DAT16 SF: HOMEWORK 8 ASSIGNMENT

Assigned: Wednesday, September 16, 2015
Due: Monday, September 21, 2015, before class
Review Due: Wednesday, September 23, 2015, before class

The purpose of this homework is to review what we've learned about Recommender Systems.

HOMEWORK QUESTIONS

DUE MONDAY:

1. Complete the second part of the Lab (install RecSys and recommend movies)
2. Use the datasets of book crossing you can find here (use the csv dump): http://www2.informatik.uni-freiburg.de/~cziegler/BX/
3. Inspect the 3 datasets and describe their content
4. Re-arrange the data in the book-user dataset to a Dataframe where users are columns and books are rows.
5. Calculate the matrix of cosine_similarity distances between books and between users.
6. Define a function that finds the K most similar users to a given user and another function that returns the K most similar items to a given item.
7. Using the RecSys package, apply SVD decomposition to the book-user ratings DataFrame of point 3.
8. Write a function that given a user recommends the next book this user should read. The function should return the title of the book, not just the id (Use the additional dataset of the book titles)
9. Parse the users dataset, extract the state or country from the "location" column and find the most common locations.

BONUS POINTS:

Find a python package that allows plotting a map of the world and put circles on the map with a size proportional to the number of users in that state or country.

DUE WEDNESDAY:

1. Go to your new assigned review-buddy's repo
2. Read through your buddy's ipython notebook and make sure you understand what he/she is doing.
3. Open an issue in his/her repo and write comments on the things you don't understand and on the things you like in his/her code.
4. Quote the instructors in the comments so that we get notified about the open issue