Classification and Prediction

——Bayesian Classification——

徐华
清华大学 计算机系 智能技术与系统国家重点实验室
xuhua@tsinghua.edu.cn

1

Classification and Prediction

- Basic Concepts
- Issues Regarding Classification and Prediction
- Decision Tree
- Bayesian Classification
- Neural Networks
- Support Vector Machine
- K-Nearest Neighbor
- Associative classification
- Classification Accuracy

2

Bayesian Theorem: Basics

- **Let X be a data sample whose class label is unknown**
- **Let H be a hypothesis that X belongs to class C**
- **For classification problems, determine P(H|X): the probability that the hypothesis holds given the observed data sample X**
- **P(H): prior probability of hypothesis H (i.e. the initial probability before we observe any data, reflects the background knowledge)**
- **P(X): probability that sample data is observed**
- **P(X|H) : probability of observing the sample X, given that the hypothesis holds**

Bayesian Theorem

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

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

- **Informally, this can be written as**
  posteriori = likelihood x prior / evidence
- **MAP (maximum posteriori) hypothesis**

$$h_{MAP} \equiv \underset{h \in H}{\operatorname{argmax}} P(h|D) = \underset{h \in H}{\operatorname{argmax}} P(D|h)P(h).$$

- **Practical difficulty: require initial knowledge of many probabilities, significant computational cost**

Naïve Bayes Classifier

- A simplified assumption: attributes are conditionally independent:

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

- The product of occurrence of 2 elements y_1 and y_2, given the current class is C, is the product of the probabilities of each element taken separately, given the same class $P([y_1,y_2], C) = P(y_1, C) * P(y_2, C)$
- No dependence relation between attributes
- Greatly reduces the computation cost, only count the class distribution.
- Once the probability $P(X|C_i)$ is known, assign X to the class with maximum $P(X|C_i) * P(C_i)$

Training dataset

- Class: C1: buys_computer= 'yes' ; C2: buys_computer= 'no'
- Data sample:
  - X =(age<=30, Income=medium, Student=yes, Credit_rating= Fair)

age	income	student	credit_rating	buys_computer
<=30	high	no	fair	no
<=30	high	no	excellent	no
30…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

- **Compute P(X|Ci) 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|Ci) : 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|Ci)*P(Ci) : P(X|buys_computer= "yes") * P(buys_computer= "yes")=0.028
P(X|buys_computer= "no") * P(buys_computer= "no")=0.007

Therefore, X belongs to class "buys_computer=yes"

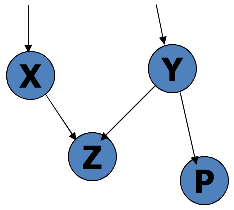
Naïve Bayesian Classifier: Comments

- **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 Bayesian Classifier
- **How to deal with these dependencies?**
  - Bayesian Belief Networks

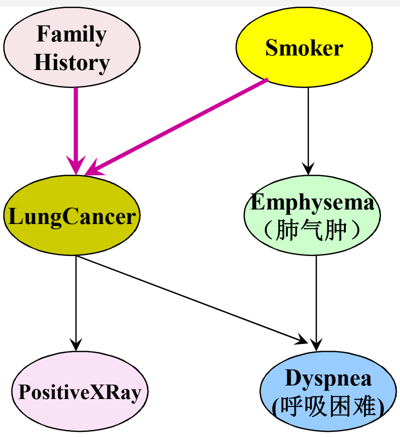
Bayesian Belief Networks

- Bayesian belief network allows a *subset* of the variables conditionally independent
- A graphical model of causal relationships
  - Represents dependency among the variables
  - Gives a specification of joint probability distribution



❑Nodes: random variables
❑Links: dependency
❑X,Y are the parents of Z, and Y is the parent of P
❑No dependency between Z and P
❑Has no loops or cycles

Bayesian Belief Network: An Example



Bayesian Belief Networks

	(FH, S)	(FH, ~S)	(~FH, S)	(~FH, ~S)
LC	0.8	0.5	0.7	0.1
~LC	0.2	0.5	0.3	0.9

The conditional probability table for the variable LungCancer:
Shows the conditional probability for each possible combination of its parents

$$P(z1,\dots,zn) = \prod_{i=1}^{n} P(z_i \mid Parents(Z_i))$$

Learning Bayesian Networks

- Several cases
  - Given both the network structure and all variables observable: learn only the CPTs
  - Network structure known, some hidden variables: method of gradient descent, analogous to neural network learning
  - Network structure unknown, all variables observable: search through the model space to reconstruct graph topology
  - Unknown structure, all hidden variables: no good algorithms known for this purpose
- D. Heckerman, Bayesian networks for data mining

  http://research.microsoft.com/adapt/MSBNx/

11

Thanks !

12