Bayes' Theorem: Basics

ใช้หาค่าความน่าจะเป็นของเบย์

- Total probability Theorem:

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

- Bayes' Theorem:

Hypothesis

training data

$$\boxed{p(H|\mathbf{X})} = \frac{p(\mathbf{X}|H)P(H)}{p(\mathbf{X})} \propto \boxed{p(\mathbf{X}|H)}\,\boxed{P(H)}$$

P test data

A? B?

posteriori probability | likelihood | prior probability

What we should choose | What we just see | What we knew previously

- **X:** a data sample ("*evidence*")
- H: X belongs to class C

Prediction can be done based on Bayes' Theorem:

Classification is to derive the maximum posteriori

ความน่าจะเป็นคลาสไหนมากที่สุด

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.

Attribute แต่ละอันต้องไม่เกี่ยวข้องกัน

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

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

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

Naïve Bayes Classifier: Categorical vs. Continuous Valued Features

- If feature $\mathrm{x_k}$ is categorical, $\mathrm{p(x_k = v_k|C_i)}$ is the # of tuples in $\mathrm{C_i}$ with $\mathrm{x_k = v_k}$, divided by $|C_{i,\,D}|$ (# of tuples of C_i in D)

$$\mathrm{p}(\mathrm{X}|C_i) = \prod_{\mathrm{k}} \mathrm{p(x_k|C_i) = p(x_1|C_i) \cdot p(x_2|C_i) \cdots\cdots p(x_n|C_i)}$$

- If feature $\mathrm{x_k}$ is continuous-valued, $\mathrm{p(x_k = v_k|C_i)}$ is usually computed based on Gaussian distribution with a mean μ and standard deviation σ

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

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)

ฟีเจอร์

age	income	student	credit_rating	buys_computer
<=30	high	no	fair	no
<=30	high	no	excellent	no
31…40	high	no	fair	yes
>40	medium	no	fair	yes
>40	low	yes	fair	yes
>40	low	yes	excellent	no
31…40	low	yes	excellent	yes
<=30	medium	no	fair	no
<=30	low	yes	fair	yes
>40	medium	yes	fair	yes
<=30	medium	yes	excellent	yes
31…40	medium	no	excellent	yes
31…40	high	yes	fair	yes
>40	medium	no	excellent	no

5

Training Data

$P(H'^{y} \mid X') = ?$

$P(H'^{N} \mid X') = ?$

Training Data

$= P(x \mid H'^{y}) \, P(H')^{y} \rightarrow \frac{9}{14}$

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

age	income	student	credit_rating	buys_computer
<=30	high	no	fair	no
<=30	high	no	excellent	no
31…40	high	no	fair	yes
>40	medium	no	fair	yes
>40	low	yes	fair	yes
>40	low	yes	excellent	no
31…40	low	yes	excellent	yes
<=30	medium	no	fair	no
<=30	low	yes	fair	yes
>40	medium	yes	fair	yes
<=30	medium	yes	excellent	yes
31…40	medium	no	excellent	yes
31…40	high	yes	fair	yes
>40	medium	no	excellent	no

- **X = (age <= 30 , income = medium, student = yes, credit_rating = fair)**

$P(X|C_i)$: 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 x 0.4 x 0.2 x 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")

$\hat{x}$ = age = 42 , student = yes

$P(x \begin{smallmatrix} \nearrow y \\ \searrow N \end{smallmatrix} \mid \hat{x}) = ?$

$3/9 \rightarrow$ yes

$6/9$

$\frac{9}{14} \rightarrow \frac{\text{yes}}{\text{ทั้งหมด}}$

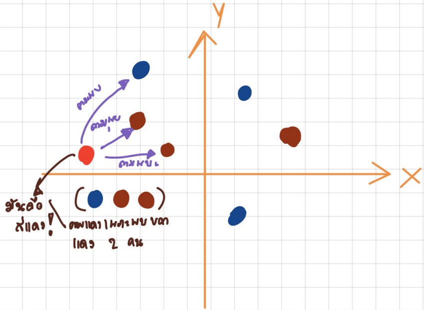
$P(H = y \mid (\text{age} = 42, \text{student} = \text{yes})) = P(\text{age} = 42 \mid y)\, P(\text{stu} = \text{yes} \mid y_{\text{by}})\, P(y)$

$= \frac{3}{9} \times \frac{6}{9} \times \frac{9}{14}$

$P(H_{\text{Buy}=N} \mid (\text{age} = 42, \text{stu} = \text{yes}) = P(\text{age} = 42 \mid N)\, P(\text{stu} = \text{yes} \mid y)\, P(N)$

K-NN

① เอาจุดมาใหม่ลงไปใน
ฟีเจอร์ space → เอาจุดลงไป
plot ในพิกัด



② ถามเพื่อนบ้าน
k คน

สมมติ $k = 3$