Chapter 8. Classification: Basic Concepts

- Classification: Basic Concepts
- Decision Tree Induction
- Bayes Classification Methods
- Linear Classifier
- Model Evaluation and Selection
- Techniques to Improve Classification Accuracy: Ensemble Methods
- Additional Concepts on Classification
- Summary

คาบที่ 19

What Is Bayesian Classification?

- A statistical classifier
  - Perform *probabilistic prediction* (*i.e.,* predict class membership probabilities)
- Foundation—Based on Bayes' Theorem
- Performance

มีคนเอาทฤษฎีมาทำเป็น classification คือ

  - A simple Bayesian classifier, *naïve Bayesian classifier*, has comparable performance with decision tree and selected neural network classifiers
- Incremental
  - Each training example can incrementally increase/decrease the probability that a hypothesis is correct—prior knowledge can be combined with observed data
- **Theoretical Standard**
  - Even when Bayesian methods are computationally intractable, they can provide a standard of optimal decision making against which other methods can be measured

Bayes' Theorem: Basics

- Total probability Theorem:

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

- Bayes' Theorem:

test data

A? B?

training data

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

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

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)

ต้องไม่เกี่ยวข้อง/สัมพันธ์กัน
ถึงจะคำนวณได้ง่าย

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

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

Naïve Bayes Classifier: Categorical vs. Continuous Valued Features

- If feature 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)

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

- 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

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

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)

X y

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

Training Data

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

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

training data

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

เรียกว่า

"prior probability"

มีโอกาสเป็น yes $= \frac{9}{14}$

ต่อมาจะคำนวณ yes โดยไม่สนใจ x

และ คำนวณ x โดยที่รู้ว่ามี yes อยู่ จะเป็นยังไง

$\hat{X}$ = age = 42 , student = yes ?

$$P(H^{\nearrow y}_{\searrow N} \mid \hat{X}) = ?$$

$$P(H = y_{buy} \mid (age = 42, student = yes)) = \overset{3/9}{\overset{\uparrow}{P(age = 42 \mid y_{buy})}} \; \overset{6/9}{\overset{\uparrow}{P(student = yes \mid y_{buy})}}$$

$$P(y_{buy}) \quad 9/14$$

$$P(H_{buy} = N \mid (age = 81, student = yes)) =$$

คาบที่ 19 บทที่ 9

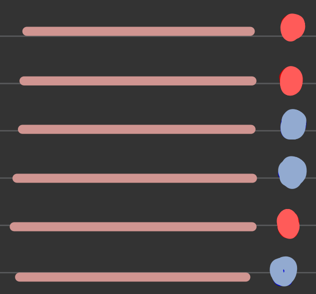
Lazy Learner: Instance-Based Methods

Lazy Learner คือ พอได้ข้อมูล train มา ก็จะเก็บไว้ พอได้ Data ใหม่มา ถึงจะทำ (หมักดอง งานไว้)

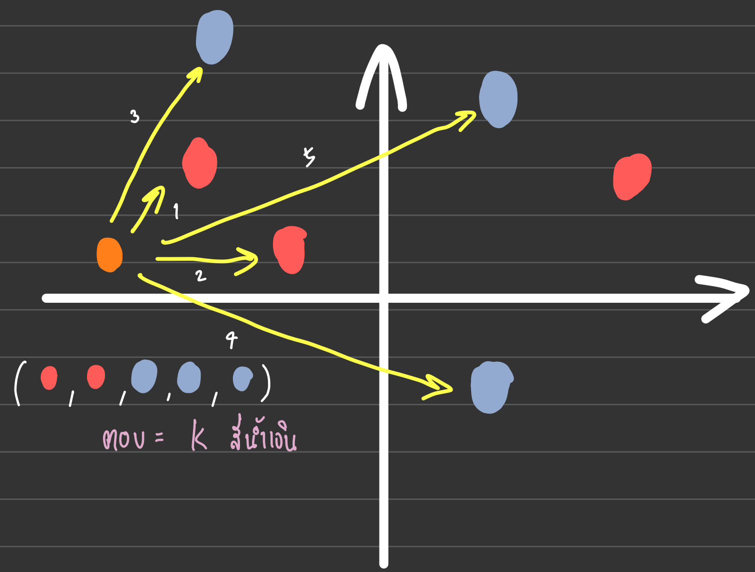
- Instance-based learning:
  - Store training examples and delay the processing ("lazy evaluation") until a new instance must be classified
- Typical approaches
  - <u>*k*-nearest neighbor approach</u>
    - Instances represented as points in a Euclidean space.
  - <u>Locally weighted regression</u>
    - Constructs local approximation
  - <u>Case-based reasoning</u>
    - Uses symbolic representations and knowledge-based inference

K-NN

K-Nearest Neighbors



test data



① plot ลงใน space

② หาเพื่อนบ้าน k คน

(k = 3)

ถ้าหาเพื่อนบ้านที่เป็นสีแดง แล้วก็จะตอบแดง

* k จะไม่เป็นเลขคู่ *

Naïve Bayes Classifier: An Example

เช่น ใน yes=9 คน (age) มีคนที่ช่วงอายุ <=30 กี่คน | ใน 9 คนมี medium ที่เป็น (income) yes กี่คน

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

- $P(C_i)$: P(buys_computer = “yes”) = $\boxed{9}$/14 = 0.643

  P(buys_computer = “no”) = 5/14= 0.357
- Compute $P(X|C_i)$ for each class ตอบ = 2 คน

  P(age = “<=30” | buys_computer = “yes”) = $\boxed{2}$/9 = 0.222

  P(age = “<= 30” | buys_computer = “no”) = 3/5 = 0.6 ตอบ = 4 คน

  P(income = “medium” | buys_computer = “yes”) = $\boxed{4}$/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
- **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 = $\boxed{0.044}$

P(X|buys_computer = “no”) = 0.6 x 0.4 x 0.2 x 0.4 = $\boxed{0.019}$

$P(X|C_i)*P(C_i)$: P(X|buys_computer = “yes”) * P(buys_computer = “yes”) = $\boxed{0.028}$

0.044 x 0.643

P(X|buys_computer = “no”) * P(buys_computer = “no”) = 0.007

Therefore, X belongs to class (“buys_computer = yes”)