Bayesian Classification

Bayesian Classification

- **A statistical classifier**: performs *probabilistic prediction, i.e.,* predicts class membership probabilities
- **Foundation:** Based on Bayes' Theorem.
- **Performance:** 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
- **Standard**: Even when Bayesian methods are computationally intractable, they can provide a standard of optimal decision making against which other methods can be measured

Basic Formulas for Probabilities

Product Rule : probability P(AB) of a conjunction of two events A and B:

$$P(A,B) = P(A \mid B)P(B) = P(B \mid A)P(A)$$

Sum Rule: probability of a disjunction of two events A and B:

$$P(A+B) = P(A) + P(B) - P(AB)$$

Theorem of Total Probability : if events A_1, …., A_n are mutually exclusive with

$$P(B) = \sum_{i=1}^{n} P(B \mid A_i)P(A_i)$$

Basic Approach

Bayes Rule:

$$P(h \mid X) = \frac{P(X \mid h)P(h)}{P(X)}$$

- $P(h)$ = prior probability of hypothesis h
- $P(X)$ = prior probability of training data X
- $P(h \mid X)$ = probability of h given X (posterior density)
- $P(X \mid h)$ = probability of X given h (likelihood of X given h)

The Goal of Bayesian Learning: the most probable hypothesis given the training data (Maximum a Posteriori (MAP) hypothesis h_{map})

$$\begin{aligned} h_{map} &= \max_{h \in H} P(h \mid X) \\ &= \max_{h \in H} \frac{P(X \mid h)P(X)}{P(X)} \\ &= \max_{h \in H} P(X \mid h)P(h) \end{aligned}$$

Towards Naïve Bayesian Classifier

- Let D be a training set of tuples and their associated class labels, and each tuple is represented by an *n*-D attribute vector **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

needs to be maximized

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

Naïve Bayesian Classifier: Training Dataset

Class:

C_1:buys_computer = 'yes'

C_2:buys_computer = 'no'

Data sample

X = (age <=30,

Income = medium,

Student = yes

Credit_rating = Fair)

age	income	student	credit_rating	_comp
<=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 Bayesian Classifier: Training Dataset

Class:

C_1:buys_computer = 'yes'

C_2:buys_computer = 'no'

Data sample

X = (age <=30,

Income = medium,

Student = yes

Credit_rating = Fair)

age	income	student	credit_rating	comp
<=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 Bayesian 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

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

P(X|buys_computer = “yes”) = 0.222 x 0.444 x 0.667 x 0.667 x 0.643

= 0.028

P(X|buys_computer = “no”) = 0.6 x 0.4 x 0.2 x 0.4 x 0.357

= 0.007

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