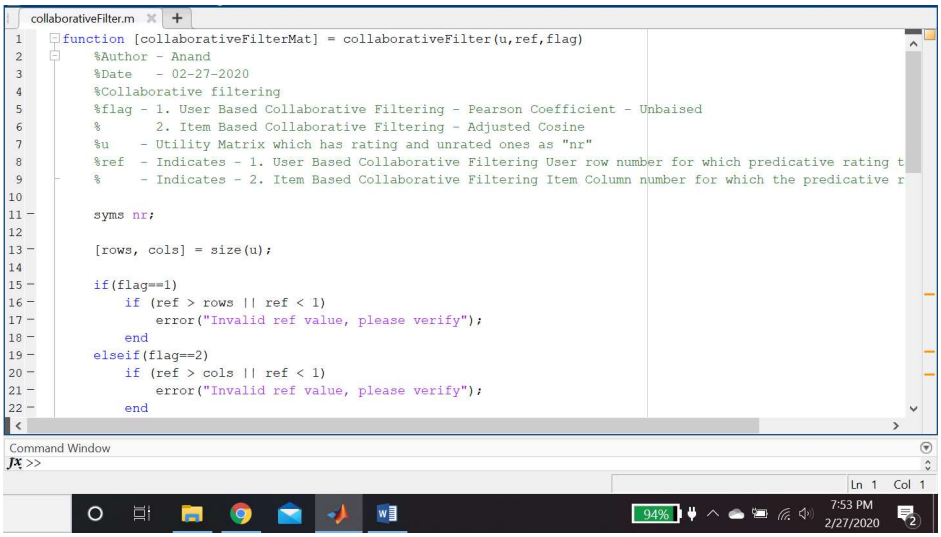
Collaborative Filtering

1. **Open collaborativeFilter.m file in matlab.**



2. **Let's prepare matrix as per the problem mentioned below :**

Table 2.1: User-user similarity computation between user 3 and other users

Item-Id ⇒ / User-Id ⇓	1	2	3	4	5	6	Mean Rating	Cosine$(i,3)$ (user-user)	Pearson$(i,3)$ (user-user)
1	7	6	7	4	5	4	5.5	0.956	0.894
2	6	7	?	4	3	4	4.8	0.981	0.939
3	?	3	3	1	1	?	2	1.0	1.0
4	1	2	2	3	3	4	2.5	0.789	-1.0
5	1	?	1	2	3	3	2	0.645	-0.817

Before preparing the utility matrix in Matlab set syms nr; it will accept symbolic letters in matrix so items which are not rated can be filled in matrix as "nr" – not rated.

```
>> syms nr;
>> A=[7 6 7 4 5 4;6 7 nr 4 3 4;nr 3 3 1 1 nr;1 2 2 3 3 4;1 nr 1 2 3 3]

A =

[  7,  6,  7, 4, 5,  4]
[  6,  7, nr, 4, 3,  4]
[ nr,  3,  3, 1, 1, nr]
[  1,  2,  2, 3, 3,  4]
[  1, nr,  1, 2, 3,  3]
```

Utility matrix is ready and its stored in variable A.

collaborativeFilter(u,ref,flag)

3. **Let's find user similarity of User 3 with other users.**

   a. **<u>User based collaborative filtering.</u>**

   U is utility matrix – which is stored in variable A
   Ref – 3 we have to find similarity of user 3 with other users
   Flag – 1 for user based collaborative filtering.
   Call in matlab as
   collaborativeFilter(A,3,1);
   Result:

```
>> collaborativeFilter(A,3,1)

ans =

    1.5000    0.5000    1.5000   -1.5000   -0.5000   -1.5000    0.8940    5.5000
    1.2000    2.2000       NaN   -0.8000   -1.8000   -0.8000    0.9380    4.8000
       NaN    1.0000    1.0000   -1.0000   -1.0000       NaN    1.0000    2.0000
   -1.5000   -0.5000   -0.5000    0.5000    0.5000    1.5000   -1.0000    2.5000
   -1.0000       NaN   -1.0000         0    1.0000    1.0000   -0.8160    2.0000
```

   It's a matrix with mean rating subtracted from all the user rating.
   Orange block- represents your centered rating matrix
   Green block – represents the pearson coefficient
   Blue block – represents the mean rating of the user.

   b. **<u>Item based collaborative filtering.</u>**

   U is utility matrix – which is stored in variable A
   Ref – 1 we have to find similarity of item 1 with other items
   Flag – 2 for item based collaborative filtering.

   Call in matlab as
   collaborativeFilter(A,1,2);
   Result:

```
>> collaborativeFilter(A,1,2)

ans =

    1.5000    0.5000    1.5000   -1.5000   -0.5000   -1.5000    5.5000
    1.2000    2.2000       NaN   -0.8000   -1.8000   -0.8000    4.8000
       NaN    1.0000    1.0000   -1.0000   -1.0000       NaN    2.0000
   -1.5000   -0.5000   -0.5000    0.5000    0.5000    1.5000    2.5000
   -1.0000       NaN   -1.0000         0    1.0000    1.0000    2.0000
    1.0000    0.7350    0.9120   -0.8480   -0.8120   -0.9900         0
```

   Orange block represents – centered rating matrix(subtracted from user mean rating)
   Blue block represents – mean rating of the user
   Green block represents – Adjusted cosine similarity for the item 1 wrt other users

Similarly let's find for item 6:

```
>> collaborativeFilter(A,6,2)

ans =

    1.5000    0.5000    1.5000   -1.5000   -0.5000   -1.5000    5.5000
    1.2000    2.2000       NaN   -0.8000   -1.8000   -0.8000    4.8000
       NaN    1.0000    1.0000   -1.0000   -1.0000       NaN    2.0000
   -1.5000   -0.5000   -0.5000    0.5000    0.5000    1.5000    2.5000
   -1.0000       NaN   -1.0000         0    1.0000    1.0000    2.0000
   -0.9900   -0.6220   -0.9120    0.8290    0.7300    1.0000         0
```

Solution for problem:

QUESTION 1

Consider the following ratings table between five users and six items:

Item-Id ⇒	1	2	3	4	5	6
1	5	6	7	4	3	?
2	4	?	3	?	5	4
3	?	3	4	1	1	?
4	7	4	3	6	?	4
5	1	?	3	2	2	5

a) Predict the values of unspecified ratings of user 2 using user-based collaborative filtering.

b) Predict the values of unspecified ratings of user 2 using item-based collaborative filtering.

Solution:

Let's prepare utility matrix for the given question in matlab.

```
>> syms nr;
>> A=[5 6 7 4 3 nr;4 nr 3 nr 5 4;nr 3 4 1 1 nr;7 4 3 6 nr 4;1 nr 3 2 2 5]

A =

[  5,  6, 7,  4,  3, nr]
[  4, nr, 3, nr,  5,  4]
[ nr,  3, 4,  1,  1, nr]
[  7,  4, 3,  6, nr,  4]
[  1, nr, 3,  2,  2,  5]
```

nr signifies - ? unrated item

a. User based collaborative filtering:

We have to find users similar to user 2 and predict rating of items in user 2 which are not rated.

```
>> collaborativeFilter(A,2,1)

ans =
```

0	1.0000	2.0000	-1.0000	-2.0000	NaN	-1.0000	5.0000
0	NaN	-1.0000	NaN	1.0000	0	1.0000	4.0000
NaN	0.7500	1.7500	-1.2500	-1.2500	NaN	-0.9860	2.2500
2.2000	-0.8000	-1.8000	1.2000	NaN	-0.8000	0.6100	4.8000
-1.6000	NaN	0.4000	-0.6000	-0.6000	2.4000	-0.2380	2.6000

With pearson coefficient we get to know there is only one user which is similar to User 2 that is User 4.

Biased Predicted rating:

$$\frac{\sum_{v \in P_u(j)} \mathrm{Sim}(u,v) \cdot (r_{vj} - \mu_v)}{\sum_{v \in P_u(j)} |\mathrm{Sim}(u,v)|}$$

= (pearson coefficient of user 4 *actual rating of item 2 by user 4)/pearson coefficient of user 4

Biased predicted rating of item 2 of user 2 is 4

```
>> (0.61*4)/0.61

ans =

     4
```

Unbiased predicted rating :

$$\mu_u + \frac{\sum_{v \in P_u(j)} \mathrm{Sim}(u,v) \cdot (r_{vj} - \mu_v)}{\sum_{v \in P_u(j)} |\mathrm{Sim}(u,v)|}$$

=Mean user rating of 2 +(pearson coefficient of user 4 *centered rating of item 2 by user 4)/pearson coefficient of user 4)

```
>> 4+((0.61*-0.8)/(0.61))

ans =

    3.2000
```

Unbiased predicted rating of item 2 of user 2 is = 3.2

Similarly let's calculate for item 4 of user 2:

Biased:

= (pearson coefficient of user 4 *actual rating of item 4 by user 4)/pearson coefficient of user 4

```
>> (0.61*6)/0.61

ans =

     6
```

Biased predicted rating of item 4 of user 2 is 6

Unbiased:

=Mean user rating of 2 +(pearson coefficient of user 4 *centered rating of item 4 by user 4)/pearson coefficient of user 4)

```
>> 4+((0.61*1.2)/(0.61))

ans =

    5.2000
```

Unbiased predicted rating of item 4 of user 2 is 5.2

b. Item based collaborative filtering

Finding adjusted cosine value for the item 2 wrt to similar items as mentioned below:

```
>> collaborativeFilter(A,2,2)

ans =

         0    1.0000    2.0000   -1.0000   -2.0000       NaN    5.0000
         0       NaN   -1.0000       NaN    1.0000         0    4.0000
       NaN    0.7500    1.7500   -1.2500   -1.2500       NaN    2.2500
    2.2000   -0.8000   -1.8000    1.2000       NaN   -0.8000    4.8000
   -1.6000       NaN    0.4000   -0.6000   -0.6000    2.4000    2.6000
   -0.6250    1.0000    0.9980   -0.9760   -0.9960    1.0000         0
```

Item 3 and item 6 are similar to item 2 using adjusted cosine value.

Predicted rating of item2 of user 2 = (adjusted cosine of item 3*user 2 rating of item 3+adjusted cosine of item 6* user 2 rating of item 6)/(adjusted cosine of item 3+adjusted cosine of item 6)

```
>> (0.998*3+1*4)/(0.998+1)

ans =

    3.5005
```

Predicted rating of item 2 of user 2 is 3.5

Similarly let's calculate adjusted cosine for the item 4:

```
>> collaborativeFilter(A,4,2)

ans =

         0    1.0000    2.0000   -1.0000   -2.0000       NaN    5.0000
         0       NaN   -1.0000       NaN    1.0000         0    4.0000
       NaN    0.7500    1.7500   -1.2500   -1.2500       NaN    2.2500
    2.2000   -0.8000   -1.8000    1.2000       NaN   -0.8000    4.8000
   -1.6000       NaN    0.4000   -0.6000   -0.6000    2.4000    2.6000
    0.7910   -0.9760   -0.9750    1.0000    0.9430   -0.7070         0
```

For item 4 item 1 and item 5 are similar from adjusted cosine value.

Predicted rating of item 4 of user 2 = (adjusted cosine value of item 1 *user2 rating of item 1 + adjusted cosine value of item 5* user 2 rating of item 5)/(adjusted cosine value of item1 +adjusted cosine value of item5)

```
>> (0.7190*4+0.943*5)/(0.7190+0.943)

ans =

    4.5674
```

Predicted rating of item 4 of user 2 is 4.57

Summary :

User	User Based		Item Based
User 2	**Biased predicted rating**	**Unbiased predicted rating**	**Predicted Rating**
Item2	4	3.2	3.5
Item4	6	5.2	4.57

Observation:

User based considers rating of other users for the item and predicts the rating, where as item based considers rating of the same user of similar items and predicts the rating. So, item-based prediction has more accuracy when compared to user based as its induced rating of other users.