ECE219 Project 3

Collaborative Filtering

Zhilai Shen, Yufei Hu, Zheang Huai, and Tianyi Liu
105023454, 404944367, 505222324, 705035425

February 19, 2019

Contents

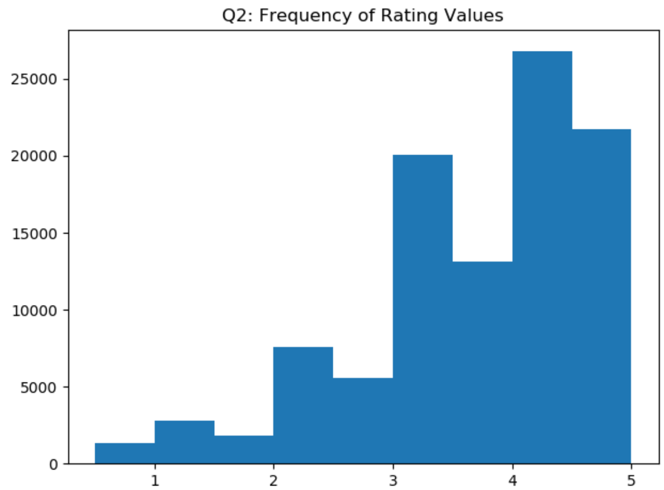

Figure 1: Frequency of the rating values

1 MovieLens dataset

Question 1:
The sparsity of the movie rating dataset is: 0.017.

Question 2:
The shape of the histogram is skewed to the right. More higher ratings fall in 4 and 5 while ratings in 1 and 2 are rather lower.

Question 3:
The distribution of the number of ratings received among movies can be seen in 2. A monotonically decreasing curve is observed.

Question 4:
The distribution of the number of ratings received among users can also be seen in 3. A monotonically decreasing curve is also observed.

Question 5:
As seen in 2, very few movies have more than 60 ratings, and most of the movies have fewer than 10 ratings. These features imply that the rating matrix is very sparse, which is the main challenge in the recommendation process as introduced in class. Also, about 500 movies have many ratings, meaning they

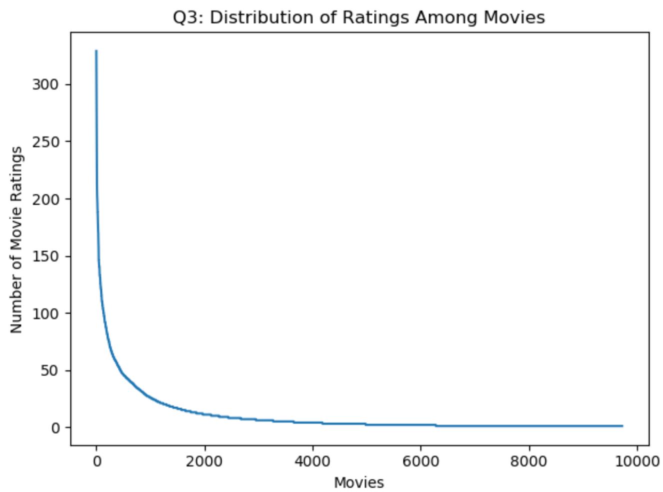

Figure 2: Distribution of ratings among movies

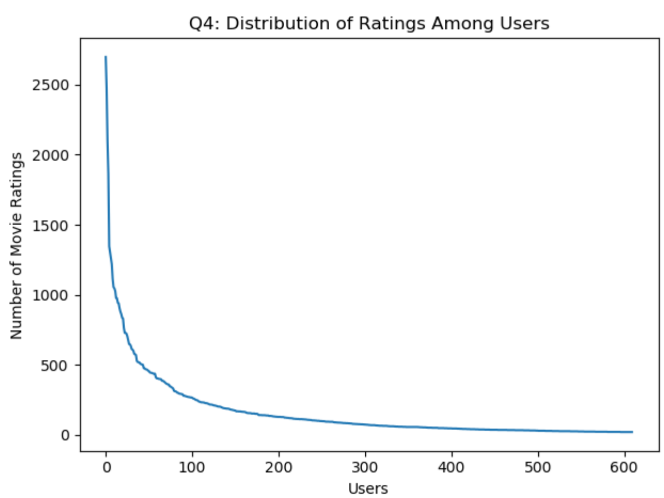

Figure 3: Distribution of ratings among users

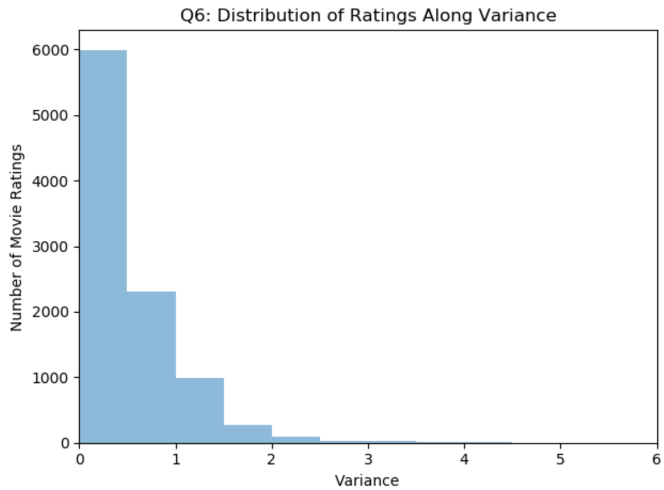

Figure 4: Variance of the rating values received by each movie

are more popular compared to the rest of the movies.

Question 6:

From 4, it is obversed that most of the movies receive similar ratings from users from the fact that their variances are less than 0.5. Very few movies have high variance of ratings among different users. This means that users have similar taste for most of the movies, which will ease the pain of designing our recommender system.

2 Neighborhood-based collaborative filtering

2.1 Pearson-correlation coefficient

Question 7:

$$\mu_u = \frac{1}{|I_u|} \sum_{k \in I_u} r_{uk}$$

Question 8:

$I_u \cap I_v$ means the set of item indices for which ratings have been specified by both user v and u. It can be an empty set when none of the items rated by u has been rated by v and vise versa.

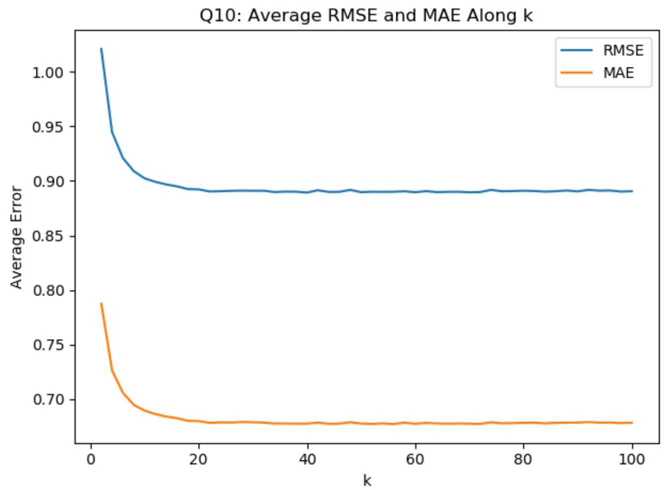

Figure 5: Number of neighbors k against average MAE and RMSE

2.2 Prediction function

Question 9:
Mean-centering effectively produces user v's preference on item j compared to his average rating. For example, if user v tends to give high ratings, we could produce an unbiased rating by subtracting v's mean rating μ_u from actual rating.

2.3 k-NN collaborative filter

2.3.1 Design and test via cross-validation

Question 10:
The plot of the number of neighbors k vs average MAE and RMSE, for k = 2 to 100 can be seen in 5.

Question 11: Minimum k of average RMSE is 22. Steady state values of average RMSE is 0.891. Minimum k of average MAE is 22. Steady state values of average MAE is 0.678.

2.4 Filter performance on trimmed test set

Question 12:
The plot of average RMSE against k, for k = 2 to 100 can be seen in Figure 6.The minimum average RMSE is 0.8724.

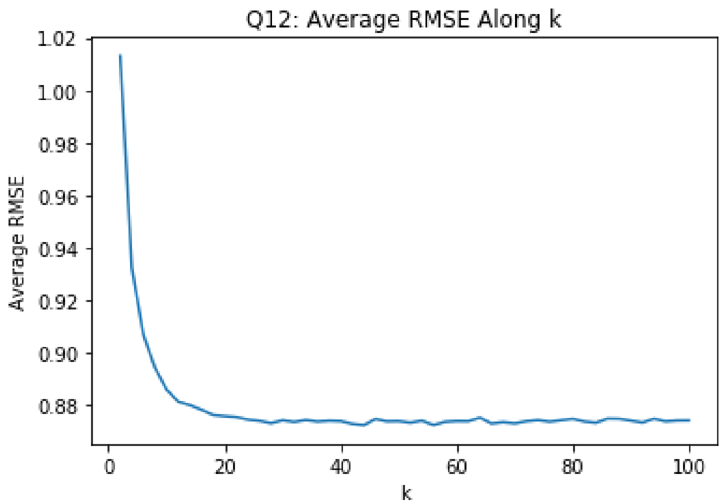

Figure 6: average RMSE of popular movie trimmed test set

Question 13:

The plot of average RMSE against k, for k = 2 to 100 can be seen in Figure 7.The minimum average RMSE is 1.1110.

Question 14:

The plot of average RMSE against k, for k = 2 to 100 can be seen in Figure 8.The minimum average RMSE is 1.5011.

2.4.1 Performance evaluation using ROC curve

Question 15:

Table 1: AUC values of different thresholds

threshold	AUC
2.5	0.7828
3	0.7813
3.5	0.7798
4	0.7756

The ROC curves for different threshold values [2.5,3,3.5,4] can be seen in Figure 9,Figure 10,Figure 11,Figure 12, respectively. The AUC values can be seen in Table 1.

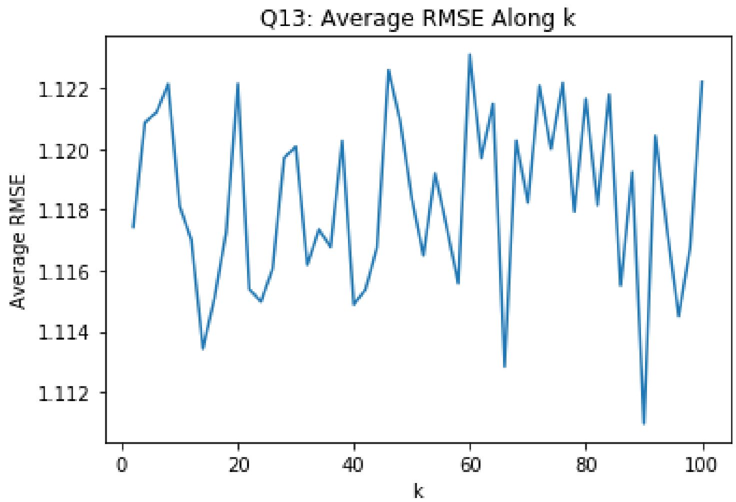

Figure 7: average RMSE of unpopular movie trimmed test set

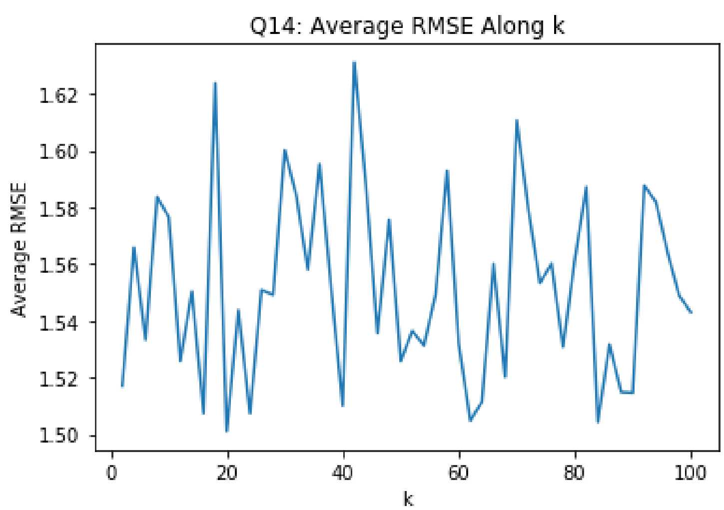

Figure 8: average RMSE of high variance movie trimmed test set

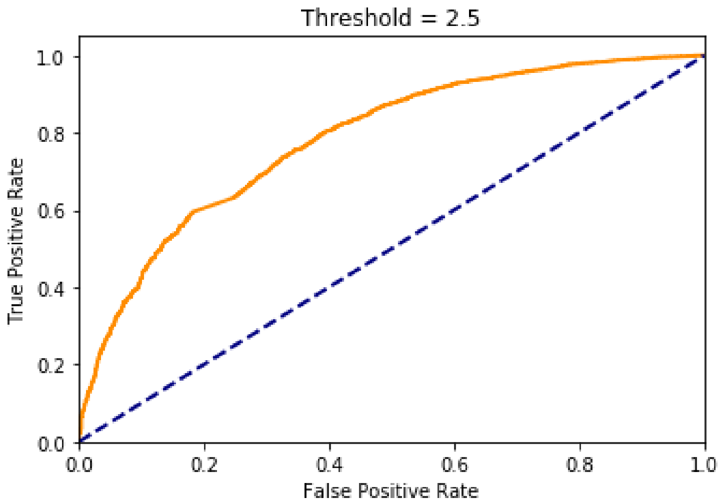

Figure 9: ROC curve for the k-NN collaborative filter, threshold=2.5

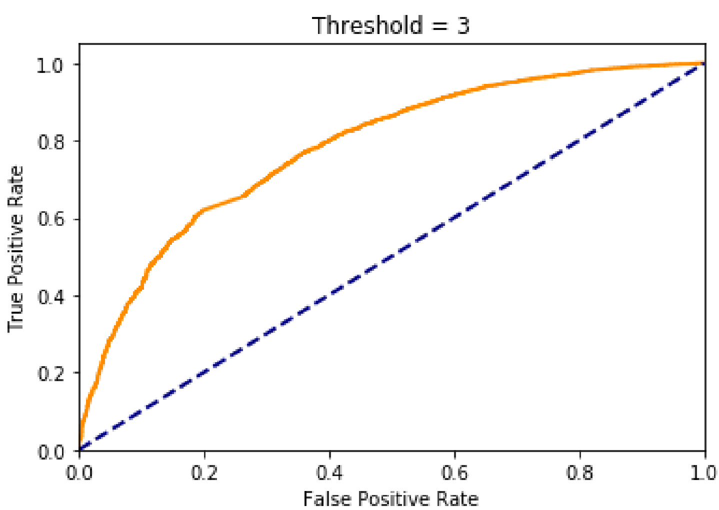

Figure 10: ROC curve for the k-NN collaborative filter, threshold=3

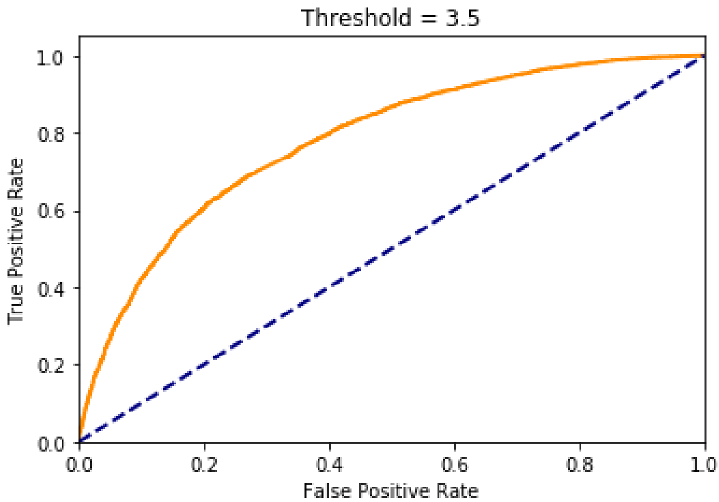

Figure 11: ROC curve for the k-NN collaborative filter, threshold=3.5

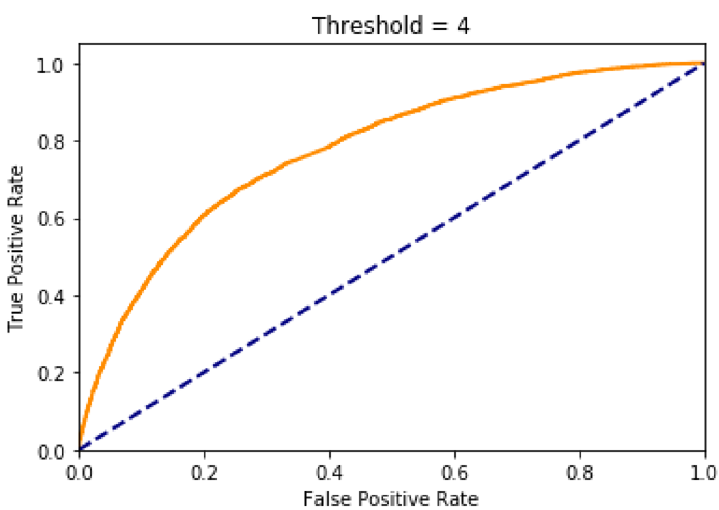

Figure 12: ROC curve for the k-NN collaborative filter, threshold=4

3 Model-based collaborative filtering

Question 16: The answer is in the picture below.

3.0.1 Design and test via cross-validation

Question 17:
The plot of the average RMSE and the average MAE against k, for k = 2 to 50 can be seen in Figure 13.
.

Question 18: The optimal number of latent factors is 20. The minimum average RMSE and MAE are 0.9204, 0.7009, respectively. As can be seen in the readme file of dataset, there are 18 genres. So the optimal number of latent factors is not exactly the same as the number of movie genres but very close to the number of movie genres.

3.0.2 NNMF filter performance on trimmed test set

Question 19:
The plot of average RMSE against k, for k = 2 to 50 can be seen in Figure 14. The minimum average RMSE is 0.8991.

Question 20:
The plot of average RMSE against k, for k = 2 to 50 can be seen in Figure 15. The minimum average RMSE is 1.1748.

Question 21:
The plot of average RMSE against k, for k = 2 to 50 can be seen in Figure 16. The minimum average RMSE is 1.6433.

3.0.3 Performance evaluation using ROC curve

Question 22: The plots are as in Figure 17, 18, 19, 20.
And the corresponding Area under Curve is in the table 2.

3.0.4 Interpretability of NNMF

Question 16

(1) Not convex. The simplest counterexample is the case when U, V are 1×1 matrices, $W_{11}=1$, and $r_{11}=0$. Then the objective function becomes $(-uv)^2 = u^2v^2$

$$\nabla(u^2v^2) = \begin{bmatrix} 2uv^2 \\ 2u^2v \end{bmatrix} \qquad \nabla^2(u^2v^2) = \begin{bmatrix} 2v^2 & 4uv \\ 4uv & 2u^2 \end{bmatrix}$$

Since $\det(\nabla^2(u^2v^2)) \le 0$, Hessian matrix is not positive semidefinite. Thus the objective function is not convex.

(2) Let the objective function be $f(U,V)$, $U^T=[u_1 \cdots u_m]$ $V^T=[v_1 \cdots v_n]$. Then $U=\begin{bmatrix} u_1^T \\ \vdots \\ u_m^T \end{bmatrix}$,

$$f(U,V) = \sum_{i=1}^{m}\sum_{j=1}^{n} W_{ij}(r_{ij} - u_i^T v_j)^2$$

$$= \sum_{i=1}^{m}\sum_{j=1}^{n} (W_{ij} r_{ij} - W_{ij} u_i^T v_j)^2$$

$$= \sum_{j=1}^{n}\sum_{i=1}^{m} (W_{ij} r_{ij} - W_{ij} u_i^T v_j)^2$$

$$= \sum_{j=1}^{n}\sum_{i=1}^{m} \left(W_{ij} r_{ij} - [0\ 0 \cdots W_{ij}u_i^T \cdots 0\ 0] \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix}\right)^2$$

where each "0" represents a $1\times k$ zero vector, and the index of $W_{ij}u_i^T$ is from $(j-1)k+1$ to jk.

If we define a $mn \times nk$ matrix A,

$$A = \begin{bmatrix} W_{11}u_1^T & 0 & - & - & - & \cdots & 0 \\ \vdots & W_{12}u_1^T & - & - & - & & 0 \\ \vdots & \vdots & & & & & \vdots \\ & & & & & & W_{1n}u_1^T \\ W_{21}u_2^T & 0 & - & - & - & \cdots & 0 \\ \vdots & & \vdots & & & & \vdots \\ 0 & - & 0 & \cdots & & - & W_{mn}u_m^T \end{bmatrix} \qquad V' = \begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix} \qquad b = \begin{bmatrix} W_{11}r_{11} \\ \vdots \\ W_{1n}r_{1n} \\ W_{21}r_{21} \\ \vdots \\ W_{2n}r_{2n} \\ \vdots \\ W_{mn}r_{mn} \end{bmatrix}$$

$$f(U,V) = \|AV' - b\|_2^2$$

So the original problem is formulated as a lease squares problem.

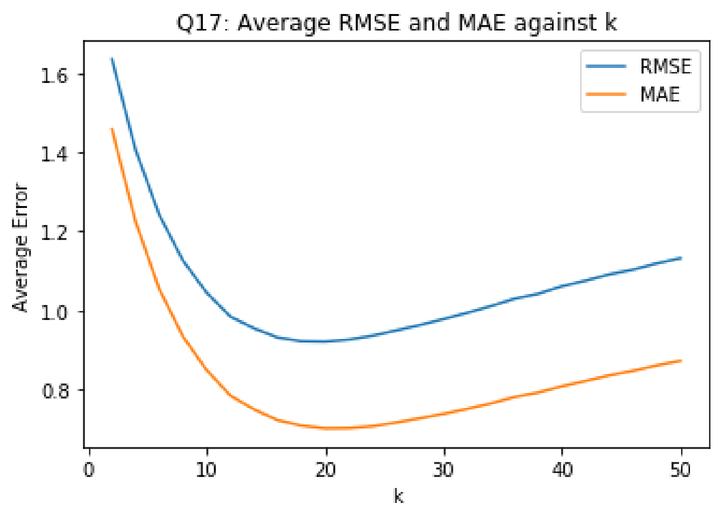

Figure 13: Average RMSE and average MAE against k

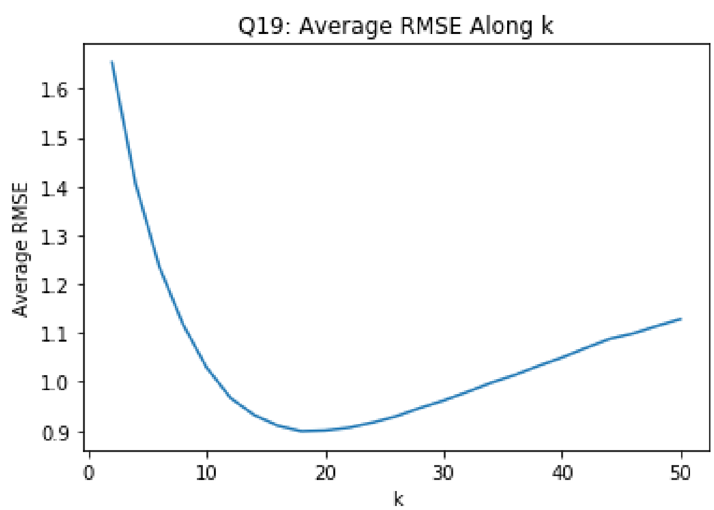

Figure 14: Average RMSE of popular movie trimmed test set

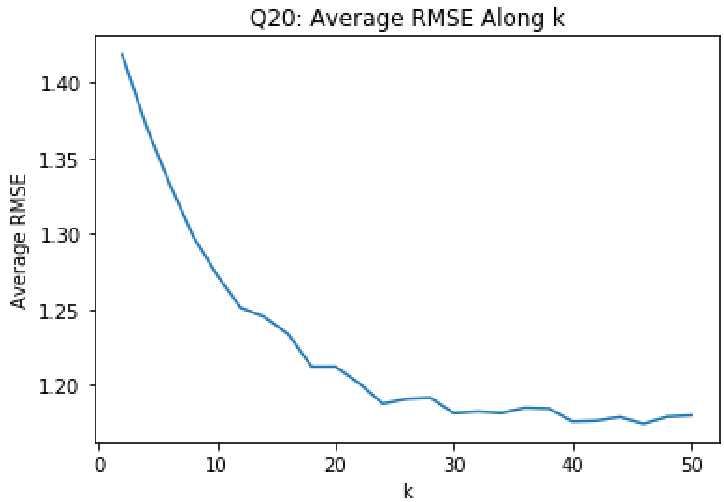

Figure 15: Average RMSE of unpopular movie trimmed test set

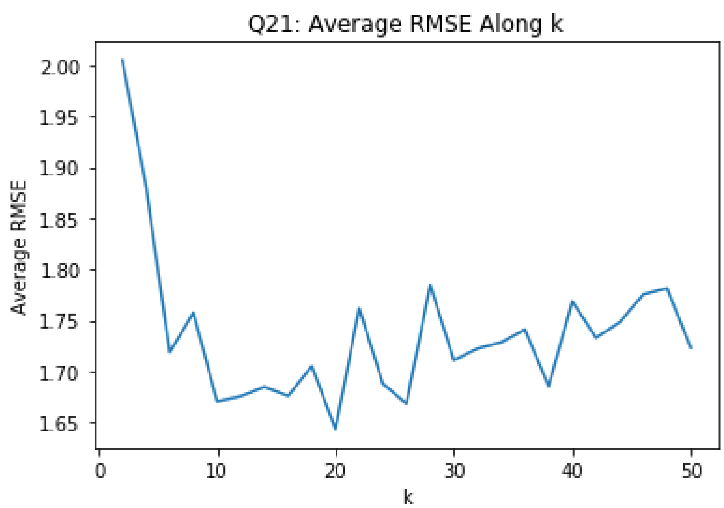

Figure 16: Average RMSE of high variance movie trimmed test set

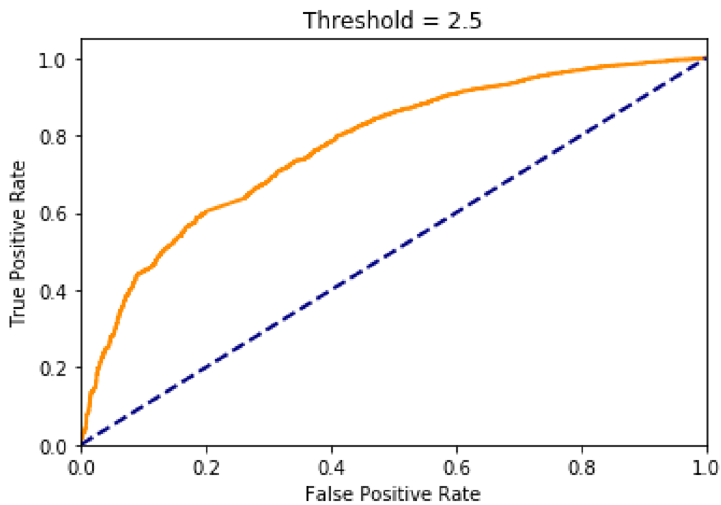

Figure 17: ROC curve for the NNMF collaborative filter, threshold=2.5

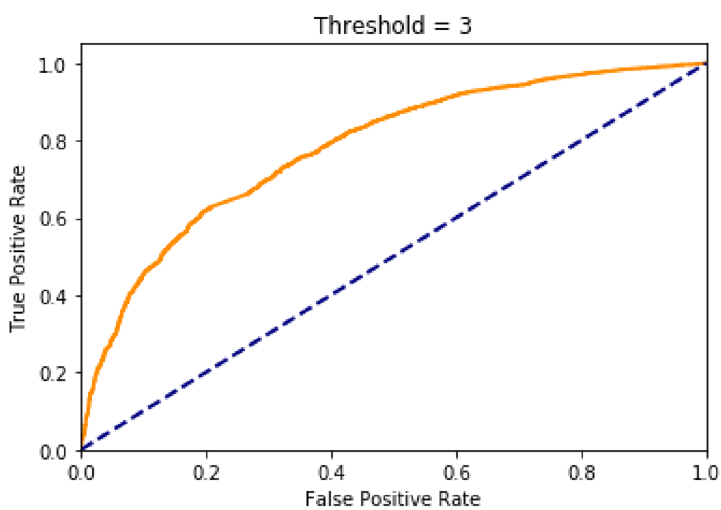

Figure 18: ROC curve for the NNMF collaborative filter, threshold=3

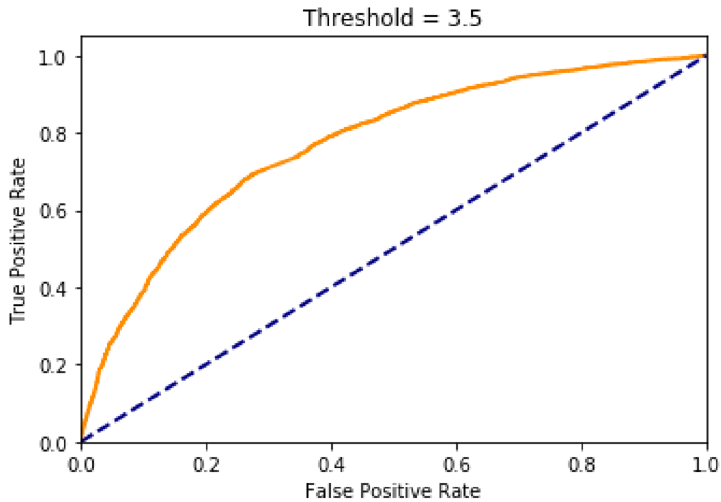

Figure 19: ROC curve for the NNMF collaborative filter, threshold=3.5

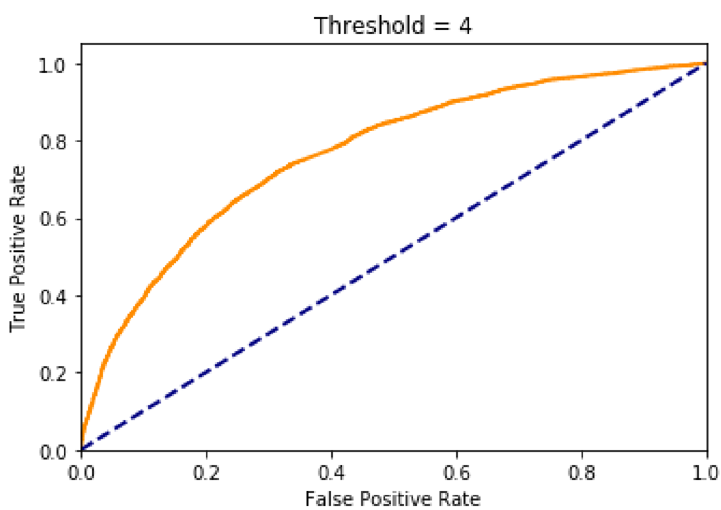

Figure 20: ROC curve for the NNMF collaborative filter, threshold=4

Table 2: AUC values of different thresholds

threshold	AUC
2.5	0.774
3	0.781
3.5	0.770
4	0.766

Question 23: Several columns are reported as following in table 3, 4, 5, 6, 7. As can been seen from the tables, most of the movies belong to one particular collection of genres: comedy and drama. Also, there is connection between latent factors and movie genres. Since the rows and columns of the rating matrix are co-related, and high rating movies are usually people's most favorite movies. As a result, after sorting the latent matrix, we find that comedy, drama and romance are the three most welcomed genres. In the matrix, these are also three genres that rating the highest.

Table 3: Top 10 movie for column 1

movieID	genres
3461	Adventure—Drama—Thriller
6201	Drama—Romance
5027	Action—Comedy—Crime—Drama—Thriller
74282	Children—Drama—Romance
2506	Comedy—Drama—Romance
177765	Adventure—Animation—Children
49272	Action—Adventure—Thriller
165483	Comedy
53161	Comedy—Drama—Romance—Sci-Fi
4178	Drama

Table 4: Top 10 movie for column 2

movieID	genres
2724	Comedy—Romance
100159	Comedy
26547	Action—Comedy—Crime—Thriller
62250	Crime—Drama
86781	Drama—Mystery—War
1643	Drama—Romance
72226	Adventure—Animation—Children—Comedy—Crime
71732	Comedy—Horror
184349	Comedy—Drama—Romance
414	Comedy

Table 5: Top 10 movie for column 3

movieID	genres
937	Comedy—Romance
4205	Comedy—Drama—Romance
61240	Drama—Fantasy—Horror—Romance
70183	Comedy—Drama—Romance
17	Drama—Romance
7932	Documentary
1356	Action—Adventure—Sci-Fi—Thriller
26777	Drama
5328	Drama—Romance
5849	Comedy

Table 6: Top 10 movie for column 4

movieID	genres
4619	Comedy
5016	Drama
183611	Action—Comedy—Crime—Horror
4849	Comedy—Drama
89745	Action—Adventure—Sci-Fi—IMAX
7299	Drama
56174	Action—Horror—Sci-Fi—Thriller—IMAX
114028	Comedy—Drama
136445	Comedy
4499	Comedy

3.1 Matrix factorization with bias (MF with bias)

3.1.1 Design and test via cross-validation

Question 24: The RMSE and MSE plots against k is in the figure 21. As seen in the figure, when k = 26, it reaches minimum average RMSE and MSE.

Question 25: The optimal number of latent factors k is 26.

3.1.2 MF with bias filter performance on trimmed test set

Question 26: The average RMSE against k is in figure 22. The minimum average RMSE is 0.857.

Question 27: The average RMSE against k is in figure 23. The minimum average RMSE is 0.971.

Question 28: The average RMSE against k is in figure 24. The minimum average RMSE is 1.435.

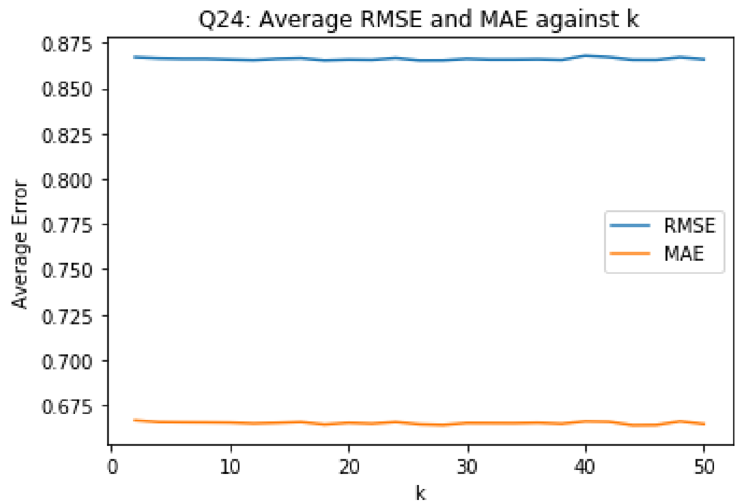

Figure 21: Average RMSE and average MAE against k

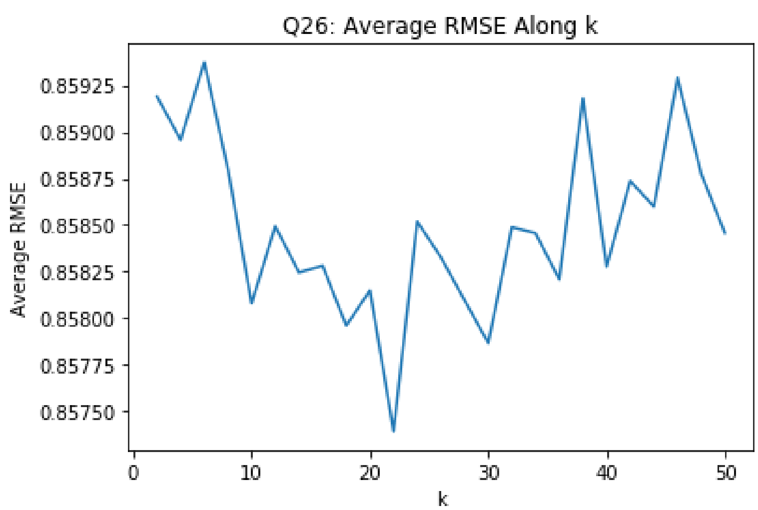

Figure 22: Average RMSE against k

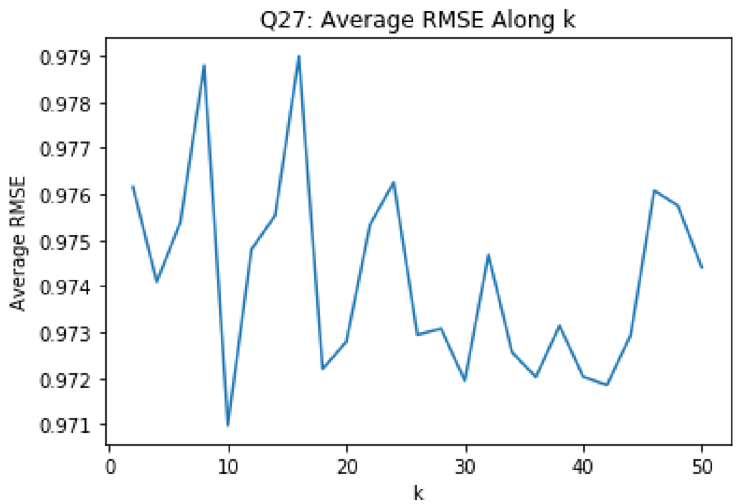

Figure 23: Average RMSE against k

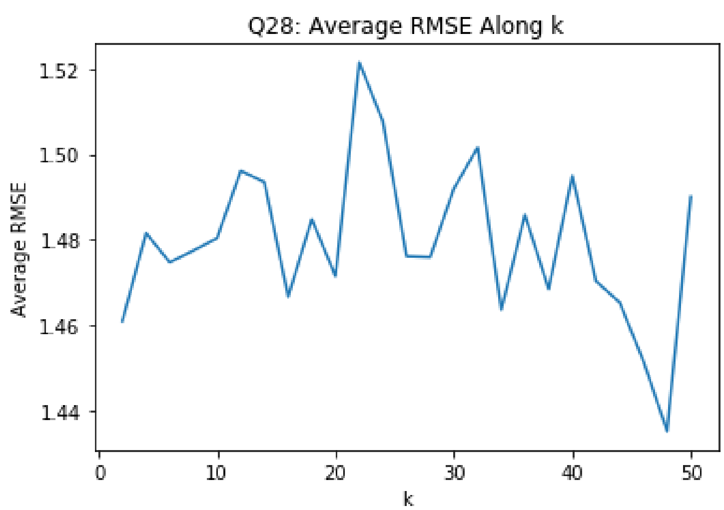

Figure 24: Average RMSE against k

Table 7: Top 10 movie for column 5

movieID	genres
27373	Drama
170355	Drama—Mystery—Romance
103810	Action—Comedy—Crime—Thriller
51834	Drama—Romance
3005	Thriller
32582	Documentary
84553	Comedy
2921	Western
80126	Drama—Thriller
52694	Comedy

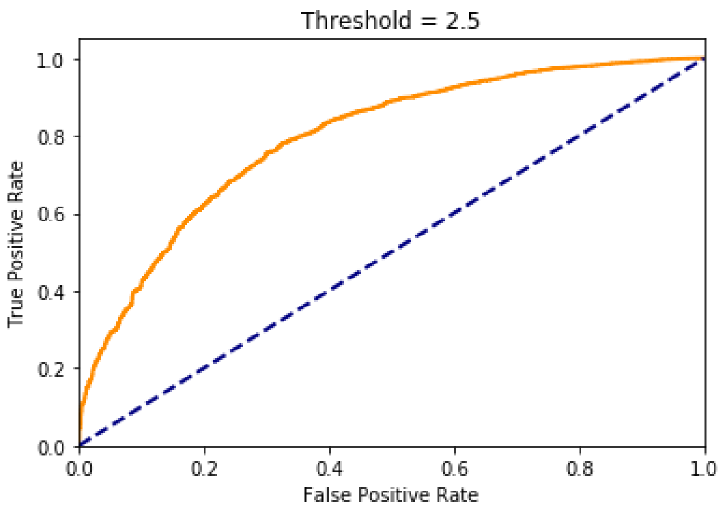

Figure 25: ROC curve for the MF collaborative filter with bias, threshold=2.5

3.1.3 Performance evaluation using ROC curve

Question 29: The plots are as in Figure 25, 26, 27, 28.
And the corresponding Area under Curve is in the table 8.

4 Naive collaborative filtering

4.1 Design and test via cross-validation

Question 30:
We designed a naive collaborative filter to predict the ratings of the movies and the average RMSE by averaging the RMSE across all 10 folds is 0.9347.

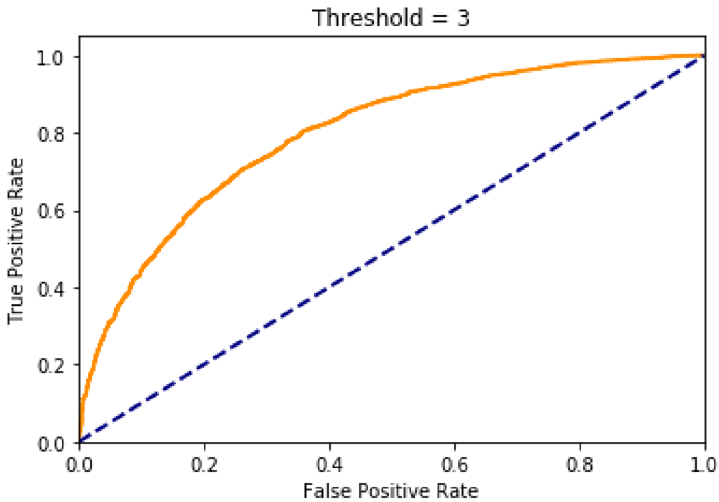

Figure 26: ROC curve for the MF collaborative filter with bias, threshold=3

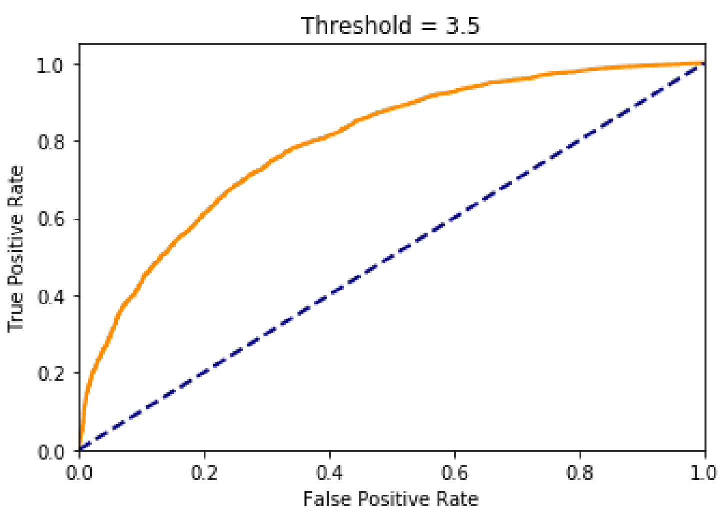

Figure 27: ROC curve for the MF collaborative filter with bias, threshold=3.5

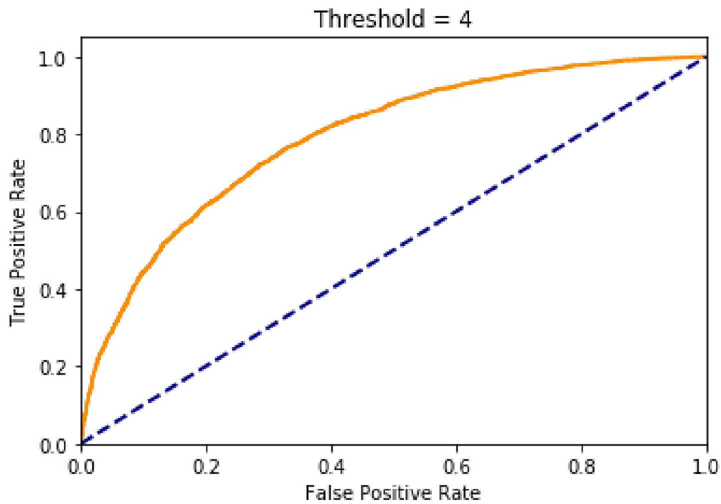

Figure 28: ROC curve for the MF collaborative filter with bias, threshold=4

Table 8: AUC values of different thresholds

threshold	AUC
2.5	0.794
3	0.797
3.5	0.792
4	0.791

4.2 Naive collaborative filter performance on trimmed test set

Question 31:

Now we will test the performance of the filter in predicting the ratings of the movies in the trimmed test set. The average RMSE for popular movies is 0.9323. The accuracy is a little bit higher than original dataset, which can be explained as there's more similarities in popular movies (that's why it's called popular movie).

Question 32:

The average RMSE for unpopular movies is 0.9323. The accuracy is a little bit lower than original dataset. As far as I am concerned, it's the same reason as the popular movies.

Question 33:

The average RMSE for high variance movies is 1.4716. The performance for high variance movies is even worse. But it's pretty obvious because it's

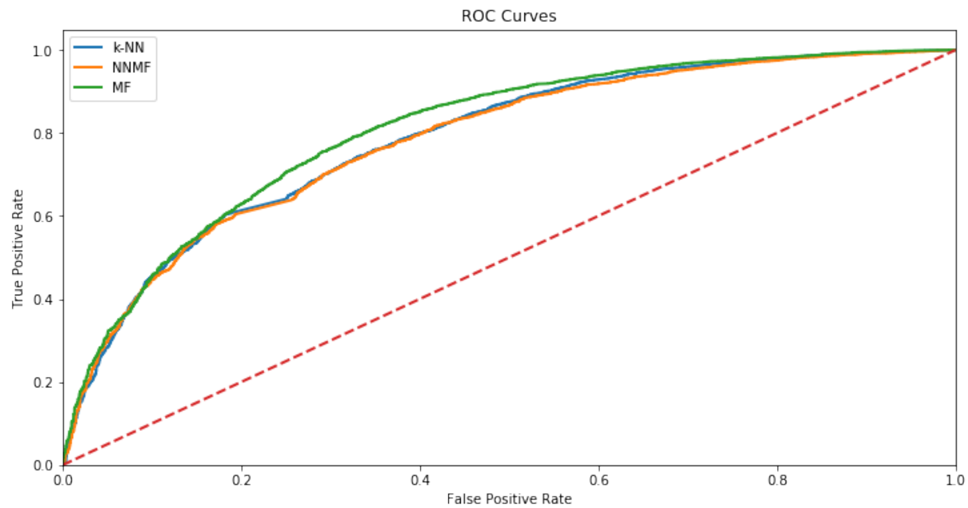

Figure 29: The ROC curves for the k-NN, NNMF, and MF

high variance, it's very hard to detect the similarities between those movies and predict very well.

5 Performance comparison

Question 34:

The ROC curves (threshold = 3) for the k-NN, NNMF, and MF with bias based collaborative filters is shown in Figure 29. We could see from the ROC curve that all those different methods give us pretty same results and it means that those model have similar performance. But if we have to choose the best one, MF collective filter seems like a little bit better.

6 Ranking

6.1 Evaluating ranking using precision-recall curve

Question 35:

Precision is the ratio of correctly predicted positive observations to the total predicted positive observations. High precision relates to the low false positive rate. In the recommendation system it represents the fraction of recommended items that are correct or liked by users.

On the other hand, recall is the ratio of correctly predicted positive observations to the all observations in actual class. It's basically the fraction of liked items that come from recommendation system.

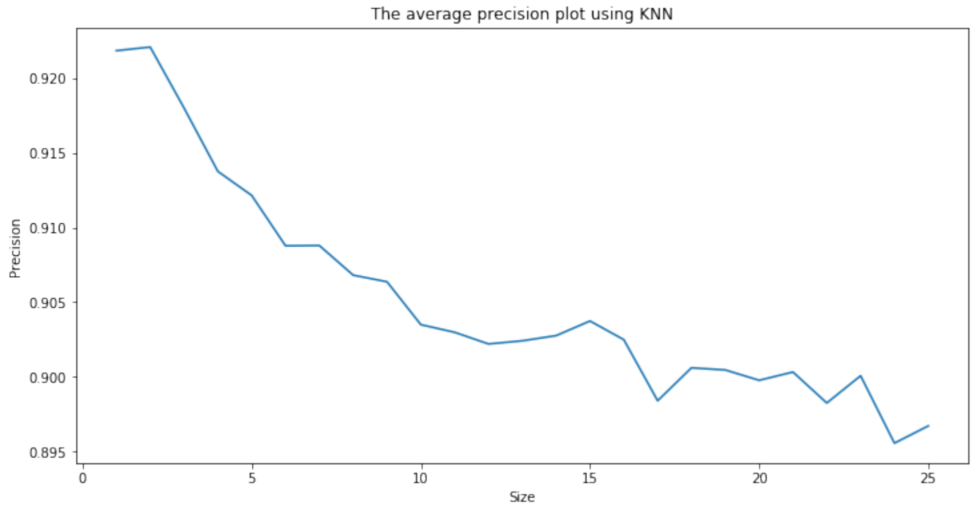

Figure 30: The average precision against size for the k-NN

Question 36:

The plot of average precision against size for the ranking obtained using k-NN collaborative filter predictions with $k = 22$ is shown in Figure 30. Also, the plot the average recall against size and average precision against average recall are Figure 31 and 32.

We can see from those plots that as the number of the recommendations increases, the precision becomes lower and lower and the recall becomes higher and higher on the contrary. This phenomenon can be also observed from the plot of the precision against recall, which is pretty obvious that if you recommend more and more, you are more likely to recommend what users like but you can not guarantee the precision of your recommendation at the same time. It's a tradeoff.

Question 37:

Similarly, the plot of average precision against size for the ranking obtained using NNMF collaborative filter predictions with $k = 20$ is shown in Figure 33. Also, the plot the average recall against size and average precision against average recall are Figure 34 and 35.

The shape of the plots follow the same trends that we discuss previously. As the number of the recommendations increases, the precision goes down but the recall goes up.

Question 38:

Same here, the plot of average precision against size for the ranking obtained using MF collaborative filter predictions with $k = 26$ is shown in Figure 36. Also,

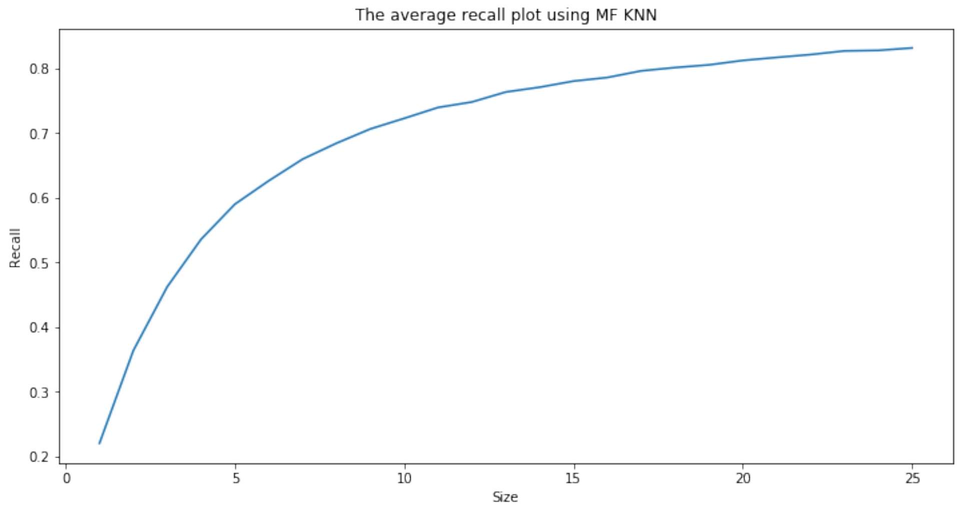

Figure 31: The average recall against size for the k-NN

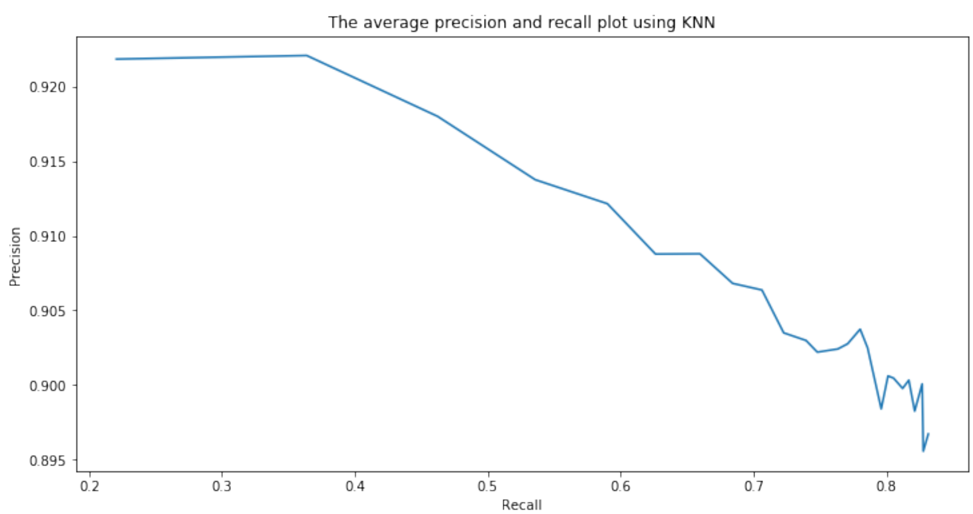

Figure 32: The average precision against recall for the k-NN

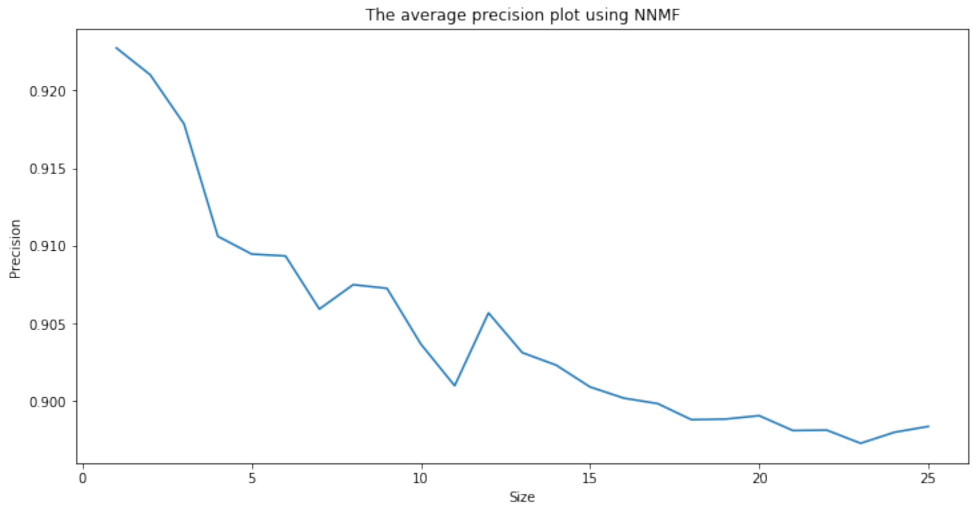

Figure 33: The average precision against size for the NNMF

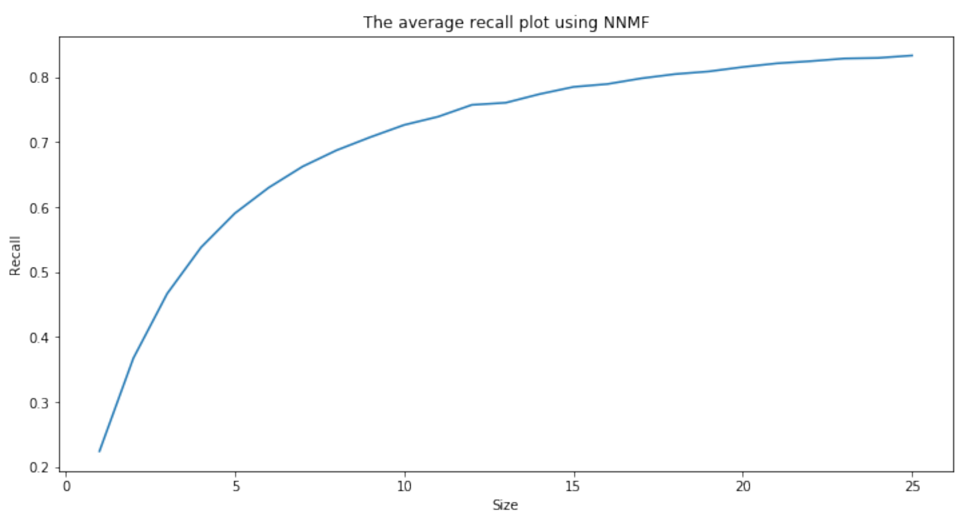

Figure 34: The average recall against size for the NNMF

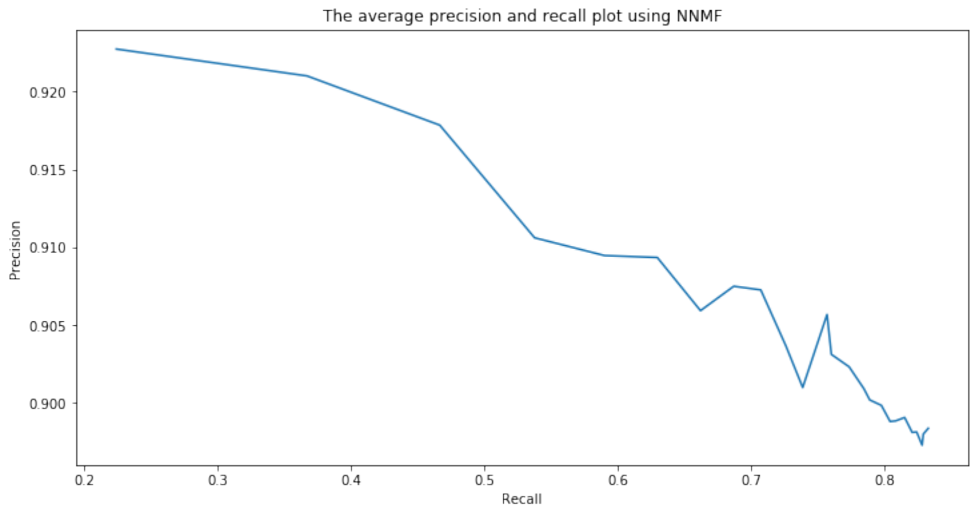

Figure 35: The average precision against recall for the NNMF

the plot the average recall against size and average precision against average recall are Figure 37 and 38.

The shape of the plots are basically still the same.follow the same trends that we discuss previously. The precision and recall go to the contradictory direction when the size growing.

Question 39:

The precision-recall curves for the k-NN, NNMF, and MF with bias based collaborative filters is shown in Figure 39. We could see that in comparison to other two methods, MF collective filter got the best performance with higher precision and recall. But overall all three methods give us similar performance.

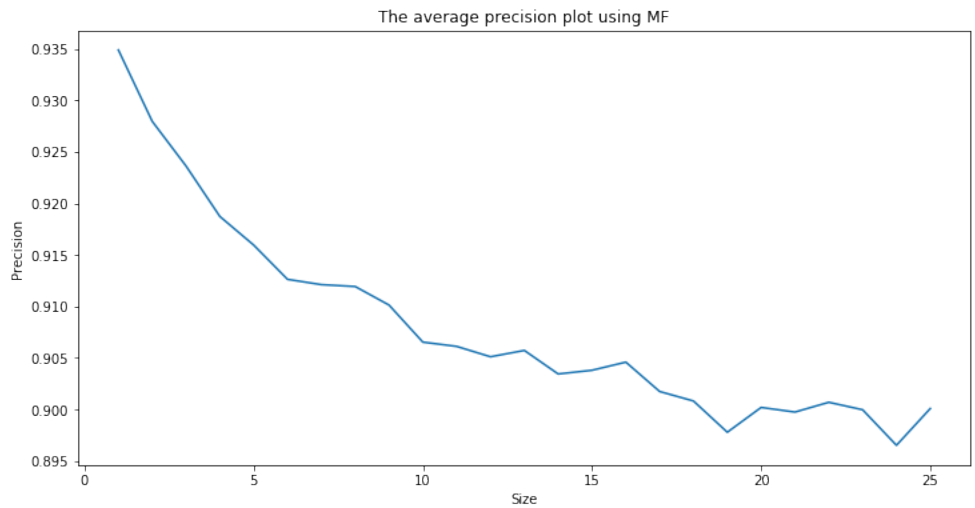

Figure 36: The average precision against size for the MF

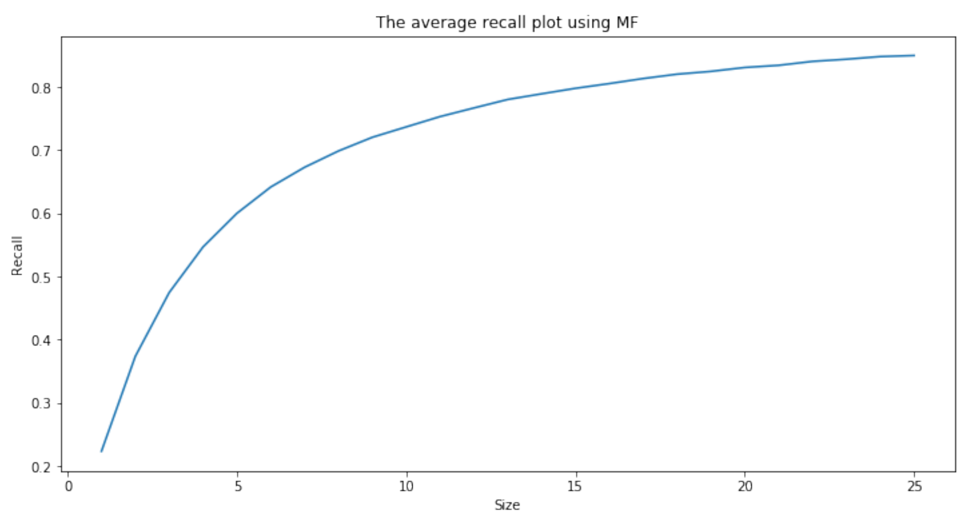

Figure 37: The average recall against size for the MF

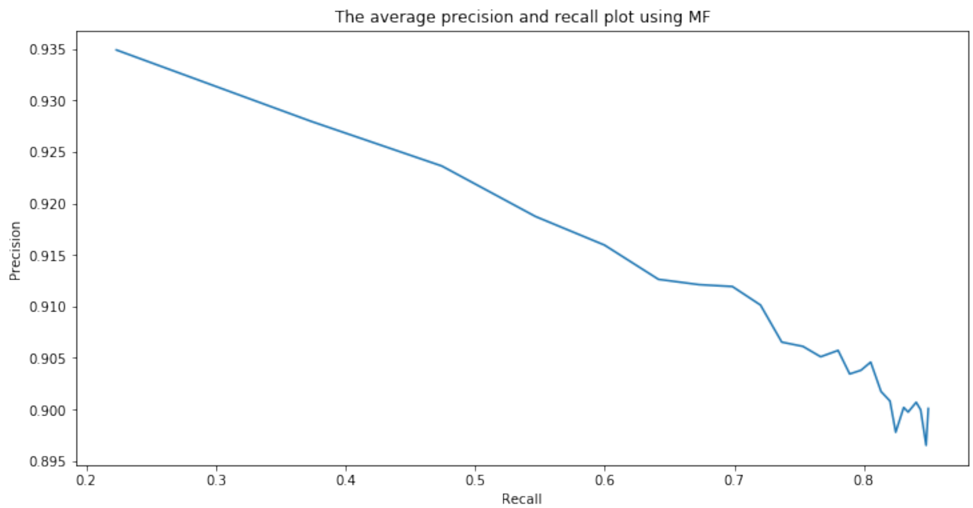

Figure 38: The average precision against recall for the MF

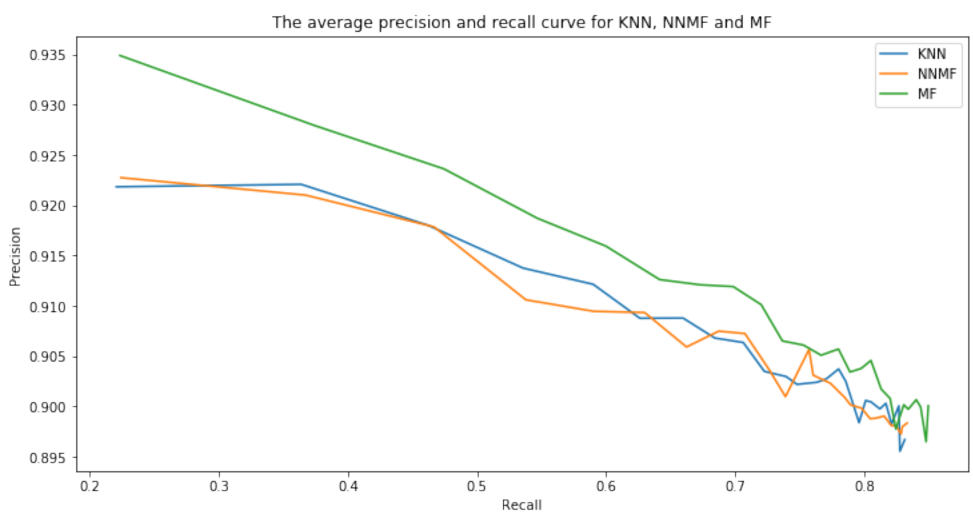

Figure 39: The precision-recall curves for the k-NN, NNMF, and MF