

Assignment 1

CIS 453/553 Data Mining, Fall 2017

due 11:59 pm, Friday April 21

1. Based on your understanding, please define what is data mining? List three major research issues of data mining. List three applications of data mining which are not mentioned in the class and textbook . If you are working for a bank, which data mining technique(s) do you think are very possible to be used for automatically approving or declining someone's credit card application? Why?

2. Based on the figure 2.5 in the text (Slide 20 of the week 2 slides), what kind of conclusion you can draw? Can you make a judgement which branch has made more profit? Why?

3. Suppose that the data about Ph.D. applicants from country C includes the attribute GPA. The values are (in increasing order): 61, 67, 69, 71, 76, 79, 81, 82, 83, 84, 85, 85, 86, 87, 87, 88, 88, 89, 90, 93, 95, 95, 95, 98. According to American GPA system, 60 is equivalent to GPA value 1.0, 90 is equivalent to value 4.0, 95 or above is equal to 4.3.

(a) Write a program to transform those values to the range [1.0, 4.3] and list all transformed values.

(b) Can decimal scaling be used for the transformation in some way? If yes, what is the value for j ? If no, why?

(c) Please show all transformed data in a boxplot. Are there potential outlier(s)? If there are outlier(s), please use one method to clean the data and show the cleaned data.

4. Suppose a bank database includes following attributes to describe customers: name, age, gender, address, phone#, credit-ranking (good and bad), year-income, job title (student, engineer, professor etc..). Based on your common knowledge, propose a possible schema with reduced number of attributes for comparing the customers with good or bad credit-ranking in the bank database. Explain your rationale.

5. Do some literature search and summarize what is DFT (discrete Fourier transform). Then explain how to use DFT for dimensionality reduction based on your understanding of the literature.

To turn in by paper version: Ask Cheri to put your answers to Prof. Dejing Dou's mailbox.

To turn in by emails: We prefer that you send in a pdf file. If you are using Word, you should be able to convert your word file to a pdf file.