CIS 635 Data Mining

Homework 1

Description

In this first assignment you are to consider some scenarios where data mining is to be done and describe the process of preparing the data to be mined. You are to identify the important details of the data and describe them. Additionally, you will install R studio and do perform some fundamental commands. Write or copy/paste your answers to the handin sheet and submit on BB.

Instructions

Part 1 – general data mining

1. Game of Thrones (GoT) is a popular set of books and HBO series about a mythic land and characters. It is inspired by the War of the Roses and other historical events. As part of the story many characters are killed. For this assignment we would like to consider predicting which characters will get killed. Write down a short list of 4 attributes that could use in a classifier. Also, what would be the class? What would be the training set?
2. assume that we want to cluster students in this class. What would you use for attributes? What would you use for a proximity measure?

part 2 – getting to know R

1. create a folder on the computer you will be using for this course (cis635 is a good name).
2. create subfolders in the main folder for homeworks and projects.
3. install R from https://www.r-project.org/
4. change the working directory to the folder where your homework1 files are
5. create a vector named iq of 100 elements of N(100,20) (normal with a mean of 100 and std dev of 20) data.
6. add 10 to every element of iq
7. calculate the mean and std dev of iq
8. reassign to iq, one at a time, vectors of 100, 1000 and 10000 elements of N(100,20):
    a) for each vector calculate the mean and std dev.
    b) as the number of samples increase what happens to the mean? std dev? Explain.