

Data Mining, Sheet 1

Exercise 1

What is data mining? In your answer, address the following:

1. Is it another hype?
2. Is it a simple transformation or application of technology developed from databases, statistics, machine learning, and pattern recognition?
3. We have presented a view that data mining is the result of the evolution of database technology. Do you think that data mining is also the result of the evolution of machine learning research? Can you present such views based on the historical progress of this discipline? Address the same for the fields of statistics and pattern recognition.
4. Describe the steps involved in data mining when viewed as a process of knowledge discovery.

Exercise 2

Define each of the following data mining functionalities: characterization, discrimination, association and correlation analysis, classification, regression, clustering, and outlier analysis. Give examples of each data mining functionality, using a real-life database that you are familiar with.

Exercise 3

Present an example where data mining is crucial to the success of a business. What data mining functionalities does this business need (e.g., think of the kinds of patterns that could be mined)? Can such patterns be generated alternatively by data query processing or simple statistical analysis?

Exercise 4

Outliers are often discarded as noise. However, one person's garbage could be another's treasure. For example, exceptions in credit card transactions can help us detect the fraudulent use of credit cards. Using fraudulence detection as an example, propose two methods that can be used to detect outliers and discuss which one is more reliable.

Exercise 5

What are the major challenges of mining a huge amount of data (e.g., billions of tuples) in comparison with mining a small amount of data (e.g., data set of a few hundred tuples)?