Discretization

Andrew Kusiak
Intelligent Systems Laboratory
2139 Seamans Center
The University of Iowa
Iowa City, IA 52242 - 1527
andrew-kusiak@uiowa.edu
http://www.icaen.uiowa.edu/~ankusiak
Tel. 319-335 5934

Discretization: Basic Concepts

Continuous values $\rightarrow$ Discrete

Discretization Algorithms

- Unsupervised vs supervised. Unsupervised algorithms do not consider the decision value.
- Global vs local. Global algorithms group values of each feature into intervals by considering other features. Local algorithms group locally.
- Static vs dynamic. Static algorithms discretize each feature in one iteration independent of other features. Dynamic algorithms search for all possible intervals for all features simultaneously.

Discretization: Main Steps

- Determine the number of discrete intervals: Normally done by a user.
- Determine the width of each interval; Normally done by the discretization algorithm.
Discretization Algorithms

- Equal interval width discretization
- Equal frequency discretization
- $k$-means clustering discretization
- One-level (1RD) decision tree discretization
- Information-theoretic discretization methods:
  - $\chi^2$ method
  - maximum entropy discretization
  - class-attribute interdependence redundancy discretization (CAIR)
  - class-attribute interdependence uncertainty and redundancy discretization (CAIUR)