

### **Question 1.**

**Discuss whether or not each of the following activities is a data mining task.**

1. Dividing the customers of a company according to their gender.

Ans: No, this is not a data mining task. We can get this information through simple interaction with a database management system.

2. Dividing the customers of a company according to their profitability.

Ans: No, this activity is not a data mining task. If profitability of customers is in customer records as an attribute, using a threshold we can divide the customers according to their profitability.

3. Computing the total sales of a company.

Ans: No, this activity is not a data mining task. We can calculate the total sales of a company by using simple calculation.

4. Sorting a student database based on student identification numbers.

Ans: No, this is not a data mining task. We can do sorting through simple database query.

5. Predicting the outcomes of tossing a (fair) pair of dice.

Ans: No, this is not a data mining task. This is probability calculation.

6. Predicting the future stock price of a company using historical records.

Ans: Yes, this is a data mining task because we want to predict the future stock price of a company based on the past data. It is a regression task because price is a continuous-valued attribute.

7. Monitoring the heart rate of a patient for abnormalities.

Ans: This activity is a data mining task. We can observe the heart rate of the patient and identify the abnormalities if the characteristics of the heart rate are different from normal observation. Here we use anomaly detection.

8. Monitoring seismic waves for earthquake activities.

Ans: Yes, this is a data mining task. Here we can build a model of different types of seismic wave behavior associated with earthquake activities and raise an alarm when different types of seismic activity observed. This is a classification task.

9. Extracting the frequencies of a sound wave.

Ans: No, this is not a data mining task. This is related to the signal processing.

## **Question 2.**

**Suppose that you are employed as a data mining consultant for an Internet search engine company. Describe how data mining can help the company by giving specific examples of how techniques, such as clustering, classification, association rule mining, and anomaly detection can be applied.**

Ans:

Data mining is a process used by companies to turn raw data into useful information. It is the process of finding anomalies, patterns and correlation within large data sets to predict outcomes.

- **Clustering:** It is the process of making a group of abstract objects into classes of similar objects. In the context of a search engine, clustering

can help to display the results that not only contain the keyword specified in the search but also related results.

- **Classification:** It is used to classify each item in a set of data into one of a predefined set of classes or groups. For example: documents on the internet are classified into predefined groups e.g. politics, science and sports etc.
- **Association rule mining:** It finds interesting associations and relationship among large sets of data item. A search engine could append additional information in its result based on the keyword entered by the user. For example: A user searching the web to buy a large screen TV might also be interested in a new home theatre system. Returning results for both TV and the home theatre system could keep the search engine one step ahead of the user.
- **Anomaly detection:** Anomaly detection or outlier detection is the task of identifying observation whose characteristics are significantly different from the rest of the data. A search engine can use anomaly detection to avoid displaying results that are not relevant to the standard keyword.

### **Question 3.**

**For each of the following data sets, explain whether or not data privacy is an important issue.**

1. Census data collected from 1900–1950.

Ans: No

2. IP addresses and visit times of web users who visit your website.

Ans: Yes

3. Images from Earth-orbiting satellites.

Ans: No

4. Names and addresses of people from the telephone book.

Ans: No

5. Names and email addresses collected from the Web.

Ans: No