

### CS 599 - Information Retrieval, Homework 3

		Titles				
		Starship Trooper (A)	Sleepless in Seattle (R)	MI-2 (A)	Matrix (A)	Titanic (R)
Users	Sammy	3	4	3	?	?
	Beatrice	3	4	3	1	1
	Dylan	3	4	3	3	4
	Mathew	4	2	3	2	5
	John	4	3	4	4	4
	Basil	5	1	5	?	?

- 1) Consider the table of movies and the ratings for 6 users.
  - a. Calculate the Pearson Correlation for the users Sammy and Basil using their neighborhood (Beatrice, Dylan, Mathew and John).
  - b. Predict the rating for users Sammy and Basil for the movies, Matrix and Titanic using the calculated Pearson Correlations.
  - c. Sammy says his actual ratings for Matrix and Titanic are 4 and 3. For Basil its 2 and 5 respectively. Now calculate the MAE and RMSE based on the predicted and actual ratings for this Collaborative Filtering method.
  
- 2) Write a simple Collaborative Filtering program in the language of your choice to solve the above problem.
  - Use the data from Beatrice, Dylan, Mathew and John as the training dataset.
  - Use the Sammy and Basil as your testing data for prediction.
  - Use Pearson Correlation and Aggregation method learned in class to predict ratings.
  - Your program should have an evaluation component to test (data from question 1 part c.) the accuracy using both MAE and RMSE for actual ratings.
  - Submissions (2 parts):
    - Your code: Submit your code to Blackboard (zip) including help.txt which explains how to run the program. Your training data should be in training.dat and testing data in the test.dat, prediction results in prediction.dat and evaluation results in eval.dat
    - Your program screenshots: Print and submit with your solutions to question 1.