Associative Classifiers
Associative Classification

• Mine association possible rules (PR) in form of condset $\rightarrow c$
  – Condset: a set of attribute-value pairs
  – C: class label

• Build classifier
  – Organize rules according to decreasing precedence based on confidence and support

• Classification
  – Use the first matching rule to classify an unknown case
Associative Classification Methods

- CBA (Classification By Association: Liu, Hsu & Ma, KDD’98)
  - Mine association possible rules in the form of
    - Cond-set (a set of attribute-value pairs) → class label
  - Build classifier: Organize rules according to decreasing precedence based on confidence and then support

- CMAR (Classification based on Multiple Association Rules: Li, Han, Pei, ICDM’01)
  - Classification: Statistical analysis on multiple rules
More Methods

- CPAR (Classification based on Predictive Association Rules: Yin & Han, SDM’03)
  - Generation of predictive rules (FOIL-like analysis)
  - High efficiency, accuracy similar to CMAR
- RCBT (Mining top-k covering rule groups for gene expression data, Cong et al. SIGMOD’05)
  - Explore high-dimensional classification, using top-k rule groups
  - Achieve high classification accuracy and high runtime efficiency
CMAR – Model Generation

- Classification based on Multiple Association Rules
- Efficiency: Uses an enhanced FP-tree that maintains the distribution of class labels among tuples satisfying each frequent itemset
- Rule pruning whenever a rule is inserted into the tree
  - Given two rules, R1 and R2, if the antecedent of R1 is more general than that of R2 and \( \text{conf}(R1) \geq \text{conf}(R2) \), then R2 is pruned
  - Prune rules where the rule antecedent and class are not positively correlated, based on a \( \chi^2 \) test of statistical significance
CMAR – Classification

- Classification based on generated/pruned rules
- If only one rule satisfies tuple X, assign the class label of the rule
- If a rule set S satisfies X, CMAR
  - Divide S into groups according to class labels
  - Use a weighted $\chi^2$ measure to find the strongest group of rules, based on the statistical correlation of rules within a group
  - Assign X the class label of the strongest group
Classification by Aggregating Emerging Patterns

• Emerging pattern (EP): A pattern frequent in one class of data but infrequent in others
  – Age<=30 is frequent in class “buys_computer=yes” and infrequent in class “buys_computer=no”
  – Rule: age<=30 $\rightarrow$ buys computer

How to Mine Emerging Patterns?

• Border differential
  – Max-patterns in D1 w.r.t. min_sup=90%
  – Max-patterns in D2 w.r.t. min_sup=10%
  – $X$ is a pattern covered by a max-pattern in D1 but not by a max-pattern in D2 $\Rightarrow X$ is an emerging pattern

• Method
  – Mine max-patterns in D1 and D2, respectively
  – Compare the two sets of borders, find the “maximal” patterns that are frequent in D1 and infrequent D2
Instance-based Methods

• Instance-based learning
  – Store training examples and delay the processing until a new instance must be classified (“lazy evaluation”)

• Typical approaches
  – K-nearest neighbor approach
    • Instances represented as points in an Euclidean space
  – Locally weighted regression
    • Construct local approximation
  – Case-based reasoning
    • Use symbolic representations and knowledge-based inference
The K-Nearest Neighbor Method

- Instances are points in an n-D space
- The k-nearest neighbors (KNN) in the Euclidean distance
  - Return the most common value among the k training examples nearest to the query point
- Discrete-/real-valued target functions
KNN Methods

- For continuous-valued target functions, return the mean value of the k nearest neighbors
- Distance-weighted nearest neighbor algorithm
  - Give greater weights to closer neighbors
  \[ w = \frac{1}{d(x_q, x_i)^2} \]
- Robust to noisy data by averaging k-nearest neighbors
- Curse of dimensionality
  - Distance could be dominated by irrelevant attributes
  - Axes stretch or elimination of the least relevant attributes
Case-based Reasoning

- Lazy evaluation + analysis of similar instances
- Methodology
  - Instances represented by rich symbolic descriptions (e.g., function graphs)
  - Combine multiple retrieved cases
  - Tight coupling between case retrieval, knowledge-based reasoning, and problem solving
Lazy vs. Eager Learning

- Efficiency: lazy learning uses less training time but more predicting time
- Accuracy
  - Lazy method effectively uses a richer hypothesis space
  - Eager: must commit to a single hypothesis that covers the entire instance space
Some Other Methods

- Genetic algorithms
- Rough set approaches
- Fuzzy set approaches
What Is Prediction?

• Prediction is similar to classification
  – Construct a model
  – Use the model to predict unknown value
    • Linear and multiple regression, non-linear regression

• Prediction models continuous-valued functions
Linear Regression

- Linear regression
  - $Y = \alpha + \beta X$

- Multiple regression
  - $Y = b_0 + b_1 X_1 + b_2 X_2 + \ldots + b_n X_n$
  - Many nonlinear functions can be transformed
Summary

- Association rules can be used for classification
- Emerging patterns for classification
- Lazy learning methods
- Prediction methods
To-Do-List

• G. Dong and J. Li. Efficient mining of emerging patterns: Discovering trends and differences. In KDD'99

• Understand how emerging patterns can be mined efficiently