Bayes' Theorem: Basics

- Total probability Theorem:

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

- Bayes' Theorem:

p test data ว่า A? ว่า B?

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

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

ฟีเจอร์ ว่าเป็นคลาสอะไร เท่ากับเท่าไหร่

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 $\boxed{\text{assumption}}$ to simplify the model, but achieve comparable performance.

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 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)

classified ว่าคนที่มีลักษณะแบบนี้ จะซื้อคอมพิวเตอร์หรือไม่

pop

$p(H|X) = ?$

$p(H|X) = ?$

training data

$p(x|H)\,P(H)$ $\frac{9}{14}$

สนใจแค่ yes ไม่สนใจ x

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

yes = 9

no = 5

ทุกๆ ฟีเจอร์ ไม่มีความสัมพันธ์กัน

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 ↝ สนใจแค่คนที่ซื้อ
  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
(0.044) (0.643)
P(X|buys_computer = "no") * P(buys_computer = "no") = 0.007

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

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

$P(\overset{Y}{\underset{N}{H}} \mid \hat{x}) = ?\quad 42\ y\ es\ 2$

$P(H = y) \mid (age = 42, student = yes)) = P(age = 42 \mid \overset{3/9}{popy}) \times P(student \mid \overset{6/9}{popy}) \times P(\overset{9/14}{popy})$

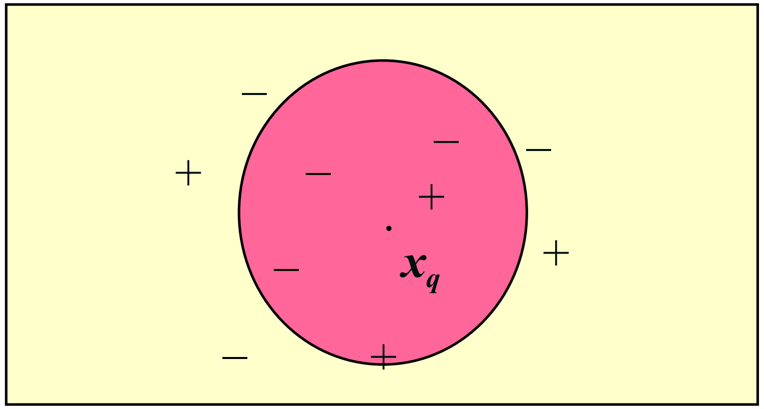
Lazy Learner: Instance-Based Methods

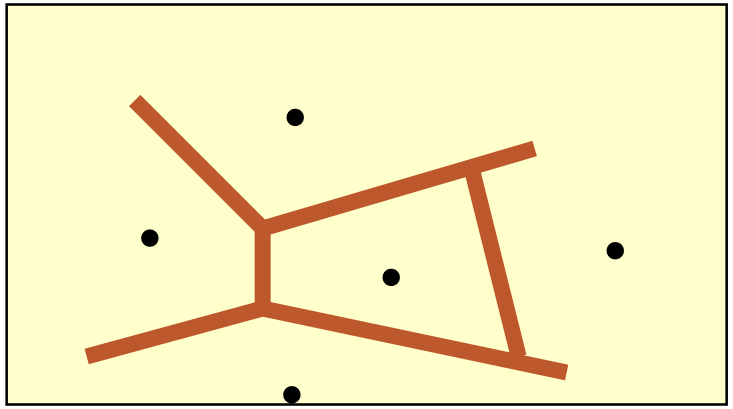
เมื่อได้ข้อมูล train มาเก็บไว้เฉยๆ เมื่อมี data เข้ามาถึงจะทำ | เก็บไว้จนสุดท้ายจริงๆค่อยทำ

- 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

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





K-NN

ฟีเจอร์

คำตอบ

test Data

ตอบแดง
เพราะแดง
เยอะกว่า

(, ,)

① plot space

ถาม 3 คนได้แดง
5 คน ได้ น้ำเงิน
เซนซิทีฟกับค่า k มากๆๆ
เชื่อเพื่อนบ้านที่อยู่ใกล้ก่อน

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