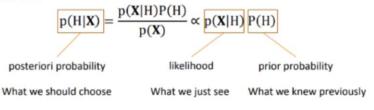
Bayes' Theorem: Basics

■ Total probability Theorem:

$$p(B) = \sum_i p\left(B|A_i\right) p(A_i)$$

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 <u>assumption</u> 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

Naïve Bayes Classifier: Categorical vs. Continuous Valued Features

☐ 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 μ and standard deviation σ

$$p(x_k = v_k | C_i) = N(x_k | \mu_{C_i}, \sigma_{C_i}) = \frac{1}{\sqrt{2\pi}\sigma_{C_i}} e^{-\frac{(x - \mu_{C_i})^2}{2\sigma^2}}$$

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

Naïve Bayes Classifier: An Example

- $P(C_i)$: $P(buys_computer = "yes") = 9/14 = 0.643$ $P(buys_computer = "no") = 5/14 = 0.357$
- □ Compute P(X|C_i) for each class

 $P(age = "<=30" | buys_computer = "yes") = 2/9 = 0.222$ $P(age = "<= 30" | buys_computer = "no") = 3/5 = 0.6$

P(income = "medium" | buys_computer = "yes") = 4/9 = 0.444 P(income = "medium" | buys_computer = "no") = 2/5 = 0.4

P(student = "yes" | buys_computer = "yes) = 6/9 = 0.667

P(student = "yes" | buys_computer = "no") = 1/5 = 0.2 P(credit_rating = "fair" | buys_computer = "yes") = 6/9 = 0.667

P(credit_rating = "fair" | buys_computer = "no") = 2/5 = 0.4

<=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

age income student credit_rating buys_computer

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")