Lecture outline

• Classification
• Naïve Bayes classifier
Bayes Theorem

- \(X, Y\) random variables
- Joint probability: \(\Pr(X=x, Y=y)\)
- Conditional probability: \(\Pr(Y=y \mid X=x)\)
- Relationship between joint and conditional probability distributions

\[
\Pr(X, Y) = \Pr(X \mid Y) \times \Pr(Y) = \Pr(Y \mid X) \times \Pr(X)
\]

- **Bayes Theorem:**

\[
\Pr(Y \mid X) = \frac{\Pr(X \mid Y) \Pr(Y)}{\Pr(X)}
\]
Bayes Theorem for Classification

- **X**: attribute set
- **Y**: class variable
- **Y** depends on **X** in a non-deterministic way
- We can capture this dependence using \( \text{Pr}(Y|X) \): Posterior probability vs \( \text{Pr}(Y) \): Prior probability
Building the Classifier

• Training phase:
  – Learning the posterior probabilities $\text{Pr}(Y|X)$ for every combination of $X$ and $Y$ based on training data

• Test phase:
  – For test record $X'$, compute the class $Y'$ that maximizes the posterior probability $\text{Pr}(Y'|X')$
Bayes Classification: Example

\[ X' = (\text{Home Owner} = \text{No}, \ \text{Marital Status} = \text{Married}, \ \text{Annual Income} = 120K) \]

Compute: \( \text{Pr}(\text{Yes}|X'), \ \text{Pr}(\text{No}|X') \) pick No or Yes with max Prob.

How can we compute these probabilities??

<table>
<thead>
<tr>
<th>Tid</th>
<th>Home Owner</th>
<th>Marital Status</th>
<th>Annual Income</th>
<th>Defaulted Borrower</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>

Figure 4.6. Training set for predicting borrowers who will default on loan payments.
Computing posterior probabilities

• Bayes Theorem

\[
\Pr(Y | X) = \frac{\Pr(X | Y) \Pr(Y)}{\Pr(X)}
\]

• \( \Pr(X) \) is constant and can be ignored
• \( \Pr(Y) \): estimated from training data; compute the fraction of training records in each class
• \( \Pr(X|Y) \)?
Naïve Bayes Classifier

\[
\Pr(X \mid Y = y) = \prod_{i=1}^{d} \Pr(X_i \mid Y = y)
\]

• Attribute set \( X = \{X_1, \ldots, X_d\} \) consists of \( d \) attributes

• Conditional independence:
  – \( X \) conditionally independent of \( Y \), given \( X \):
    \[
    \Pr(X \mid Y, Z) = \Pr(X \mid Z)
    \]
  – \( \Pr(X, Y \mid Z) = \Pr(X \mid Z) \times \Pr(Y \mid Z) \)
Naïve Bayes Classifier

\[ \Pr(X|Y = y) = \prod_{i=1}^{d} \Pr(X_i|Y = y) \]

- Attribute set \( X = \{X_1, \ldots, X_d\} \) consists of \( d \) attributes

\[ \Pr(X|Y) = \frac{\Pr(Y) \prod_{i=1}^{d} \Pr(X_i|Y)}{\Pr(X)} \]
Conditional probabilities for categorical attributes

- Categorical attribute $X_i$
- $\Pr(X_i = x_i|Y=y)$: fraction of training instances in class $y$ that take value $x_i$ on the $i$-th attribute

$\Pr(\text{homeOwner} = \text{yes}|\text{No}) = \frac{3}{7}$

$\Pr(\text{MaritalStatus} = \text{Single}|\text{Yes}) = \frac{2}{3}$
Estimating conditional probabilities for continuous attributes?

- Discretization?

- How can we discretize?
Naïve Bayes Classifier: Example

• $X' = (\text{HomeOwner} = \text{No}, \text{MaritalStatus} = \text{Married}, \text{Income}=120K)$

• Need to compute $Pr(Y|X')$ or $Pr(Y) \times Pr(X'|Y)$

• But $Pr(X'|Y)$ is
  
  – $Y = \text{No}$:
    
    • $Pr(\text{HO}=\text{No}|\text{No}) \times Pr(\text{MS}=\text{Married}|\text{No}) \times Pr(\text{Inc}=120K|\text{No}) = \frac{4}{7} \times \frac{4}{7} \times 0.0072 = 0.0024$

  – $Y=\text{Yes}$:
    
    • $Pr(\text{HO}=\text{No}|\text{Yes}) \times Pr(\text{MS}=\text{Married}|\text{Yes}) \times Pr(\text{Inc}=120K|\text{Yes}) = 1 \times 0 \times 1.2 \times 10^{-9} = 0$
Naïve Bayes Classifier: Example

• $X' = (\text{HomeOwner} = \text{No}, \text{MaritalStatus} = \text{Married}, \text{Income} = 120K)$
• Need to compute $\Pr(Y|X')$ or $\Pr(Y) \times \Pr(X'|Y)$
• But $\Pr(X'|Y = \text{Yes})$ is 0?
• Correction process:

$$\Pr(X_i = x_i \mid Y = y_j) = \frac{n_c + mp}{n + m}$$

$n_c$: number of training examples from class $y_j$ that take value $x_i$
$n$: total number of instances from class $y_j$
m: equivalent sample size (balance between prior and posterior)
p: user-specified parameter (prior probability)
Characteristics of Naïve Bayes Classifier

• Robust to isolated noise points
  – noise points are averaged out
• Handles missing values
  – Ignoring missing-value examples
• Robust to irrelevant attributes
  – If $X_i$ is irrelevant, $P(X_i|Y)$ becomes almost uniform
• Correlated attributes degrade the performance of NB classifier