Data Mining

Clustering II – Agglomerative Hierarchical Algorithm (Part B)

Dr. Jason T.L. Wang, Professor

Department of Computer Science
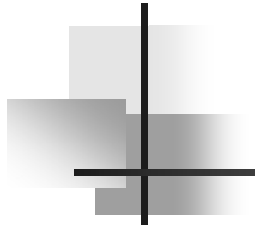
New Jersey Institute of Technology

/

Outlier Discovery

- What are outliers?
  - Objects that are considerably dissimilar from the remainder of the data
- Applications:
  - Credit card fraud detection
  - Telecom fraud detection
  - A newly formed cancer protein

Outlier Discovery Techniques

- Distance-based outlier: A DB(p, D)-outlier is an object O in a dataset T such that at least a fraction p of the objects in T lie at a distance greater than D from O.

- Algorithms for mining distance-based outliers
  - Index-based algorithm
  - Nested-loop algorithm

End of
Agglomerative Hierarchical Clustering Algorithm Module (Part B)