Data Mining

Classification III – Naïve Bayes (Part A)

Dr. Jason T.L. Wang, Professor

Department of Computer Science

New Jersey Institute of Technology

Bayes Theorem

- Conditional probability

$$P(A,B) = P(A|B)P(B) = P(B|A)P(A)$$

$$\Rightarrow P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

Bayes rule

Naïve Bayes Classifier

- Applying Bayes rule

$$P(Yi|X1\ldots Xn) = \frac{P(Yi)\ P(X1\ldots Xn|Yi)}{P(X1\ldots Xn)}$$

$$= \frac{P(Yi)\ P(X1|Yi)P(X2|Yi)\ldots P(Xn|Yi)}{P(X1\ldots Xn)}$$

$$Y \leftarrow \arg\max P(Yi)\ P(X1|Yi)P(X2|Yi)\ldots P(Xn|Yi)$$

Assumption: Xi's are conditionally independent given Y

Terminology

P(Y|**X**): posterior probability for Y

P(Y): prior probability

P(**X**|Y): class-conditional probability

P(**X**): evidence

Bayes theorem (Bayes rule) allows us to calculate the posterior probability P(Y|**X**) using the prior probability P(Y), the class-conditional probability P(**X**|Y) and the evidence P(**X**) (which is constant and ignored).

Algorithmic Details (Training Phase)

- Naïve Bayes Algorithm

  - Learning/training phase: given a training data set T,

    For each label Yi

    P(Yi) ← calculate P(Y=Yi) using training examples in T;

    For each attribute value Xj

    calculate P(Xj|Yi) using training examples in T.

Training Data Set

Student	Assignment	Project	Exam	Label
1	Good	A	High	Pass
2	Good	B	High	Pass
3	Bad	B	Low	Fail
4	Bad	C	High	Fail
5	Good	C	Low	Fail
6	Good	C	High	Pass
7	Bad	B	High	Pass
8	Good	A	Low	Pass
9	Bad	A	Low	Fail
10	Good	B	Low	Pass

Training Phase

P(Pass) = 6/10; P(Fail) = 4/10

P(Good|Pass) = 5/6; P(Bad|Pass) = 1/6

P(High|Fail) = 1/4; P(Low|Fail) = 3/4

…

Algorithmic Details (Testing Phase)

- Naïve Bayes Algorithm
  - Testing phase: given an unlabeled test record $X^*=<X1...Xn>$
  - Assign label Y to X* based on $\max\{P(Yi)P(X1|Yi)P(X2|Yi)\ldots P(Xn|Yi)\}$

End of
Naïve Bayes Module (Part A)