Isfahan University of Technology
Information Technology Fundamentals
Dr. Manshaei
Homework 2: Module 5 Web System
Deadline: Tuesday 3^{rd} **Khordad, at** 23 : 59

We only accept the homework **delivered via Yekta(yekta.iut.ac.ir), before the deadline**.

Overview

There are some packages for implementing recommender systems. These packages simplify the process of creating a recommender system. One of these packages is Surprise. Surprise is a Python scikit for building and analyzing recommender systems that deal with explicit rating data. Surprise does not support implicit ratings or content-based information. With Surprise, we have the following features:

- Alleviate the pain of Dataset handling. Users can use both built-in datasets (Movielens, Jester) and their custom datasets.
- Provide various ready-to-use prediction algorithms such as baseline algorithms, neighborhood methods, matrix factorization-based (e.g., SVD, PMF, SVD++, and NMF), etc. Also, various similarity measures (e.g., cosine, MSD, and Pearson) are built-in.

You can use python files or a Jupyter Notebook file for this homework. (For example, you can use Jupyter Notebook in PyCharm)

Goals and Instructions

1. Search and study **SVD** and **KNN** algorithm. Then, write a summary of both and compare them. (Please avoid copying the documentation from the Internet)
2. Install Surprise and its requirements and load the movielens-100k dataset.
3. Create SVD and KNNBasic algorithm.
4. Use 4-fold cross-validation for evaluating **RMSE** and **MAE** with SVD and KNNBasic algorithm. Please read this link if you don't know what RMSE and MAE are. Analyze the output of cross-validation in your report.
5. We need to have two datasets. The first dataset is for training the model, and the second is for testing the models. So we need to split our data into two parts. Split the data into train and test with the train-test-split function and split size of 0.25.
6. Fit the trainset and use testset for prediction of both SVD and KNNBasic algorithm. Then, check the accuracy. (link)
7. Now predict ratings by directly calling the predict() method to predict the rating for the **testset**. (Other methods for predication could be applied as well.)
8. Analyze the result of the predictions.

Which dataset will you use for this homework?

You just need to import the dataset from Surprise and load "ml-100k". This is a straightforward and ready dataset.

What should you upload as your homework?

- Your codes.
- A report of what you did.
- A video for explaining whatever you did. You should analyze and explain your solutions. (This video must be at most 15 minutes).

What should you do if you need help?

- Available links in this file
- The Surprise documentation
- GitHub
- Contact the TA (Reach out to Ghazaleh Zehtab at ghazalze@yahoo.com or Skype)