1   Introduction

There are two options for this project.

- A classification project involving various types of glass.

- A Recommender system project involving beer ratings.

2   Classification Project

The classification problem involves a dataset from the UC-Irvine Machine Learning Repository. Full information can be found here.

The dataset, `glass.csv`, which I've cleaned for our purposes, consists of 9 numerical variables: RI, the refractive index, and 8 oxides, as a percentage by weight. The variable for classification is the type of glass, of which there are 6:

A: building windows float processed

B: building windows non float processed

C: vehicle windows float processed

D: vehicle windows non float processed (none in this database)

E: containers

F: tableware

G: headlamps

   Following the example in the sample code, you will split your dataset into a training and testing set. You will use the training set to produce the model, and you will use your model on the test set, with the goal of getting the highest accuracy possible. You should compare at least two different models. (Note: for the confusionMatrix function to work, you will need to make sure that the Type is a *factor* variable.

3   Beer Recommendations

BeerAdvocate.com is a beer rating website. The Stanford Network Analysis Project once hosted a dataset containing over 1.5 million beer reviews by consumers. For this project I have collected a list of 10 frequently rated beers and 10 reviewers who have assigned ratings to all of them. The full review matrix is in the file `fullratings.csv`. I have removed 8 of the 100 ratings, saving this matrix as `partialratings.csv`. Your goal is to approximate the missing ratings as closely as possible. This closeness can be measured by the norm of the difference between the true and approximated ratings matrices. You should use at least two methods. You can code it entirely by hand, or you can use the `recommenderlab` package that we saw earlier in the semester.

4   Coding help

The `Project2Help.R` file contains sample code to help with some of the work. Note that the `FL21movies.csv` file is not part of the project itself. It is there to help you see how some of the code works.

5   To turn in

Submit your work to TLearn. Your completed project will include

- A short paper discussing your project and results.

- Reproducible code so that I can reproduce your results.