SMAI - Interim Evaluation 1

Project Abstract

Problem Statement : Given the user's preferences in movies and a general dataset of all the users and their ratings we have to predict how the user would rate an arbitrarily selected movie. To solve this problem we would use collaborative filtering which is a mechanism to filter massive amounts of data based upon a previous interactions of a large number of users. In this project we will analyze and benchmark several collaborative filtering methods. These methods are :- Stochastic Gradient Descent (SGD), Alternating Least Squares (ALS), Biased Stochastic Gradient Descent (B-SGD), Weighted Alternating Least Squares (W-ALS).

Paper which we are following is :
http://cs229.stanford.edu/proj2014/Christopher%20Aberger,%20Recommender.pdf

Input Dataset

Downloaded from :- http://grouplens.org/datasets/movielens/1m
The dataset contains around 1,000,209 ratings of approximately 3,900 movies
made by 6,040 MovieLens users who joined MovieLens in 2000.

The dataset is separated into three tables - ratings, users, movies. Ratings contain UserID, MovieID, Rating (1-5). User table contains UserID, Gender, Age Group and Occupation. Movies table contains MovieID, Title and Genres. However, we will only use genre and age group only if time permits later.

Project Plan

First, we will convert the dataset into a matrix of users vs. ratings. Then, we will implement low rank decomposition using four methods SGD, B-SGD, ALS, W-ALS.
We will cover SGD, ALS by Second Interim Evaluation and the other two by the Final Evaluation.

Platform

Project will implemented in Python and no special tools will be required for these tasks.