Question 5

For this part, first we find predicted R matrix by supplying R as a 0-1 matrix where 1 is when a rating is available and 0 otherwise. On the other hand, the weight matrix comprised of the actual ratings that the users had given.

While performing a 10 fold cross validation, we kept a track of the predicted ratings corresponding to the known data points. The next step is to sort the ratings for every user in the descending order to get the top L movies for every user.

Hit Rate is the movies recommended by the system which are liked by the user. For this, we calculate the number of movies in L for each user that has a value above threshold. This gives us the hit rate. On the other hand, the movies recommended by the system which are not liked by the user are counted as a false alarm. Thus any rating present in L falling below the threshold which indicates the user did not like it falls under this category. We got different values of Hit Rate and False Alarm Rate by increasing the L from 1 onwards.

When the value of L hit 5, the average precision of the algorithm based on the following formula was calculated:

Precision = 𝑡𝑟𝑢𝑒 𝑝𝑜𝑠𝑖𝑡𝑖𝑣𝑒/ 𝑡𝑟𝑢𝑒 𝑝𝑜𝑠𝑖𝑡𝑖𝑣𝑒 + 𝑓𝑎𝑙𝑠𝑒 𝑝𝑜𝑠𝑖𝑡𝑖𝑣𝑒

The entire process was repeated for k = 10, 50 and 100, and the following results were obtained:

Average precision for L = 5

|  |  |
| --- | --- |
| Dimension of K | Average Precision |
| 10 | 0.531707317073171 |
| 50 | 0.549946977730648 |
| 100 | 0.556945917285261 |

As we can notice, the average precision increases with an increase in the value of k.

The Hit Rate vs False Alarm Rate curve is also an increasing one in such a manner that it approaches 1.

