Collaborative Filtering [CSE640]
Assignment 2 Report

Name: Yash Bhargava
Roll no: 2019289

MAE values for user-based clustering

	Crisp K means clustering			Fuzzy C means clustering		
Fold #	K=5	K=10	K=20	C=5	C=10	C=20
Fold 1	0.8275260343	0.8290220104	0.846270308	0.822262018	0.8215345011	0.8218702781
Fold 2	0.7987841112	0.8051159367	0.8250531789	0.7923345897	0.7927887017	0.7925162345
Fold 3	0.8046721723	0.8151504001	0.8169304556	0.8051444043	0.8048285199	0.805234657
Fold 4	0.8067929382	0.8138310459	0.8467261905	0.8000812206	0.7999651912	0.799559088
Fold 5	0.8064691492	0.8053738827	0.82175991	0.7967957708	0.7962023951	0.7957169058
Average	0.808848881	0.8136986552	0.8313480086	0.8033236007	0.8030638618	0.8029794327

Crisp K means Clustering

$$p(a,j) = \frac{\sum_{i \epsilon N} cos\_sim(a,i) \times rating(i,j)}{\sum_{i \epsilon N} cos\_sim(a,i)}$$

where

p(a,j) : predicted rating by active user ‘a’ of movie ‘j’
cos_sim(a,i) : cosine similarity between user ‘a’ and user ‘i’
rating(i,j) : rating by user ‘i’ of movie ‘j’ (co-rated)
N : subset of the cluster to which user ‘a’ belongs and comprising of users ‘i’ with cos_sim(a,i)>0.5

Fuzzy C means Clustering

$$cluster\_rating(k,j) = \frac{\sum_{all\ users\ i} membership(i,k) \times cos\_sim(a,i) \times rating(i,j)}{\sum_{all\ users\ i} membership(i,k) \times cos\_sim(a,i)}$$

$$p(a,j) = \frac{\sum_{all\ clusters\ k} membership(a,k) \times cluster\_rating(k,j)}{\sum_{all\ clusters\ k} membership(a,k)}$$

where

p(a,j) : predicted rating by active user ‘a’ of movie ‘j’
membership(i,k) : membership score of user ‘i’ in cluster ‘k’
cos_sim(a,i) : cosine similarity between user ‘a’ and user ‘i’ (more than 0.5)
rating(i,j) : rating by user ‘i’ of movie ‘j’ (co-rated)
membership(a,k) : membership score of active user ‘a’ in cluster ‘k’
cluster_rating(k,j) : rating by cluster ‘k’ of movie ‘j’