

Data Mining

Classification II – Other Methods (Part A)

**Dr. Jason T.L. Wang, Professor
Department of Computer Science
New Jersey Institute of Technology**

Generalization to N Classes where $N > 2$

- Consider class i and non-class i for each i , $i = 1, 2, \dots$
- Apply ID3 algorithm to class i and non-class i .
- Use the resulting N trees (classification rules) to classify an unlabeled object.

Other Classification Methods

- K-nearest neighbor classifier
 - $K = 1$
 - $K > 1$
- Neural networks

Accuracy of Classification

- Binary partition
 - use two independent data sets, e.g., training set (2/3), test set (1/3).
- Cross-validation
 - divide the data set into k partitions and use $k-1$ partitions as training data and one partition as test data – k -fold cross-validation ($k = 10$ is most common).

Research Issues

- Data cleaning, data reduction (relevance analysis, feature selection, dimensionality reduction), classification accuracy, speed, scalability, usability and quality of classification rules, robustness (ability to handle noise and incomplete data)



End of Other Classification Methods Module (Part A)