Bayes' Theorem: Basics

- Total probability Theorem: $P(B) = \sum_{i=1}^{M} P(B|A_i)P(A_i)$
- Bayes' Theorem: $\boxed{P(H|\mathbf{X}) = \frac{P(\mathbf{X}|H)P(H)}{P(\mathbf{X})}} = P(\mathbf{X}|H) \times P(H)/P(\mathbf{X})$
  - Let **X** be a data sample ("*evidence*"): class label is unknown
  - Let H be a *hypothesis* that X belongs to class C
  - Classification is to determine P(H|**X**), (i.e., *posteriori probability*): the probability that the hypothesis holds given the observed data sample **X**
  - P(H) (*prior probability*): the initial probability
    - E.g., **X** will buy computer, regardless of age, income, …
  - P(**X**): probability that sample data is observed
  - P(**X**|H) (likelihood): the probability of observing the sample **X**, given that the hypothesis holds
    - E.g., Given that **X** will buy computer, the prob. that X is 31..40, medium income

Prediction Based on Bayes' Theorem

- Given training data **X**, *posteriori probability of a hypothesis* H, P(H|**X**), follows the Bayes' theorem

$$P(H|\mathbf{X}) = \frac{P(\mathbf{X}|H)P(H)}{P(\mathbf{X})} = P(\mathbf{X}|H) \times P(H)/P(\mathbf{X})$$

- Informally, this can be viewed as

  posteriori = likelihood x prior/evidence

- Predicts **X** belongs to C_i iff the probability P(C_i|**X**) is the highest among all the P(C_k|X) for all the *k* classes
- Practical difficulty: It requires initial knowledge of many probabilities, involving significant computational cost

classification--naive bayes.mp4

Classification Is to Derive the Maximum Posteriori

- Let D be a training set of tuples and their associated class labels, and each tuple is represented by an n-D attribute vector $\mathbf{X} = (x_1, x_2, \ldots, x_n)$
- Suppose there are *m* classes $C_1, C_2, \ldots, C_m$.
- Classification is to derive the maximum posteriori, i.e., the maximal $P(C_i|\mathbf{X})$
- This can be derived from Bayes' theorem

$$P(C_i|\mathbf{X}) = \frac{P(\mathbf{X}|C_i)P(C_i)}{P(\mathbf{X})}$$

- Since P(X) is constant for all classes, only

$$P(C_i|\mathbf{X}) = P(\mathbf{X}|C_i)P(C_i)$$

needs to be maximized

11:23 / 1:14:06

Naïve Bayes Classifier

- A simplified assumption: attributes are conditionally independent (i.e., no dependence relation between attributes):

$$P(\mathbf{X}|C_i) = \prod_{k=1}^{n} P(x_k|C_i) = P(x_1|C_i) \times P(x_2|C_i) \times \ldots \times P(x_n|C_i)$$

- This greatly reduces the computation cost: Only counts the class distribution
- If A_k is categorical, $P(x_k|C_i)$ is the # of tuples in C_i having value x_k for A_k divided by $|C_{i,D}|$ (# of tuples of C_i in D)
- If A_k is continous-valued, $P(x_k|C_i)$ is usually computed based on Gaussian distribution with a mean μ and standard deviation σ

$$g(x, \mu, \sigma) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

and $P(x_k|C_i)$ is

$$P(\mathbf{X}|C_i) = g(x_k, \mu_{C_i}, \sigma_{C_i})$$

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	com
<=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

Naïve Bayes Classifier: An Example

- We need to compute:

P(buys_comp="yes"| age <=30, Inc = med, Stud = yes, CR=Fair) = ?
P(buys_comp="no"| age <=30, Inc = med, Stud = yes, CR=Fair) = ?

- According to Naïve Bayes assumption these will be calculated as:

P(buys_comp=yes| age <=30, Inc = med, Stud = yes, CR=Fair) =
P(age=<=30|buys_comp=yes)* P(inc=med|buys_comp=yes)*
P(stud=yes|buys_comp=yes)* P(cr=fair|buys_comp=yes)*
P(buys_com =yes)

P(buys_comp=no| age <=30, Inc = med, Stud = yes, CR=Fair) =
P(age=<=30|buys_comp=no)* P(inc=med|buys_comp=no)*
P(stud=yes|buys_comp=no)*P(cr=fair|buys_comp=no)*
P(buys_com =no)

1 P(buys_comp=yes| age <=30, Inc = med, Stud = yes, CR=Fair) =

P(age=<=30|buys_comp=yes)* P(inc=med|buys_comp=yes)*
P(stud=yes|buys_comp=yes)* P(cr=fair|buys_comp=yes)*
P(buys_com =yes)

P(buys_computer = “yes”) = 9/14 = 0.643
P(age =<=30| buys_comp=yes) = 2/9 = 0.222
P(inc=med|buys_comp=yes) = 4/9 = 0.444
P(stud=yes|buys_comp=yes) = 6/9 = 0.667
P(CR=fair|buys_comp=yes) = 6/9 = 0.667

age	income	studen	credit_rating	com
<=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 , inc = med, student = yes, CR = fair)

$P(X|C_i)$:

1 P(X|buys_comp = “yes”) = 0.222 x 0.444 x 0.667 x 0.667 = 0.044
2 P(X|buys_comp = “no”) = 0.6 x 0.4 x 0.2 x 0.4 = 0.019

$P(X|C_i)*P(C_i)$:

1 P(X|buys_comp= “yes”) * P(buys_comp = “yes”) = 0.028
2 P(X|buys_comp = “no”) * P(buys_comp = “no”) = 0.007

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

Q:

- Classify tuple:

X = (age =>40,
Income = low,
Student = yes
CR = excellent)

- Find:

P(b_c=yes|X)=?
P(b_c=no|X)=?

age	income	student	credit_rating	com
<=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=>40,Inc=low,Stu=y,CR=ex)

① **P(b_c=yes|X)**=P(age=>40|b_c=yes)*
P(inc=low|b_c=yes)*P(stu=y|b_c=yes)*
P(CD=ex|b_c=yes)*P(b_c=yes)

P(b_c=yes|X)=3/9*3/9*6/9*3/9*9/14

② P(b_c=no|X)= P(age=>40|b_c=no)*
P(inc=low|b_c=no)*P(stu=y|b_c=no)*
P(CD=ex|b_c=no)*P(b_c=no)

P(b_c=no|X)=2/5*1/5*4/5*3/5*5/14

age	income	student	credit_rating	com
<=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

Avoiding the Zero-Probability Problem

- Naïve Bayesian prediction requires each conditional prob. be **non-zero**. Otherwise, the predicted prob. will be zero

$$P(X \mid C_i) = \prod_{k=1}^{n} P(x_k \mid C_i)$$

- Ex. Suppose a dataset with 1000 tuples, income=low (0), income= medium (990), and income = high (10)
- Use **Laplacian correction** (or Laplacian estimator)
  - *Adding 1 to each case*

    +3

    Prob(income = low) = 1/1003

    Prob(income = medium) = 991/1003

    Prob(income = high) = 11/1003
  - The "corrected" prob. estimates are close to their "uncorrected" counterparts

X=(age=31-40,Inc=low,Stu=y,CR=ex)

1) **P(b_c=yes|X)**=P(age=31-40|b_c=yes)*
P(inc=low|b_c=yes)*P(stu=y|b_c=yes)*
P(CD=ex|b_c=yes)*P(b_c=yes)

P(b_c=yes|X)=4/9*3/9*6/9*3/9*9/14

2) P(b_c=no|X)= P(age=31-40|b_c=no)*
P(inc=low|b_c=no)*P(stu=y|b_c=no)*
P(CD=ex|b_c=no)*P(b_c=no)

P(b_c=no|X)=0/5*1/5*4/5*3/5*5/14

Thus,

P(b_c=no|X)=1/8*1/5*4/5*3/5*5/14

age	income	student	credit_rating	com
<=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

Classification: Mammals vs. Non-mammals

Name	Give Birth	Can Fly	Live in Water	Have Legs	Class
uman	yes	no	no	yes	mammals
ython	no	no	no	no	non-mammals
almon	no	no	yes	no	non-mammals
hale	yes	no	yes	no	mammals
rog	no	no	sometimes	yes	non-mammals
omodo	no	no	no	yes	non-mammals
at	yes	yes	no	yes	mammals
igeon	no	yes	no	yes	non-mammals
at	yes	no	no	yes	mammals
eopard shark	yes	no	yes	no	non-mammals
urtle	no	no	sometimes	yes	non-mammals
enguin	no	no	sometimes	yes	non-mammals
orcupine	yes	no	no	yes	mammals
el	no	no	yes	no	non-mammals
alamander	no	no	sometimes	yes	non-mammals
ila monster	no	no	no	yes	non-mammals
latypus	no	no	no	yes	mammals
wl	no	yes	no	yes	non-mammals
olphin	yes	no	yes	no	mammals
agle	no	yes	no	yes	non-mammals

- Train the model (learn the parameters) using the given data set.
- Apply the learned model on new cases.

20

Predict class for (GiveBirth=yes, CanFly=no, LivesinWater=yes, HaveLegs=no)

Naïve Bayes Classifier: Comments

- Advantages
  - Easy to implement
  - Good results obtained in most of the cases
- Disadvantages
  - Assumption: class conditional independence, therefore loss of accuracy
  - Practically, dependencies exist among variables
    - E.g., hospitals: patients: Profile: age, family history, etc.
      Symptoms: fever, cough etc., Disease: lung cancer, diabetes, etc.
    - Dependencies among these cannot be modeled by Naïve Bayes Classifier
- How to deal with these dependencies? Bayesian Belief Networks