

Homework 1

Overview

Objectives

- Read through the course syllabus
- Identify nominal versus ordinal attributes
- Consider general data mining tasks

Grading

- Uploaded requested files, 5%
- File is properly/clearly formatted, 5%
 - Proper section headers for each part of your homework.
 - You clearly indicate which question each of your responses are associated with.
- Part A., 10%
- Part B., 30%
- Part C., 50%

Deliverables

- **.pdf file** with your responses

Formatting

In the document that you turn in for this homework, include a heading for each Part (Part A, Part B, Part C). Ensure that you **clearly** indicate which question each of your responses are associated with.

Part A. Syllabus

1. Read the course syllabus (posted on blackboard). For this question (on the document that you turn in), confirm that you have read the syllabus.
2. Read the academic honesty policy for the GVSU School of Computing: <https://www.gvsu.edu/computing/academic-honesty-30.htm>. For this question (on the document that you turn in), confirm that you have read and agree to abide by the academic honesty policy.

Part B. Attribute descriptions

You are given a data file (see assignment attachments on blackboard): **hw01data.csv**. This is a table of ratings for teaching assistants (TA) at some university (this is modified data that originally came from <https://archive.ics.uci.edu/ml/datasets.php>). There are 6 attributes:

- **id**: unique number to identify each rating
- **instructor**: id number that identifies the instructor
- **classNbr**: id number that identifies the course
- **semester**: 1=summer, 2=fall, 3=winter
- **classSize**: number of students in the class
- **rating**: 1=low, 2=medium, 3=high

For each attribute, describe whether it is nominal or ordinal.

1. `id`
2. `instructor`
3. `classNbr`
4. `semester`
5. `classSize`
6. `rating`

Part C. Thinking about common data mining applications

1. This question asks you to think about clustering. Considering an average kitchen, identify 4 clusters of objects. Name them according to their utility. For example, if I were asked to identify 4 clusters in a sporting goods store, I might choose apparel (e.g., shoes, shirts), containers (e.g., backpacks, coolers), sports equipment (e.g., basketball, tennis racquets), and fishing (e.g., rods, lures).
2. Pick one of the following forms of predictive modeling: regression, classification, or ranking.
 - Describe an application domain or a dataset where you think it would be useful to apply the chosen predictive modeling approach.
 - What attribute(s) would you be predicting?
 - What attributes would you use to make the prediction?