CSE 640: Collaborative Filtering

Assignment-2

22nd March 2022

Name: Sakshat Mali

Roll No.: 2019327

Crisp K-Means

For predicting the rating of new users I am following the below logic and steps:

- Fit the training dataset into the sklearn.cluster KMeans.
- After fitting the training dataset into the kmeans model we will get the labels for each of the users in the training dataset which will tell us that which user belongs to which cluster.
- Now for each user in the training dataset, we will traverse over all the movies that he/she has watched and this we will do for all the users belonging to the same cluster.
- We will take the average of all the ratings on that particular movie.
- Now after this we will get the matrix of *Clusters* ***x*** *Movies*, this matrix tells us that for a particular cluster what is the rating of every movie.
- Now, whenever a new user comes from the test dataset we predict its cluster from our kmeans model.
- Then we traverse on all the movies watched by that user and append find its actual rating and append this into the actual list.
- For the prediction list, we append the value that is present in the *Cluster* ***x*** *Movie* matrix for that cluster in which the new user belongs.
- Now we have both actual and predicted ratings we simply calculate MAE.

Fuzzy C-Means

For predicting the rating of new users I am following the below logic and steps:

- Fit the training dataset into the fcmeans FCM (module in python).
- After fitting the training dataset into the fcmeans model we will get the membership score for each of the users in the training dataset which will tell us how close a particular user is with a particular cluster.
- Now for each user in the training dataset, we will traverse over all the movies that he/she has watched and this we will do for all the users belonging to the same cluster.
- We will take the weighted average of all the ratings on that particular movie.
- For weights, we will consider the membership score and by using this we will multiply the rating by the membership score of that particular score and will divide by the sum of all membership scores (weighted average).
- Now after this we will get the matrix of *Clusters **x** Movies*, this matrix tells us that for a particular cluster what is the rating of every movie.
- Now, whenever a new user comes from the test dataset we predict its closest cluster from our fcmeans model.
- Then we traverse on all the movies watched by that user and append find its actual rating and append this into the actual list.
- For the prediction list, we append the value that is present in the *Cluster **x** Movie* matrix for that cluster in which the new user belongs.
- Now we have both actual and predicted ratings we simply calculate MAE.

MAE Tables

Table: Crisp K-Means Clustering

	K=5	K=10	K=20
Fold 1	0.814964	0.833601	0.864075
Fold 2	0.804468	0.822784	0.841602
Fold 3	0.814117	0.833988	0.850382
Fold 4	0.779515	0.797933	0.812925
Fold 5	0.790836	0.806714	0.831904
Average	0.800780	0.819004	0.840177

Table: Fuzzy C-Means Clustering

	C=5	C=10	C=20
Fold 1	0.797933	0.797933	0.797933
Fold 2	0.784090	0.784140	0.784190
Fold 3	0.801517	0.801563	0.801517
Fold 4	0.767736	0.767843	0.767896
Fold 5	0.768341	0.768341	0.768243
Average	0.783923	0.783964	0.783956