- A **grid-based method** first quantizes the object space into a finite number of cells that form a grid structure, and then performs clustering on the grid structure. STING is a typical example of a grid-based method based on statistical information stored in grid cells. CLIQUE and WaveCluster are two clustering algorithms that are both grid-based and density-based.
- A **model-based method** hypothesizes a model for each of the clusters and finds the best fit of the data to that model. Typical model-based methods involve statistical approaches (such as COBWEB, CLASSIT, and AutoClass) or neural network approaches (such as competitive learning and self-organizing feature maps).
- *One person's noise could be another person's signal.* **Outlier detection and analysis** are very useful for fraud detection, customized marketing, medical analysis, and many other tasks. Computer-based outlier analysis methods typically follow either a *statistical approach*, a *distance-based approach*, or a *deviation-based approach*.

Exercises

8.1 Briefly outline how to compute the dissimilarity between objects described by the following types of variables:

(a) Asymmetric binary variables
(b) Nominal variables
(c) Ratio-scaled variables
(d) Numerical (interval-scaled) variables

8.2 Given the following measurements for the variable *age*:

18, 22, 25, 42, 28, 43, 33, 35, 56, 28,

standardize the variable by the following:

(a) Compute the mean absolute deviation of *age*.
(b) Compute the z-score for the first four measurements.

8.3 Given two objects represented by the tuples (22, 1, 42, 10) and (20, 0, 36, 8):

(a) Compute the *Euclidean distance* between the two objects.
(b) Compute the *Manhattan distance* between the two objects.
(c) Compute the *Minkowski distance* between the two objects, using $q = 3$.