14/10/64

What Is Bayesian Classification?

- A statistical classifier
  - Perform *probabilistic prediction* (*i.e.,* predict class membership probabilities)
- Foundation—Based on Bayes' Theorem
- Performance
  - ทุกๆ ฟีเจอร์ไม่มีความสัมพันธ์กัน
  - 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: เวลาจะสร้าง naive Bay ก็จะสร้างสมการนี้

test data

Class A เท่าไร

Class B เท่าไร

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

features

เท่ากันเสมอ, ละเว้นได้

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

ระบุฟีเจอร์ว่าเป็น class อะไร เท่ากับเท่าไร

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)

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

- 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

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

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

$P(H^{y}_{N}|\hat{X})$ = ?   42 yes 2

P(H = y | (age = 42, student = yes))

= P(age=42|popy) P(student|popy) P(popy)

3/9 x 6/9 x 9/14

= P(H_Buy = N | (age=41, studen=yes)

= 0.142857

Naïve Bayes Classifier: Training Dataset

Class:

C1:buys_computer = 'yes'

C2:buys_computer = 'no'

Classified ว่าคนที่มีลักษณะแบบนี้จะซื้อหรือไม่

Data to be classified: คำถาม

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

P(H|X) = ?

P(H|X) = ?

= P(X|H^y) P(H^y) → (9/14)

Training Data ← → มีโอกาสเป็น yes เท่าไร

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

yes = 9

No = 5

Taining Data

นาอีฟ เบย์

Naïve Bayes Classifier: An Example

→ ทุกๆ ฟีเจอร์ไม่มีความสัมพันธ์กัน

- $P(C_i)$: 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)**

เอาตัวเลขที่หาได้มาคูณกัน

$P(X|C_i)$: 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   (0.044, 0.643)

P(X|buys_computer = "no") * P(buys_computer = "no") = 0.007

Therefore, X belongs to class ("buys_computer = yes")

Chapter 9

Lazy Learner: Instance-Based Methods

เลซี่ เลินเนอร์

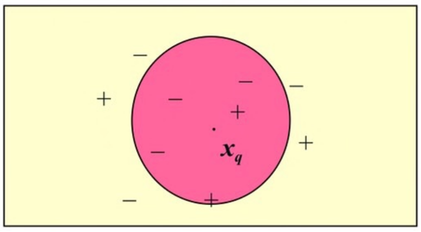
↳ เมื่อได้ข้อมูล train มาเก็บเอาไว้เฉยๆ เมื่อมี data เข้ามาถึงจะทำ | เก็บไว้จนสุดท้ายจริงๆ ค่อยทำ

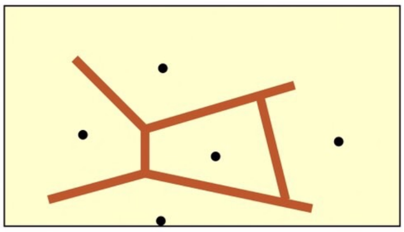
- Instance-based learning:
  - Store training examples and delay the processing (“lazy evaluation”) until a new instance must be classified
- Typical approaches
  - <u>*k*-nearest neighbor approach</u>
    - Instances represented as points in a Euclidean space.
  - <u>Locally weighted regression</u>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

65

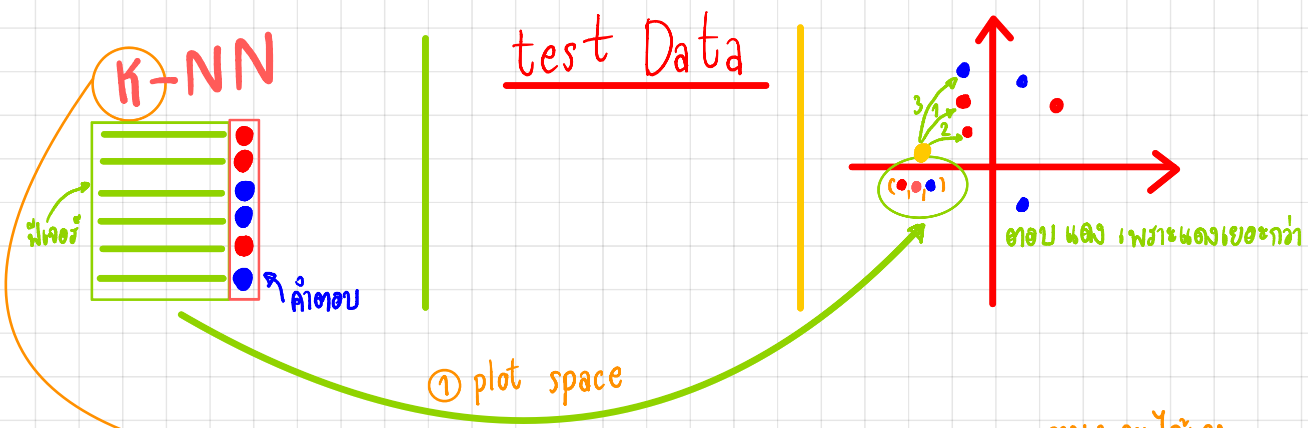
The *k*-Nearest Neighbor Algorithm

- All instances correspond to points in the n-D space
- The nearest neighbor are defined in terms of Euclidean distance, dist($\mathbf{X}_1$, $\mathbf{X}_2$)
- Target function could be discrete- or real- valued
- For discrete-valued, *k*-NN returns the most common value among the *k* training examples nearest to x_q
- Vonoroi diagram: the decision surface induced by 1-NN for a typical set of training examples





66



② ถามเพื่อนบ้าน
k คน
คนที่ 1 แดง
คนที่ 2 แดง
คนที่ 3 น้ำเงิน

ถาม 3 คน ได้แดง
5 คนได้ น้ำเงิน
เซนซิทีฟกับค่า K มากๆ
เชื่อเพื่อนบ้านที่อยู่ใกล้ก่อน