

CIS 335 Data Mining

Homework 1

Details

Due: section 1: 1/18/18, section 2: 1/17/18

Files: hw01data.txt and hw01handin.doc

Textbook questions

Answer the following questions with concise answers.

1. Fill in the missing tasks of the knowledge discovery process
 - a) data cleaning
 - b) _____
 - c) selection
 - d) _____
 - e) pattern evaluation
2. Name two tasks of data pre-processing
3. Which data mining function is used to predict classes of objects?

R practice

Follow the instructions below to install R. The instructions will also explain how to do the basic tasks.

1. create a folder on the computer you will be using for this course (cis335 is a good name).
2. create subfolders in the main folder for homeworks and projects.
3. copy the file hw01data.txt from BlackBoard to the folder for homework1
4. install R from <https://www.r-project.org/>
5. change the working directory to the folder where your homework1 files are
6. use `list.files()` to show the files in the folder
7. use `ls` to show the objects in the workspace

8. create a vector named iq of 100 elements of $N(100,20)$ (normal with a mean of 100 and std dev of 20) data. ex. `iq <- rnorm(100,100,20)`
9. add 10 to every element of iq
10. calculate the mean and std dev of iq
11. create a random vector of numbers using `sample(1:10)`
12. create a vector called a with the values 3, 4, 7, 2, 1
13. create a vector called b with the values 1, 0, 0, 1, 1
14. read the file hw1data.txt into the matrix x using `read.table("hw01data.txt",header=TRUE)`
15. show just first row of x using `x[1,]`
16. show just the second and third columns using `x[,2:3]`
17. calculate the column means of rows 1 through 4 using `colMeans`
18. show just the value of x at row 25, column 3