Machine Learning

Dr. Rajesh Kumar

PhD, PDF (NUS, Singapore)
SMIEEE, FIET (UK), FIETE, FIE (I), SMIACSIT, LMISTE, MIAENG
Professor, Department of Electrical Engineering
Professor, Centre of Energy and Environment
Malaviya National Institute of Technology, Jaipur, India, 302017
Tel: (91) 9549654481
http://drrajeshkumar.wordpress.com
rkumar.ee@mnit.ac.in, rkumar.ee@gmail.com
Naive Bayes
Content

- Background
- Probability Basics
- Probabilistic Classification
- Naïve Bayes
  - Principle and Algorithms
  - Play Tennis
- Zero Conditional Probability
- Summary
Probability Basics

- Prior, conditional and joint probability for random variables
  - Prior probability: \( P(x) \)
  - Conditional probability: \( P(x_1|x_2), P(x_2|x_1) \)
  - Joint probability: \( x = (x_1, x_2), P(x) = P(x_1, x_2) \)
  - Relationship: \( P(x_1, x_2) = P(x_2|x_1)P(x_1) = P(x_1|x_2)P(x_2) \)
  - Independence: \( P(x_2|x_1) = P(x_2), P(x_1|x_2) = P(x_1), P(x_1,x_2) = \)

- Bayesian Rule

\[
P(c|x) = \frac{P(x|c)P(c)}{P(x)}
\]

\[
Posterior = \frac{Likelihood \times Prior}{Evidence}
\]
Probabilistic Classification

- **Maximum A Posterior (MAP) classification rule**
  - For an input $\mathbf{x}$, find the largest one from $L$ probabilities output by a discriminative probabilistic classifier $P(c_1|\mathbf{x}), \ldots, P(c_L|\mathbf{x})$.
  - Assign $\mathbf{x}$ to label $c^*$ if $P(c^*|\mathbf{x})$ is the largest.

- **Generative classification with the MAP rule**
  - Apply Bayesian rule to convert them into posterior probabilities
    \[
    P(c_i|\mathbf{x}) = \frac{P(\mathbf{x}|c_i)P(c_i)}{P(\mathbf{x})} \propto P(\mathbf{x}|c_i)P(c_i)
    \]
    \[\text{for } i = 1, 2, \ldots, L\]
  - Then apply the MAP rule to assign a label
Spam Detector

“Buy” and “Cheap”

Spam

No spam

https://drrajeshkumar.wordpress.com
Spam Detector

“Buy” and “Cheap”

<table>
<thead>
<tr>
<th>Spam</th>
<th>No spam</th>
</tr>
</thead>
<tbody>
<tr>
<td>100%</td>
<td>0%</td>
</tr>
</tbody>
</table>

**Quiz:** If an e-mail contains the words “buy” and “cheap”, what is the probability that it is spam?

- 40%
- 60%
- 80%
- 100%
Solution: Collect more data?
Spam Detector

“Buy” and “Cheap”

Spam

12 e-mails

No spam

0 e-mails?

Guess?
Spam Detector

25 e-mails
20 “Buy”
15 Cheap

4/5 \times 25 = 12 “Buy”
3/5 \times 25 = 12 “Cheap”

\frac{12}{25} \times 25 = 12 “Buy” and “Cheap”
Spam Detector

No spam

75 e-mails
5 “Buy” 1/15
10 “Cheap” 2/15 2/225
Spam Detector

No spam

75 e-mails
5 “Buy” 1/15
10 “Cheap” 2/15

2/225 x 75 = 2/3 “Buy” and “Cheap”
Spam Detector

“Buy” and “Cheap”

<table>
<thead>
<tr>
<th>Spam</th>
<th>No spam</th>
</tr>
</thead>
<tbody>
<tr>
<td>![Alert] 12</td>
<td>![Alert] 2/3</td>
</tr>
<tr>
<td>94.737%</td>
<td>5.263%</td>
</tr>
</tbody>
</table>

Quiz: If an e-mail contains the words “buy” and “cheap”, what is the probability that it is spam?

\[
\frac{12}{12 + \frac{2}{3}} = \frac{36}{38} = 94.737\%
\]
Naive Bayes

<table>
<thead>
<tr>
<th></th>
<th>Spam</th>
<th>No spam</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Total</strong></td>
<td>25</td>
<td>75</td>
</tr>
<tr>
<td><strong>Buy</strong></td>
<td>20</td>
<td>4/5</td>
</tr>
<tr>
<td><strong>Cheap</strong></td>
<td>15</td>
<td>3/5</td>
</tr>
<tr>
<td><strong>Buy &amp; Cheap</strong></td>
<td>12</td>
<td>12/25</td>
</tr>
</tbody>
</table>

\[
\frac{12}{12 + \frac{2}{3}} = \frac{36}{38}
\]
# Naive Bayes

<table>
<thead>
<tr>
<th></th>
<th>Spam</th>
<th>No Spam</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Total</strong></td>
<td>25</td>
<td>75</td>
</tr>
<tr>
<td><strong>Buy</strong></td>
<td>20</td>
<td>4/5</td>
</tr>
<tr>
<td><strong>Cheap</strong></td>
<td>15</td>
<td>3/5</td>
</tr>
<tr>
<td><strong>Work</strong></td>
<td>5</td>
<td>1/5</td>
</tr>
<tr>
<td><strong>Buy, Cheap, &amp; Work</strong></td>
<td>12/5</td>
<td>12/125</td>
</tr>
</tbody>
</table>

\[
\frac{12/5}{12/5 + 4/15} = \frac{36}{40} = 90\%
\]
Bayes Theorem

S: Spam
H: Ham (not spam)
B: ‘Buy’

\[
P(S \mid B) = \frac{P(B \mid S) \cdot P(S)}{P(B \mid S) \cdot P(S) + P(B \mid H) \cdot P(H)}
\]

\[
P(\text{spam if “Buy”}) = \frac{\frac{20}{25} \cdot \frac{25}{100}}{\frac{20}{25} \cdot \frac{25}{100} + \frac{5}{75} \cdot \frac{75}{100}} = 80\%
\]
Naive Bayes

S: Spam
H: Ham (not spam)
B: ‘Buy’
C: ‘Cheap’

\[
P(S \mid B \cap C) = \frac{P(B \mid S)P(C \mid S)P(S)}{P(B \mid S)P(C \mid S)P(S) + P(B \mid H)P(C \mid H)P(H)}
\]

\[
P(\text{spam if "Buy" & "Cheap"}) = \frac{20}{25} \cdot \frac{15}{25} \cdot \frac{25}{100} + \frac{5}{75} \cdot \frac{10}{75} \cdot \frac{75}{100}
\]

= 94.737%
Example: Play Tennis

PlayTennis: training examples

<table>
<thead>
<tr>
<th>Day</th>
<th>Outlook</th>
<th>Temperature</th>
<th>Humidity</th>
<th>Wind</th>
<th>PlayTennis</th>
</tr>
</thead>
<tbody>
<tr>
<td>D1</td>
<td>Sunny</td>
<td>Hot</td>
<td>High</td>
<td>Weak</td>
<td>No</td>
</tr>
<tr>
<td>D2</td>
<td>Sunny</td>
<td>Hot</td>
<td>High</td>
<td>Strong</td>
<td>No</td>
</tr>
<tr>
<td>D3</td>
<td>Overcast</td>
<td>Hot</td>
<td>High</td>
<td>Weak</td>
<td>Yes</td>
</tr>
<tr>
<td>D4</td>
<td>Rain</td>
<td>Mild</td>
<td>High</td>
<td>Weak</td>
<td>Yes</td>
</tr>
<tr>
<td>D5</td>
<td>Rain</td>
<td>Cool</td>
<td>Normal</td>
<td>Weak</td>
<td>Yes</td>
</tr>
<tr>
<td>D6</td>
<td>Rain</td>
<td>Cool</td>
<td>Normal</td>
<td>Strong</td>
<td>No</td>
</tr>
<tr>
<td>D7</td>
<td>Overcast</td>
<td>Cool</td>
<td>Normal</td>
<td>Strong</td>
<td>Yes</td>
</tr>
<tr>
<td>D8</td>
<td>Sunny</td>
<td>Mild</td>
<td>High</td>
<td>Weak</td>
<td>No</td>
</tr>
<tr>
<td>D9</td>
<td>Sunny</td>
<td>Cool</td>
<td>Normal</td>
<td>Weak</td>
<td>Yes</td>
</tr>
<tr>
<td>D10</td>
<td>Rain</td>
<td>Mild</td>
<td>Normal</td>
<td>Weak</td>
<td>Yes</td>
</tr>
<tr>
<td>D11</td>
<td>Sunny</td>
<td>Mild</td>
<td>Normal</td>
<td>Strong</td>
<td>Yes</td>
</tr>
<tr>
<td>D12</td>
<td>Overcast</td>
<td>Mild</td>
<td>High</td>
<td>Strong</td>
<td>Yes</td>
</tr>
<tr>
<td>D13</td>
<td>Overcast</td>
<td>Hot</td>
<td>Normal</td>
<td>Weak</td>
<td>Yes</td>
</tr>
<tr>
<td>D14</td>
<td>Rain</td>
<td>Mild</td>
<td>High</td>
<td>Strong</td>
<td>No</td>
</tr>
</tbody>
</table>
Example: Play Tennis

- **Learning Phase**

<table>
<thead>
<tr>
<th>Outlook</th>
<th>Play=Yes</th>
<th>Play=No</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sunny</td>
<td>2/9</td>
<td>3/5</td>
</tr>
<tr>
<td>Overcast</td>
<td>4/9</td>
<td>0/5</td>
</tr>
<tr>
<td>Rain</td>
<td>3/9</td>
<td>2/5</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Temperature</th>
<th>Play=Yes</th>
<th>Play=No</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hot</td>
<td>2/9</td>
<td>2/5</td>
</tr>
<tr>
<td>Mild</td>
<td>4/9</td>
<td>2/5</td>
</tr>
<tr>
<td>Cool</td>
<td>3/9</td>
<td>1/5</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Humidity</th>
<th>Play=Yes</th>
<th>Play=No</th>
</tr>
</thead>
<tbody>
<tr>
<td>High</td>
<td>3/9</td>
<td>4/5</td>
</tr>
<tr>
<td>Normal</td>
<td>6/9</td>
<td>1/5</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Wind</th>
<th>Play=Yes</th>
<th>Play=No</th>
</tr>
</thead>
<tbody>
<tr>
<td>Strong</td>
<td>3/9</td>
<td>3/5</td>
</tr>
<tr>
<td>Weak</td>
<td>6/9</td>
<td>2/5</td>
</tr>
</tbody>
</table>

\[ P(\text{Play}=\text{Yes}) = \frac{9}{14} \quad P(\text{Play}=\text{No}) = \frac{5}{14} \]
Example: Play Tennis

- **Test Phase**
  - Given a new instance, predict its label
    \[ x'=(\text{Outlook}=\text{Sunny}, \text{Temperature}=\text{Cool}, \text{Humidity}=\text{High}, \text{Wind}=\text{Strong}) \]
  - Look up tables achieved in the learning phrase
    
    \[
    \begin{align*}
    P(\text{Outlook}=\text{Sunny}|\text{Play}=\text{Yes}) &= 2/9 & P(\text{Outlook}=\text{Sunny}|\text{Play}=\text{No}) &= 3/5 \\
    P(\text{Temperature}=\text{Cool}|\text{Play}=\text{Yes}) &= 3/9 & P(\text{Temperature}=\text{Cool}|\text{Play}=\text{No}) &= 1/5 \\
    P(\text{Humidity}=\text{High}|\text{Play}=\text{Yes}) &= 3/9 & P(\text{Humidity}=\text{High}|\text{Play}=\text{No}) &= 4/5 \\
    P(\text{Wind}=\text{Strong}|\text{Play}=\text{Yes}) &= 3/9 & P(\text{Wind}=\text{Strong}|\text{Play}=\text{No}) &= 3/5 \\
    P(\text{Play}=\text{Yes}) &= 9/14 & P(\text{Play}=\text{No}) &= 5/14
    \end{align*}
    \]

  - Decision making with the MAP rule
    \[
    \begin{align*}
    P(\text{Yes}|x') &\approx [P(\text{Sunny}|\text{Yes})P(\text{Cool}|\text{Yes})P(\text{High}|\text{Yes})P(\text{Strong}|\text{Yes})]P(\text{Play}=\text{Yes}) = 0.0053 \\
    P(\text{No}|x') &\approx [P(\text{Sunny}|\text{No}) P(\text{Cool}|\text{No})P(\text{High}|\text{No})P(\text{Strong}|\text{No})]P(\text{Play}=\text{No}) = 0.0206
    \end{align*}
    \]

  
  Given the fact \( P(\text{Yes}|x') < P(\text{No}|x') \), we label \( x' \) to be “No”.
Naive Bayes

- Example: Continuous-valued Features
  - Temperature is naturally of continuous value.
  
  **Yes:** 25.2, 19.3, 18.5, 21.7, 20.1, 24.3, 22.8, 23.1, 19.8
  
  **No:** 27.3, 30.1, 17.4, 29.5, 15.1

- Estimate mean and variance for each class

\[ \mu = \frac{1}{N} \sum_{n=1}^{N} x_n, \quad \sigma^2 = \frac{1}{N} \sum_{n=1}^{N} (x_n - \mu)^2 \]

- **Learning Phase:** output two Gaussian models for \( P(\text{temp}|C) \)

\[
\hat{P}(x|\text{Yes}) = \frac{1}{2.35\sqrt{2\pi}} \exp\left(-\frac{(x-21.64)^2}{2 \times 2.35^2}\right) = \frac{1}{2.35\sqrt{2\pi}} \exp\left(-\frac{(x-21.64)^2}{11.09}\right)
\]

\[
\hat{P}(x|\text{No}) = \frac{1}{7.09\sqrt{2\pi}} \exp\left(-\frac{(x-23.88)^2}{2 \times 7.09^2}\right) = \frac{1}{7.09\sqrt{2\pi}} \exp\left(-\frac{(x-23.88)^2}{50.25}\right)
\]