td>
<td>yes</td>
<td>yes</td>
<td>no</td>
<td>non-mammals</td>
</tr>
<tr>
<td>dolphin</td>
<td>no</td>
<td>yes</td>
<td>yes</td>
<td>no</td>
<td>non-mammals</td>
</tr>
<tr>
<td>eagle</td>
<td>no</td>
<td>yes</td>
<td>yes</td>
<td>no</td>
<td>non-mammals</td>
</tr>
</tbody>
</table>

A: attributes
M: mammals
N: non-mammals

\[
P(A \mid M) = \frac{6}{7} \times \frac{6}{7} \times \frac{2}{7} = 0.06
\]

\[
P(A \mid N) = \frac{1}{13} \times \frac{10}{13} \times \frac{3}{13} \times \frac{4}{13} = 0.0042
\]

\[
P(A \mid M)P(M) = 0.06 \times \frac{7}{20} = 0.021
\]

\[
P(A \mid N)P(N) = 0.004 \times \frac{13}{20} = 0.0027
\]

\[
P(A \mid M)P(M) > P(A \mid N)P(N)
\]

=> Mammals
Naïve Bayes (Summary)

• Robust to isolated noise points

• Handle missing values by ignoring the instance during probability estimate calculations

• Robust to irrelevant attributes

• 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
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
Conditional Independence

- Naïve Bayes assumption:

\[ Y \]
\[ X_1 \quad X_2 \quad X_3 \quad X_4 \quad \ldots \quad X_d \]

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    Diet=Healthy 0.25
Exercise=No 0.3     Diet=Unhealthy 0.75

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