นางสาวอุมาพร คำภิชัย รหัสนักศึกษา 623020547-0 คณะวิทยาศาสตร์ สาขาสถิติ หลักสูตรสารสนเทศสถิติ ชั้นปีที่ 3

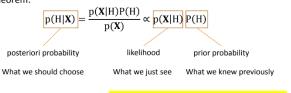
## สรุป Chapter 8 (ต่อ)

### **Bayes' Theorem: Basics**

■ Total probability Theorem:

$$p(B) = \sum_{i} p(B|A_{i})p(A_{i})$$

■ Bayes' Theorem:



- X: a data sample ("evidence")
- Prediction can be done based on Bayes' Theorem:
- ☐ H: X belongs to class C

Classification is to derive the maximum posteriori

#### Naïve Bayes Classifier: Making a Naïve Assumption

- Practical difficulty of Naïve Bayes inference: It requires initial knowledge of many probabilities, which may not be available or involving significant computational cost
- A Naïve Special Case
- Make an additional assumption to simplify the model, but achieve comparable performance.

attributes are conditionally independent (i.e., no dependence relation between attributes)

$$p(X|C_i) = \prod_k p(x_k|C_i) = p(x_1|C_i) \cdot p(x_2|C_i) \cdot \cdots \cdot p(x_n|C_i)$$

Only need to count the class distribution w.r.t. features

- สูตรในการหาค่า

-จะต้องมีความสัมพันธ์กันถึงจะหาค่าได้

# Naïve Bayes Classifier: Categorical vs. Continuous Valued Features

 $\Box$  If feature  $x_k$  is categorical,  $p(x_k = v_k | C_i)$  is the # of tuples in  $C_i$  with  $x_k = v_k$ , divided by  $|C_i|_D$  (# of tuples of  $C_i$  in D)

$$p(X|C_i) = \prod_k p(x_k|C_i) = p(x_1|C_i) \cdot p(x_2|C_i) \cdot \cdots \cdot p(x_n|C_i)$$

 $\Box$  If feature  $x_k$  is continuous-valued,  $p(x_k=v_k|C_i)$  is usually computed based on Gaussian distribution with a mean  $\mu$  and standard deviation  $\sigma$ 

$$p(\mathbf{x}_{k} = \mathbf{v}_{k} | C_{i}) = N(\mathbf{x}_{k} | \mu_{C_{i}}, \sigma_{C_{i}}) = \frac{1}{\sqrt{2\pi}\sigma_{C_{i}}} e^{-\frac{(x-\mu_{C_{i}})^{2}}{2\sigma^{2}}}$$

# Naïve Bayes Classifier: Training Dataset

Class:

C1:buys\_computer = 'yes' C2:buys\_computer = 'no'

Data to be classified:

X = (age <=30, Income = medium, Student = yes, Credit rating = Fair)

age	income	student	credit_rating	buys_computer
<=30	high	no	fair	no
<=30	high	no	excellent	no
3140	high	no	fair	yes
>40	medium	no	fair	yes
>40	low	yes	fair	yes
>40	low	yes	excellent	no
3140	low	yes	excellent	yes
<=30	medium	no	fair	no
<=30	low	yes	fair	yes
>40	medium	yes	fair	yes
<=30	medium	yes	excellent	yes
3140	medium	no	excellent	yes
3140	high	yes	fair	yes
>40	medium	no	excellent	no

- ตัวอย่างในการศึกษา
- แบ่งออกเป็น X และ Y

## Naïve Bayes Classifier: An Example

=30 high no excellent

40 low yes excellent

1...40 low yes excellent =30 medium no fair

40 medium yes fair =30 medium yes excellent

31...40 high no fair 40 medium no fair

40 low yes fair

=30 low yes fair

no

yes

yes

no

yes

yes

yes no

- $P(C_i)$ : P(buys computer = "yes") = 9/14 = 0.643  $P(buys\_computer = "no") = 5/14 = 0.357$
- $\square$  Compute  $P(X|C_i)$  for each class

```
P(age = "<=30" | buys\_computer = "yes") = 2/9 = 0.222
```

P(student = "yes" | buys\_computer = "no") = 
$$1/5 = 0.2$$

P(student = "yes" | buys\_computer = "no") = 
$$1/5 = 0.2$$

P(credit rating = "fair" | buys computer = "no") = 2/5 = 0.4

X = (age <= 30, income = medium, student = yes, credit\_rating = fair)</p>

 $P(X|C_i)$ :  $P(X|buys\_computer = "yes") = 0.222 x 0.444 x 0.667 x 0.667 = 0.044$  $P(X|buys computer = "no") = 0.6 \times 0.4 \times 0.2 \times 0.4 = 0.019$ 

 $P(X|C_i)*P(C_i): P(X|buys computer = "yes") * P(buys computer = "yes") = 0.028$ 

P(X|buys\_computer = "no") \* P(buys\_computer = "no") = 0.007

Therefore, X belongs to class ("buys computer = yes")

- ตัวอย่างในการหาค่าต่าง ๆ