

Example of Exam Questions

January 30, 2012

1. To detect fraudulent usage of credit cards, the following data mining task should be used:
 - feature selection
 - prediction
 - outlier analysis
2. Correlation analysis is used to:
 - extract association rules
 - define support and confidence values
 - eliminate misleading rules
3. In a scenario of clustering text documents using frequent patterns, you aim at choosing the subset of frequent term sets that minimizes the overlap between clusters. To do so, you compute the information entropy based on:
 - the number of documents covered by the subset of the frequent term sets
 - the number of all frequent term sets for each document
 - the number of documents covered by each cluster
4. Which of the following methods is not designed to handle outlier detection in multi-dimensional spaces
 - distance-based methods
 - statistical-based methods
 - density-based methods
5. To detect clusters in a dataset having different densities, you should use
 - global density parameters
 - a small ϵ -neighborhood
 - a large ϵ -neighborhood

- an ordering of the objects based on their core and reachability distance
- an ordering of the objects based on the number of objects falling in their ϵ -neighborhood