

Instructions for Exam 1

By Rafael Becerril Arreola, Associate Professor of Marketing, DMSB

In this exam, you will implement a simplified version of the user-to-user (or user-based) collaborative filtering algorithm. This is a technique used to predict the items that a user might purchase on the basis of purchases of that item by the other users who have purchase histories similar to that of the target user.

A basic algorithm can be represented with the following pseudocode:

- 1) for each customer j in the data
 - a) compute the distances to every other customer k using their purchase histories on P products, $x_k = [x_{1k}, \dots, x_{Pk}]'$, such that

$$d(j, k) = \sqrt{\sum_{i=1}^P (x_{ij} - x_{ik})^2}$$

- b) identify the customer s (or set of customers, S , if ties exist)

$$s = \left\{ \underset{k, k \neq j}{\operatorname{argmin}} d(j, k) \right\}$$

who is (are) most similar to customer j

- c) make a vector L of items purchased by s (S)
 - d) remove duplicates in the vector L
 - e) remove from vector L the products that customer j has bought in the past
 - f) report vector L

To implement and test the algorithm, please follow the steps below.

1. Use the data file provided together with these instructions. Each row $j=1, \dots, 8$ represents a customer, each column $i=1, \dots, 10$ represents a product. The entries x_{ij} are either “0” or “1”. A “0” represents “no purchase” and a “1” represents “purchase”. For example, if the data are loaded into a data frame called “purchHist”, then `purchHist$x2[1]` equals “1” if and only if customer 1 has purchased product 2.
2. Translate the pseudocode above into R code and store it in file called “E1_X_Y.r”, where X should be your first name and Y your last name. To report the recommended products, simply let your code print them with R’s print command. Note that the number of products recommended can vary across customers and it can be zero if there are no new products to recommend. That is fine.

Reminders

Do not forget to:

1. Use comments in your code to explain lines of code that are not self-explanatory.
2. Upload your complete and self-contained R script as a file with extension “.r”
3. Include `rm(list=ls())` at the very top of your script
4. Ensure that the script produces the right results when called with the `source` command.