**Kushagra Rastogi**

**304640248**

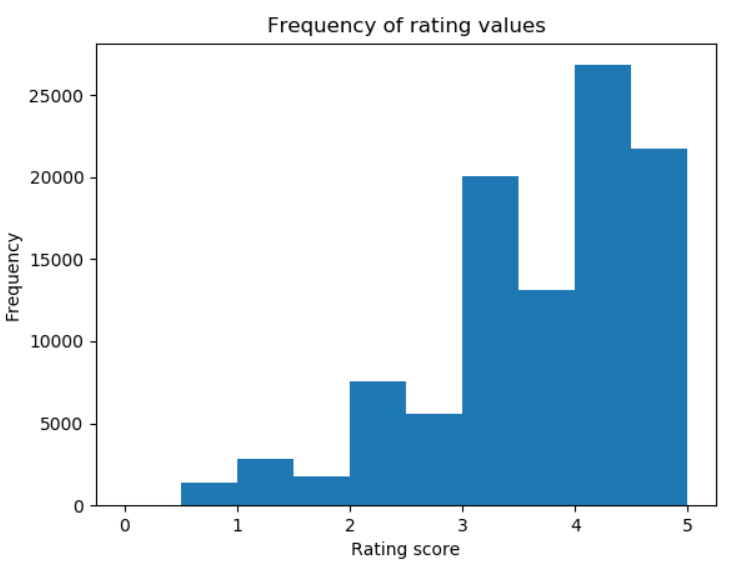
**ECE 219**

**Project 3: Collaborative Filtering**

**QUESTION 1:**

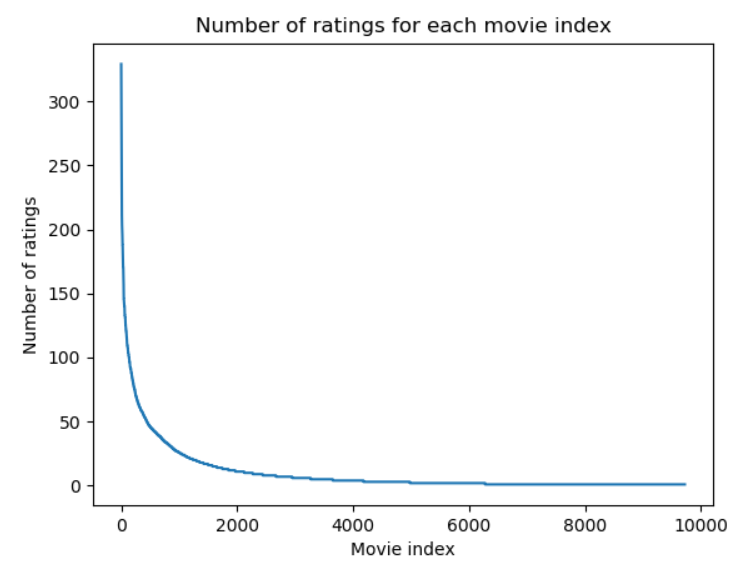


**QUESTION 2:**

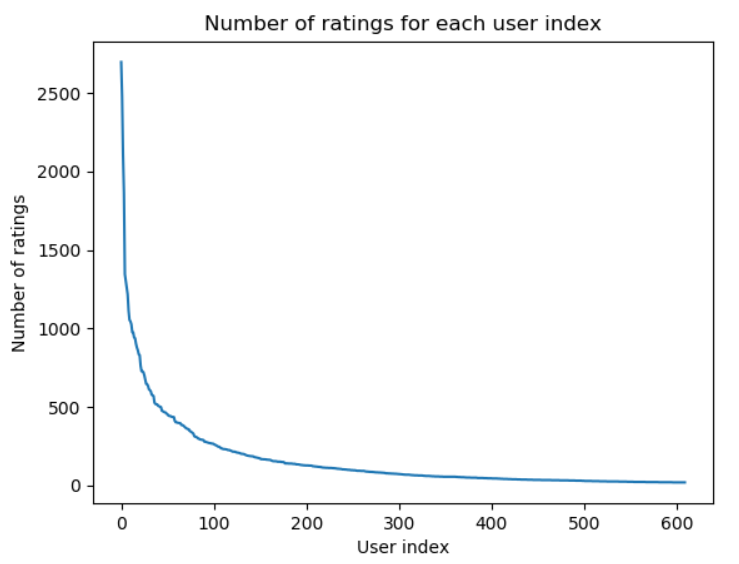


The histogram has an upward growth trend. This is indicative of the fact that more movies were rated highly than lowly. Most users gave ratings between 3 and 5 suggesting that a majority of the users liked the movies they watched.

**QUESTION 3:**



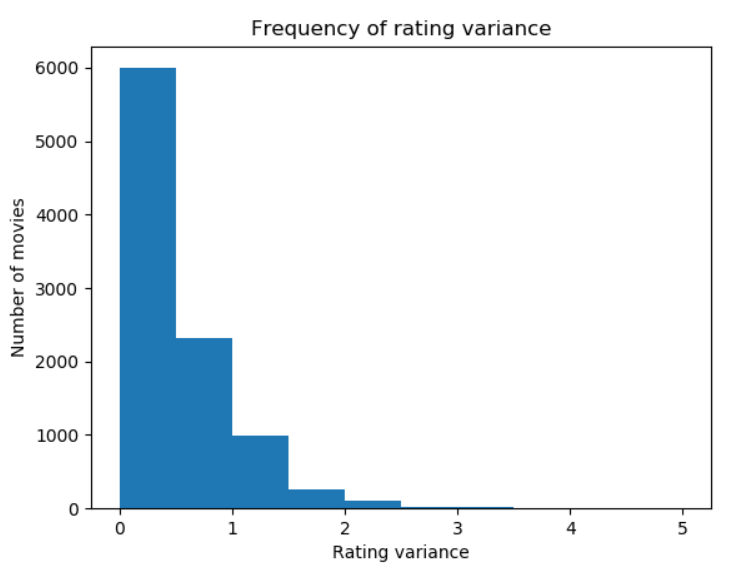
**QUESTION 4:**



**QUESTION 5:**

The number of ratings has a reciprocal relationship with the movie index. This means that a small number of movies received a majority of the ratings. This implies that a lot of movies received a very small number of ratings. Hence, this implies that the rating matrix R is sparse which means heavy regularization needs to be added to the recommendation process to prevent overfitting and false links.

**QUESTION 6:**



It can be seen from the histogram that the rating variance for most movies lies between 0 and 2. This means that most ratings are reliable and consistent.

**QUESTION 7:**

**QUESTION 8:**

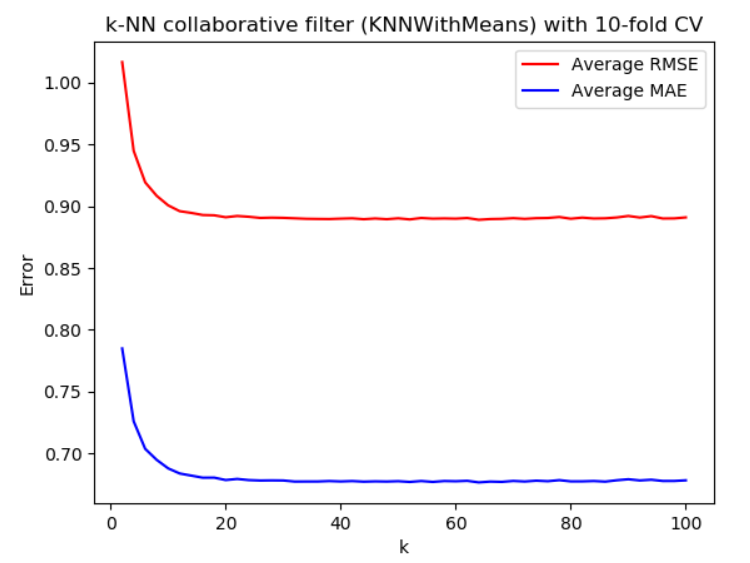
The quantity corresponds to movies that have been rated by both user and user . Since the rating matrix R is sparse, it is highly likely that because it is very likely that both users did not watch the same movie and/or did not rate the movie.

**QUESTION 9:**

Mean-centering the raw ratings in the prediction function helps reduce bias and remove extreme data points. For example, users who either rate all items highly or poorly are usually giving extreme opinions which is biased and can be considered noisy. Thus, we can make a more accurate prediction if we mean-center the ratings.

**QUESTION 10:**

I implemented the k-NN collaborative filter using KNNWithMeans from surprise python package.



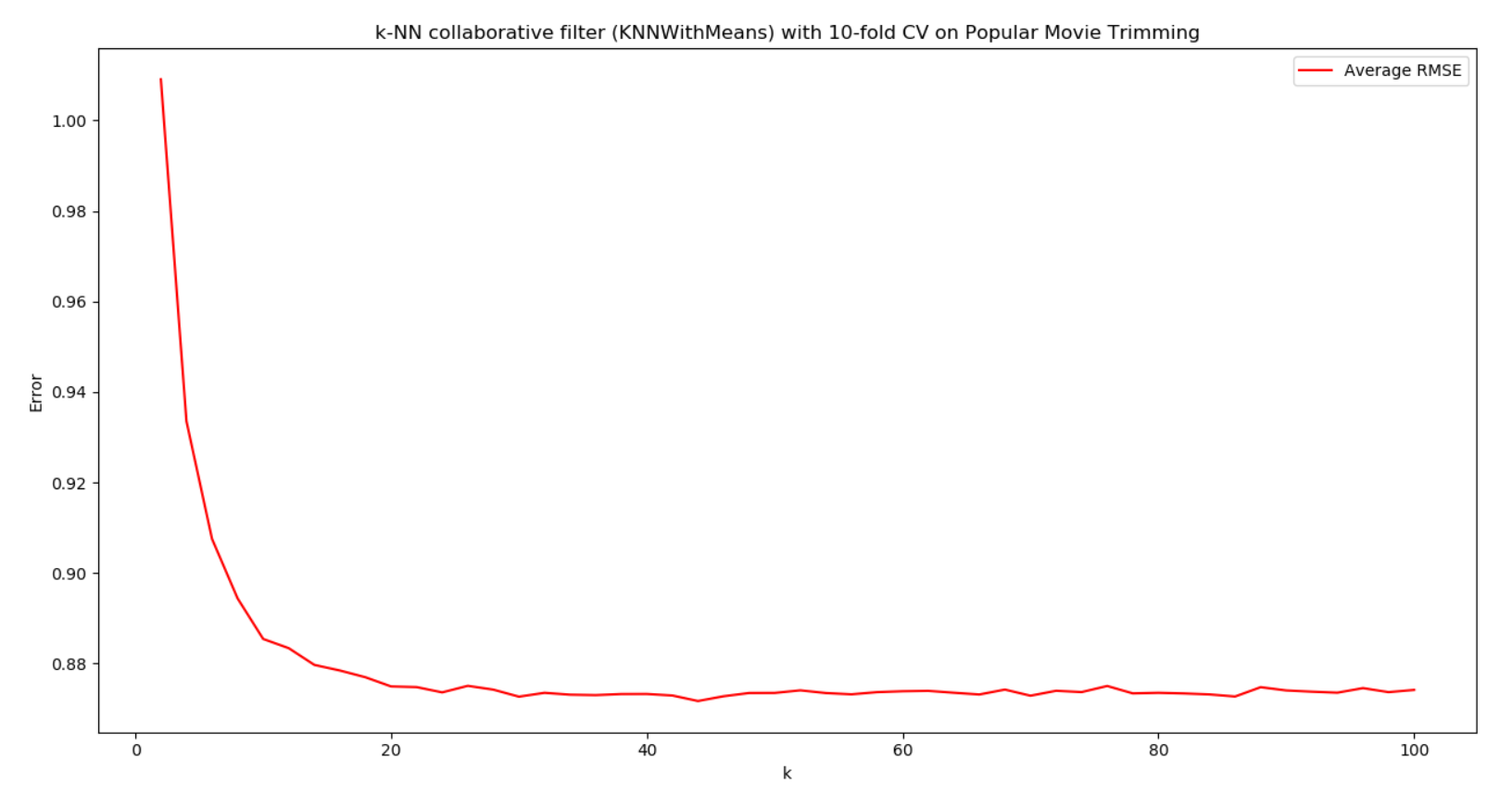
**QUESTION 11:**

The minimum is about 20 since both RMSE and MAE reach their respective steady state values at that value.

Steady-state RMSE = 0.8884

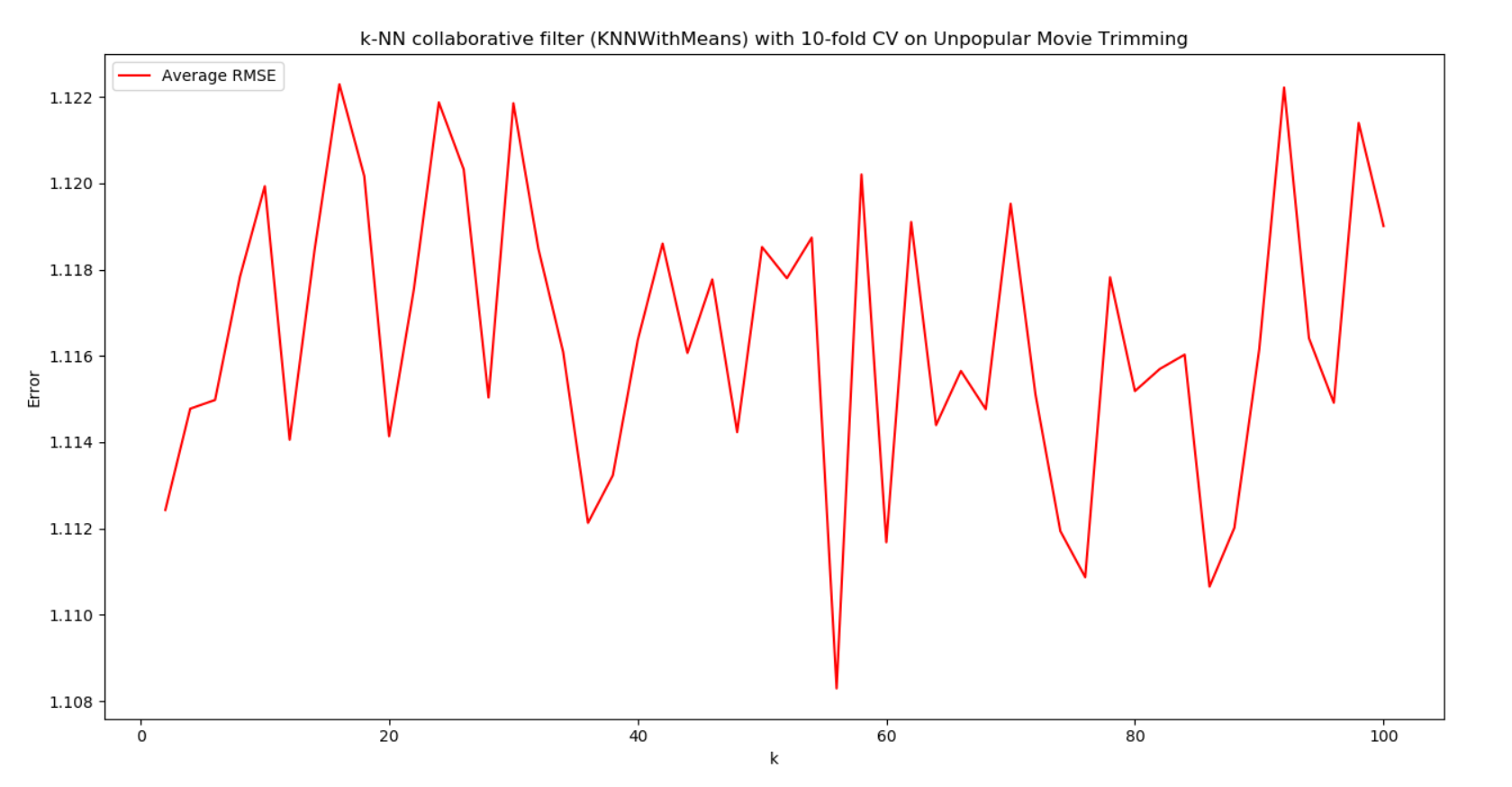
Steady-state MAE = 0.6781

**QUESTION 12:**



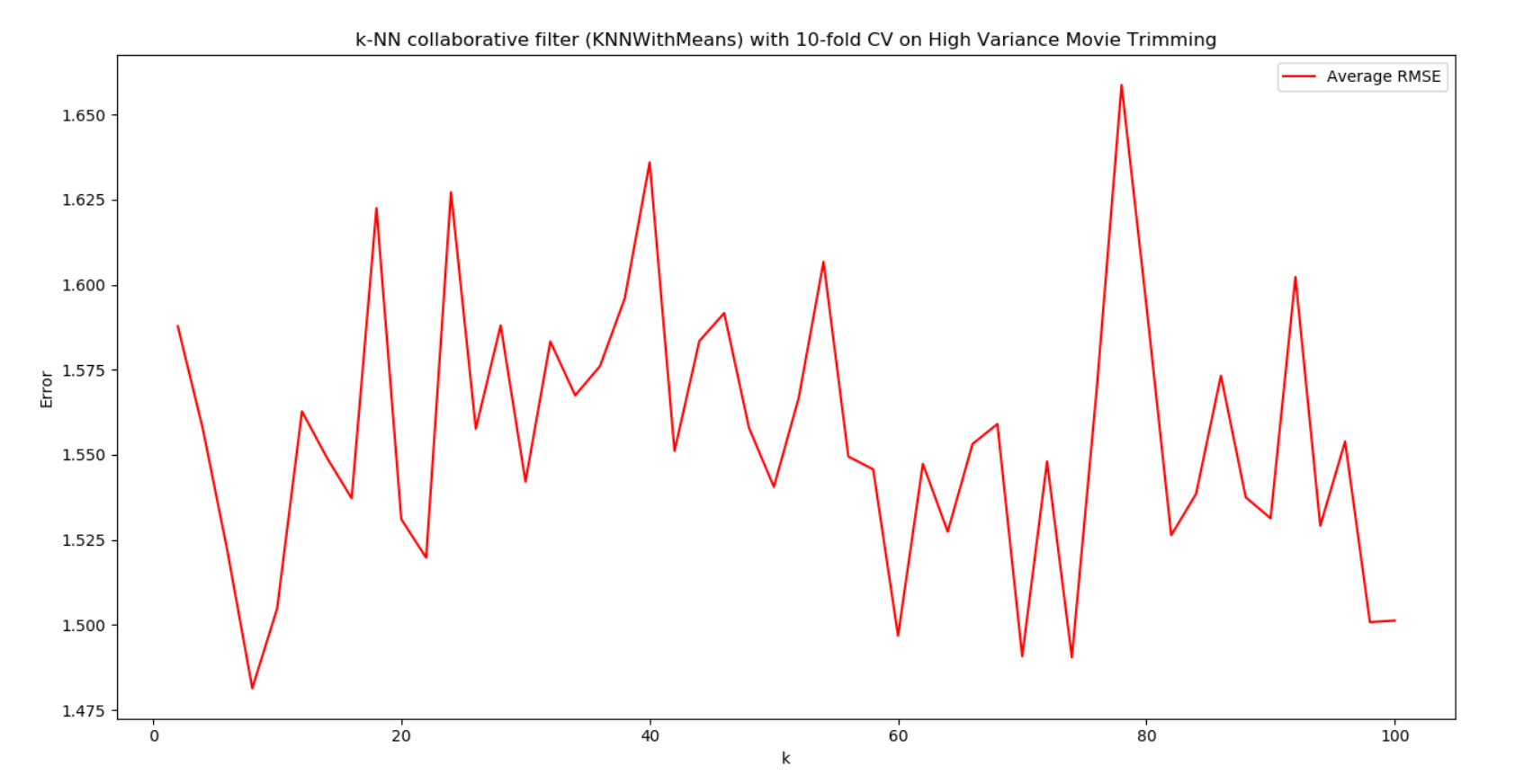


**QUESTION 13:**



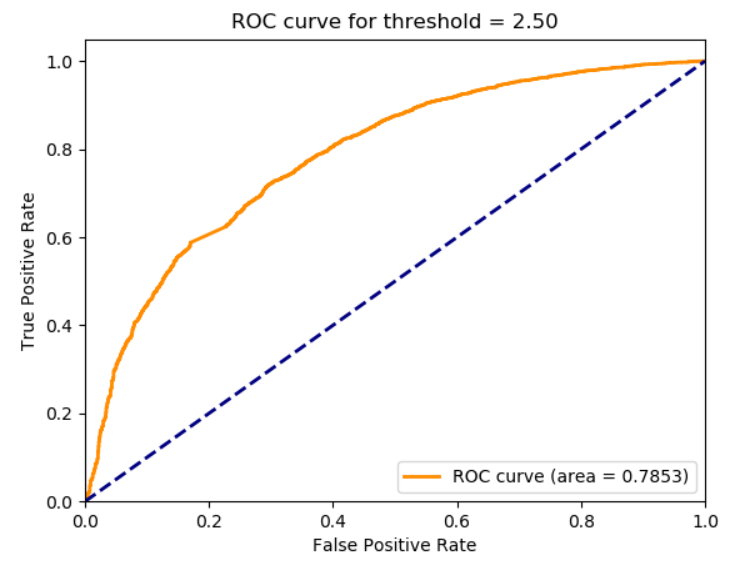


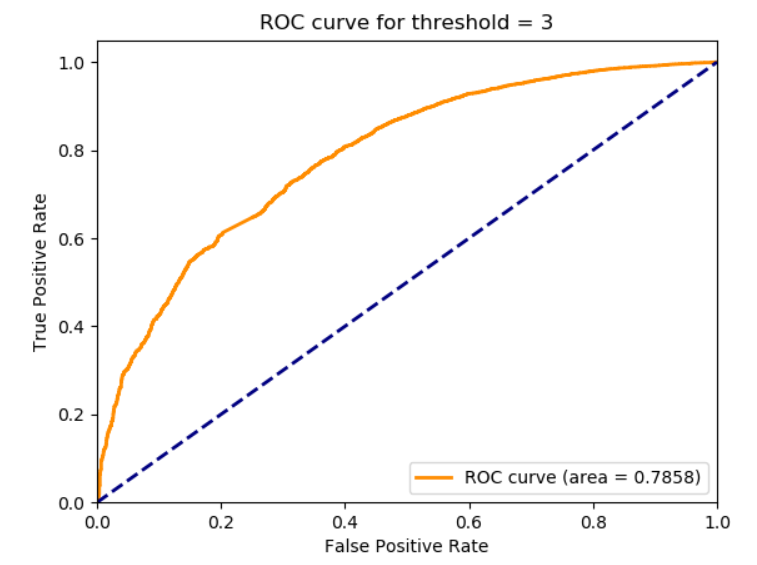
**QUESTION 14:**

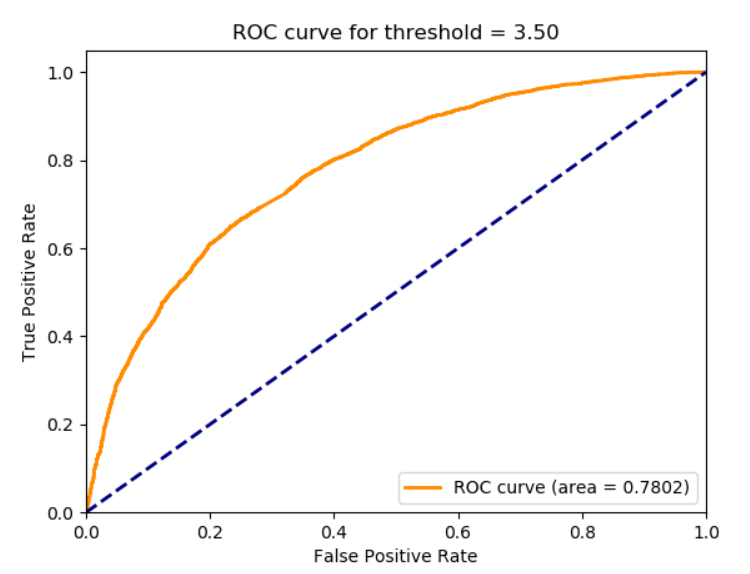


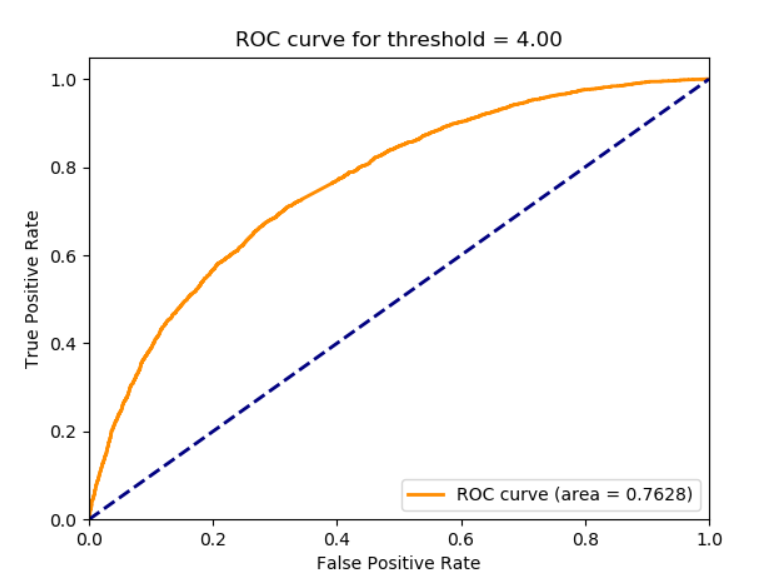


**QUESTION 15:**







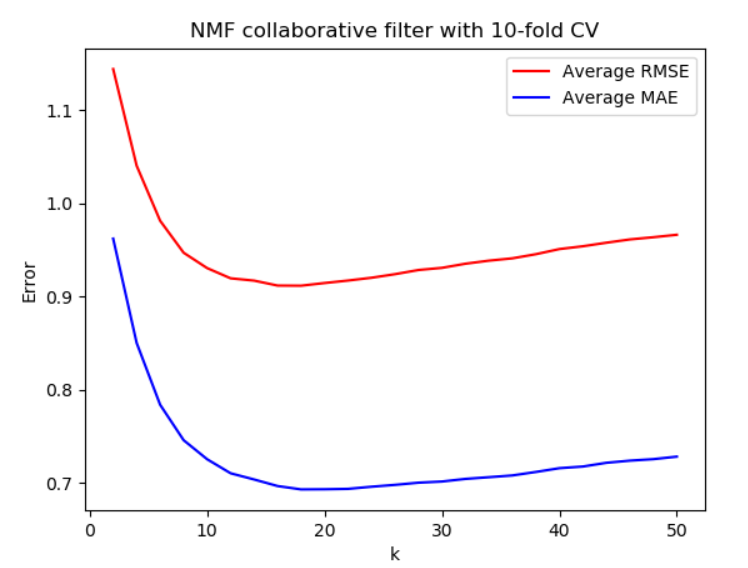


**QUESTION 16:**

The optimization problem given by equation 5 is not convex because of the matrix. For a fixed , the least-squares formulation is:

For a fixed , the least-squares formulation is:

**QUESTION 17:**

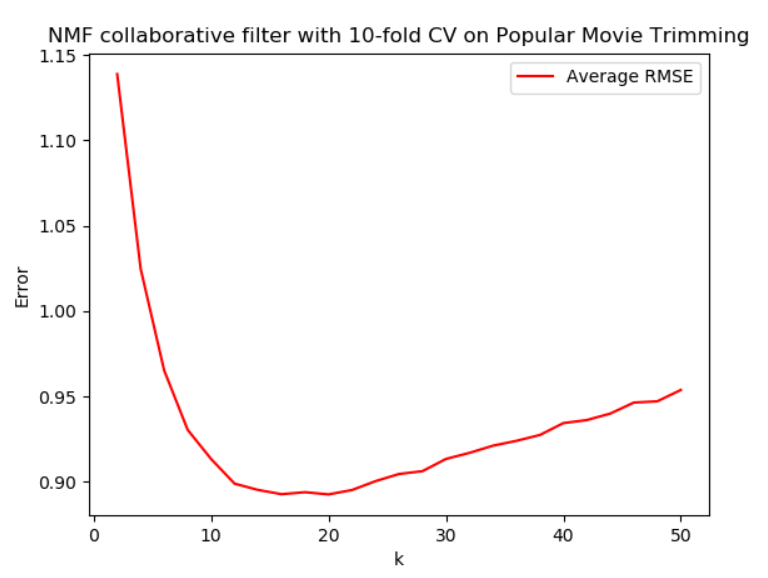


**QUESTION 18:**



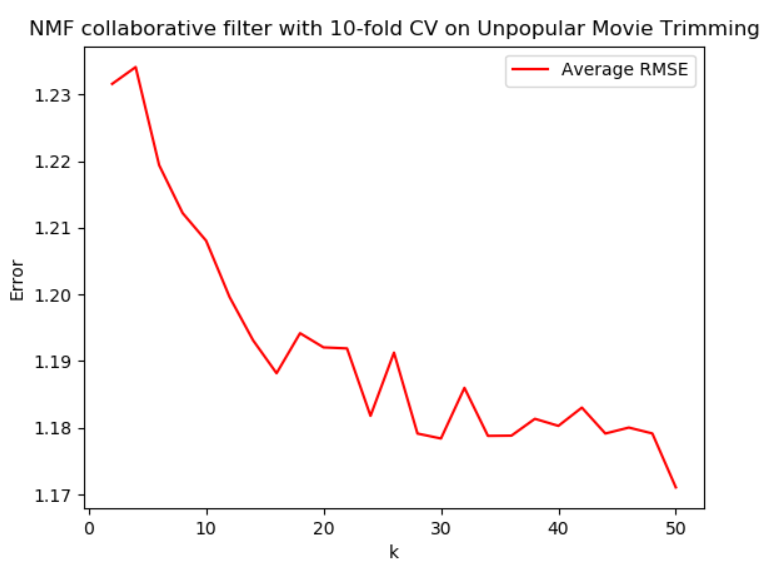
The optimal number of latent factors that gives minimum average RMSE is . The optimal number of latent factors that gives minimum average MAE is . The number of movie genres in the dataset is 19. Thus, the optimal number of latent factors is roughly about the same as the number of movie genres.

**QUESTION 19:**



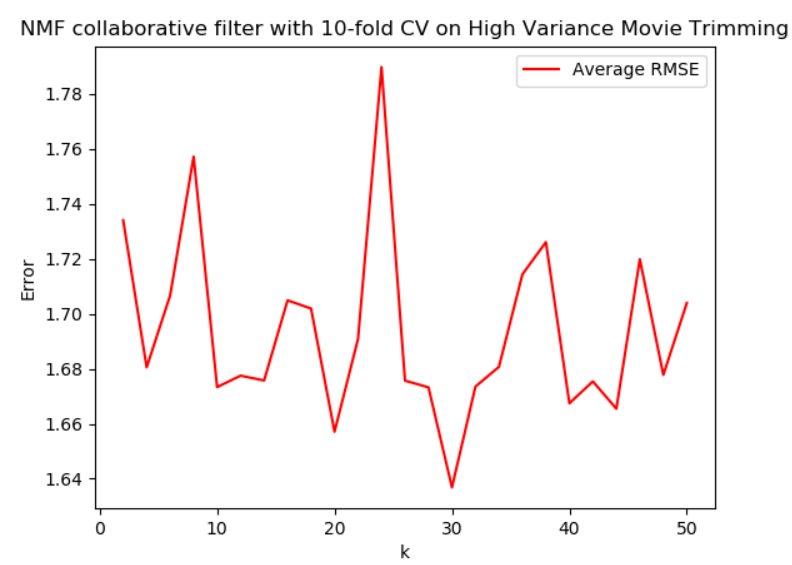


**QUESTION 20:**



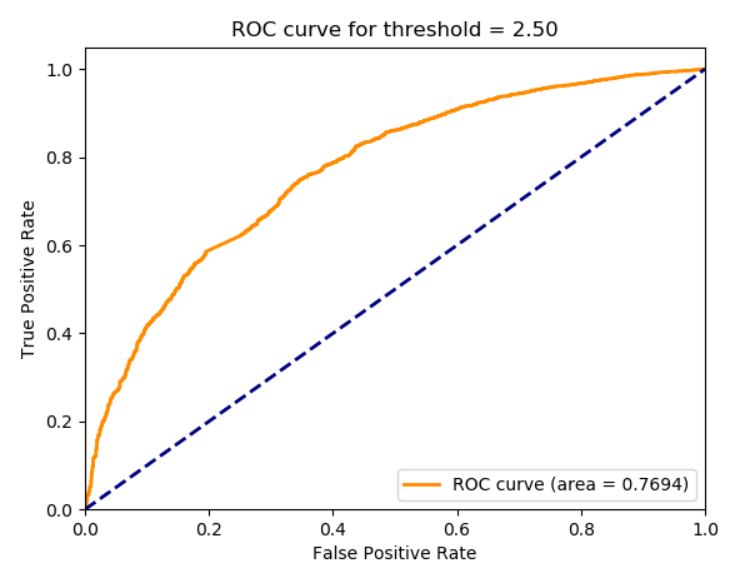


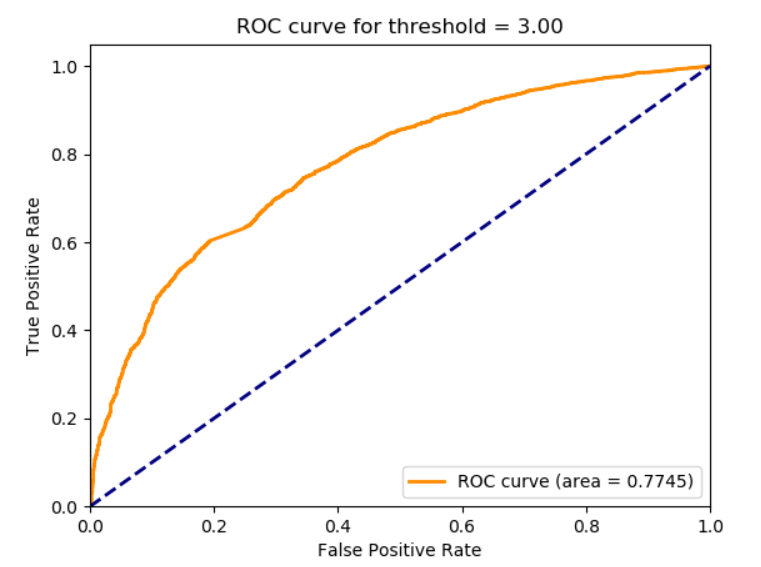
**QUESTION 21:**

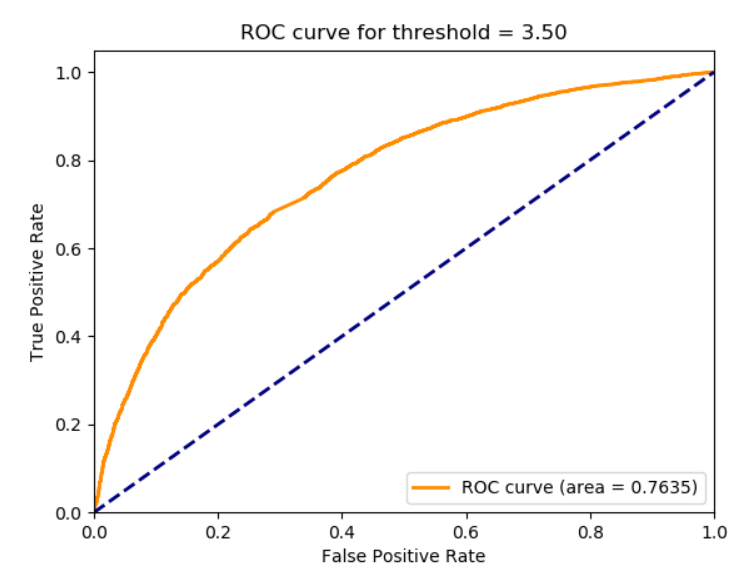


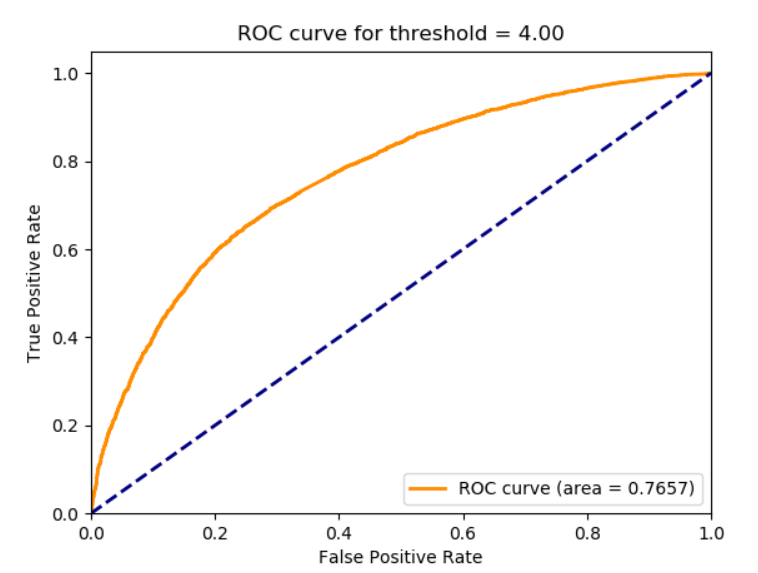


**QUESTION 22:**







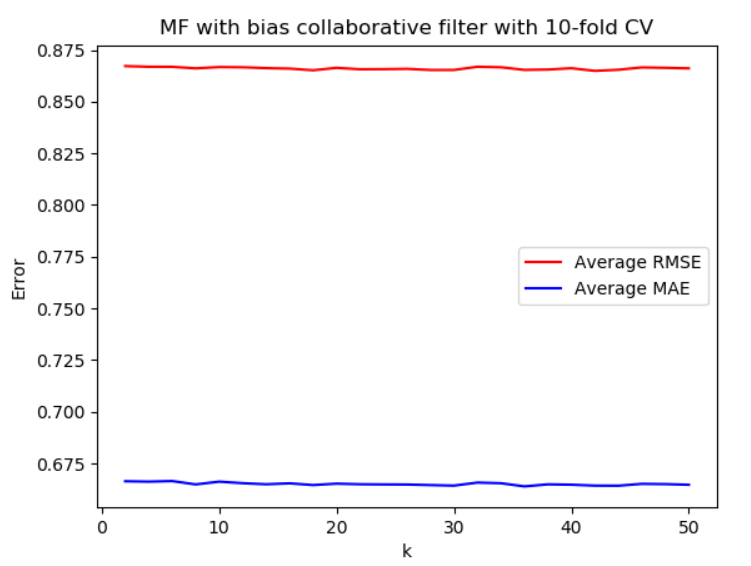


**QUESTION 23:**

|  |  |
| --- | --- |
| V[:, 0] | Comedy|Drama  Action|Crime  Comedy  Crime|Drama  Action|Crime|Thriller  Comedy|Romance  Adventure|Comedy|Western  Action|Comedy|Crime  Comedy|Drama|Romance  Drama |
| V[:, 5] | Sci-Fi  Comedy|Romance  Comedy|Drama  Comedy|Fantasy|Musical  Comedy|Drama  Documentary  Children|Drama  Action|Adventure|Drama|Fantasy|Thriller  Mystery|Thriller  Animation|Children |
| V[:, 10] | Comedy|Romance  Comedy  Drama|Horror  Drama|Romance  Comedy|Drama  Comedy|Drama|Romance  Action|Crime|Drama  Action|Adventure  Comedy|Drama|Romance  Documentary |
| V[:, 15] | Action|Adventure  Drama  Documentary  Musical|Romance|War  Drama  Action|Comedy  Comedy  Action|Horror|Thriller  Drama  Crime|Drama|Musical |
| V[:, 16] | Action|Adventure|Comedy|Fantasy|Horror|Thriller  Action|Comedy|Crime  Documentary|War  Action|Adventure|Drama  Action|Crime  Adventure|Animation|Comedy  Comedy|Romance  Comedy|Drama|Romance  Action|Comedy  Drama|Romance |
| V[:, 17] | Drama|Romance  Action|Adventure|Drama|Thriller  Horror|Sci-Fi|Thriller  Drama  Drama  Crime|Drama|Romance|Thriller  Adventure|Comedy|Drama  Comedy|War  Drama  Comedy |
| V[:, 18] | Comedy  Children|Comedy  Drama  Comedy|Fantasy|Romance  Comedy  Drama  Drama|Thriller  Drama|Romance  Comedy  Horror|Thriller |
| V[:, 19] | Adventure|Crime|Drama  Horror|Sci-Fi|Thriller  Comedy|Horror  Horror|Mystery|Thriller  Comedy  Mystery|Thriller  Comedy|Drama  Action|Drama|War  Animation|Children|Fantasy  Drama|Mystery |

The table shows the genres of the top 10 movies for a particular number of latent factors. For all the latent factors, the top 10 movies belong to a small collection of genres. As the number of latent factors increases, the number of distinct movie genres decreases. For example, when the number of latent factors = 5, the number of distinct movie genres = 13. When the number of latent factors = 18, the number of distinct movie genres = 7. Thus, movie genres are clustered more closely together as the number of latent factors increases.

**QUESTION 24:**



**QUESTION 25:**

The minimum average RMSE occurs at . The minimum average MAE occurs at . Thus, the number of optimal latent factors is 18.

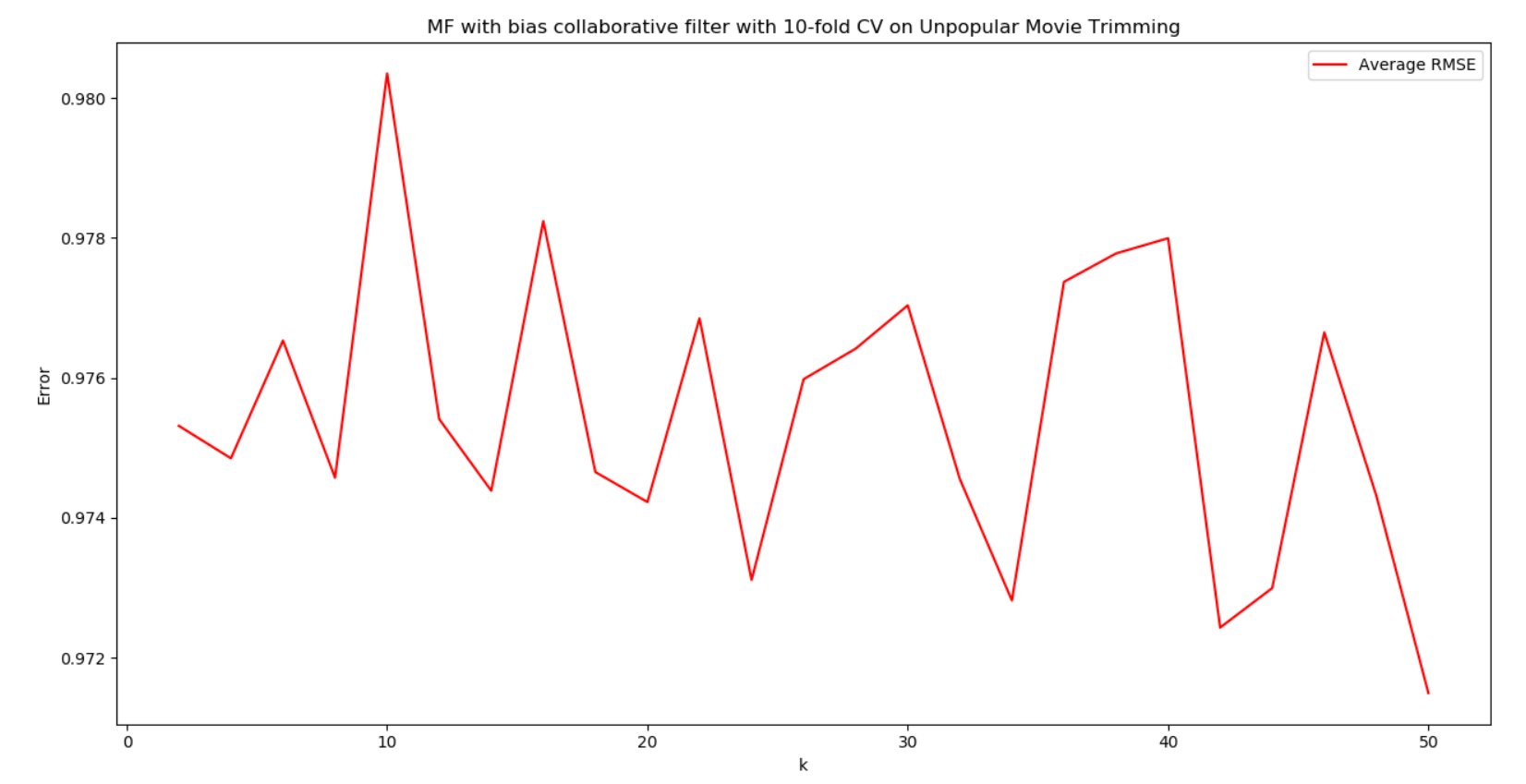


**QUESTION 26:**



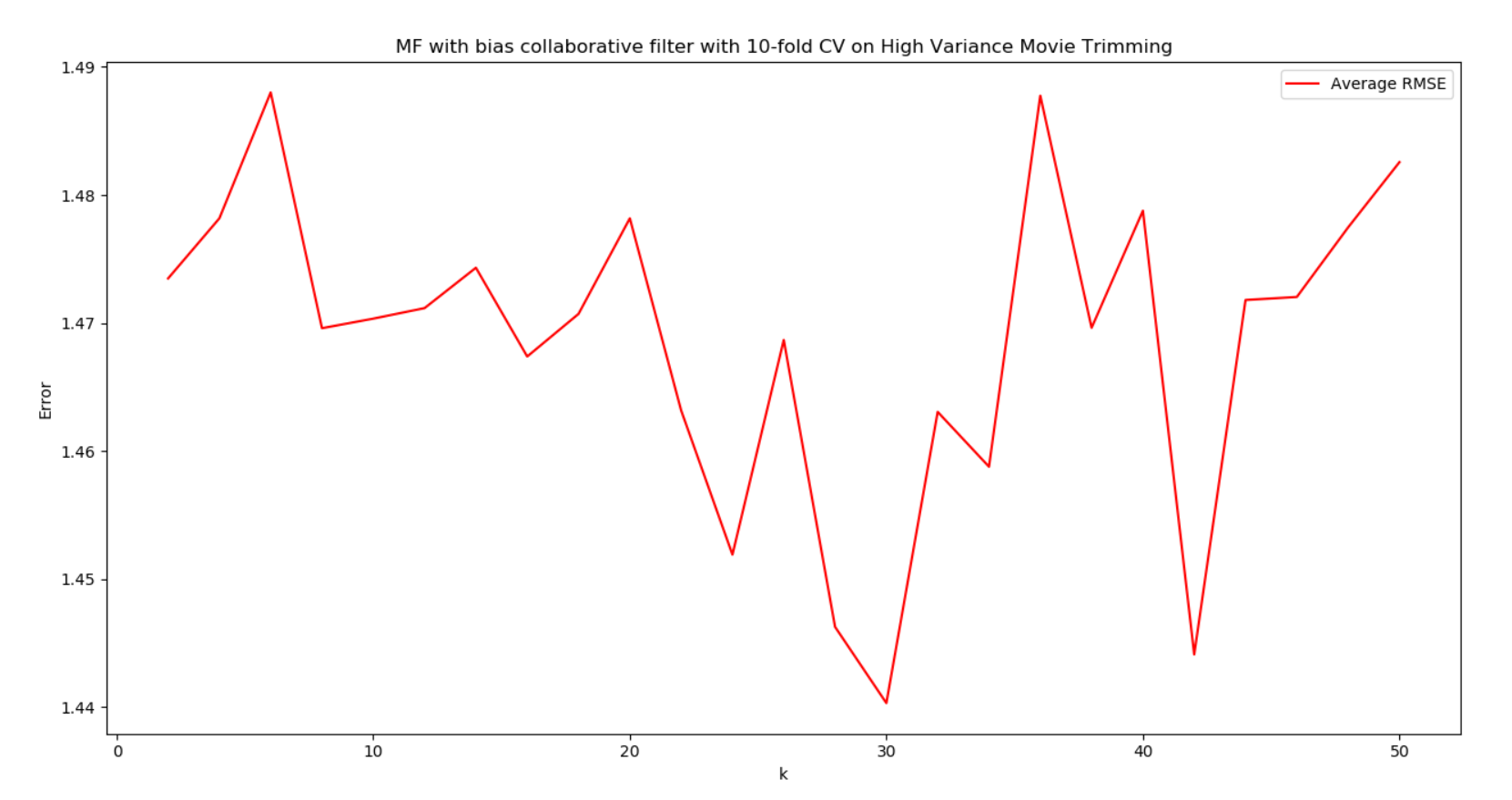


**QUESTION 27:**



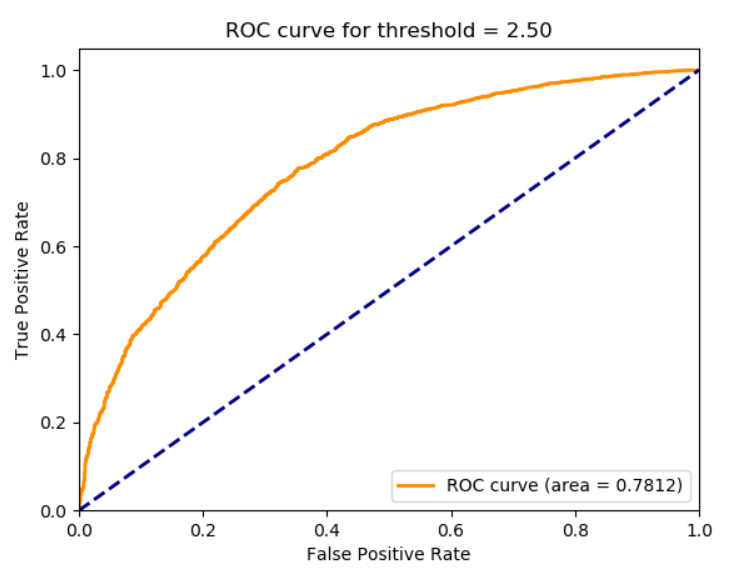


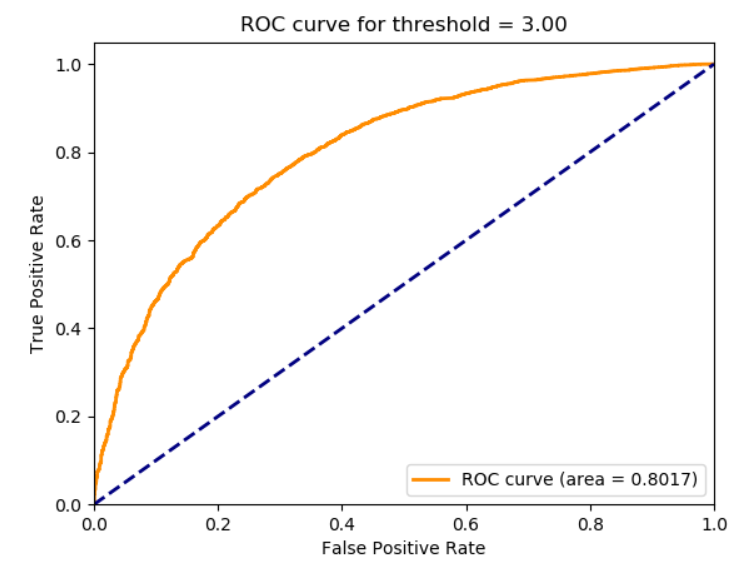
**QUESTION 28:**

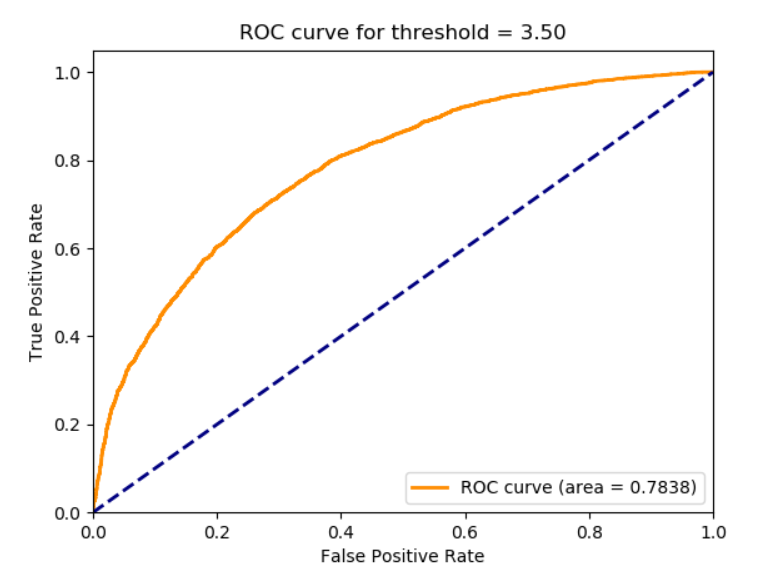


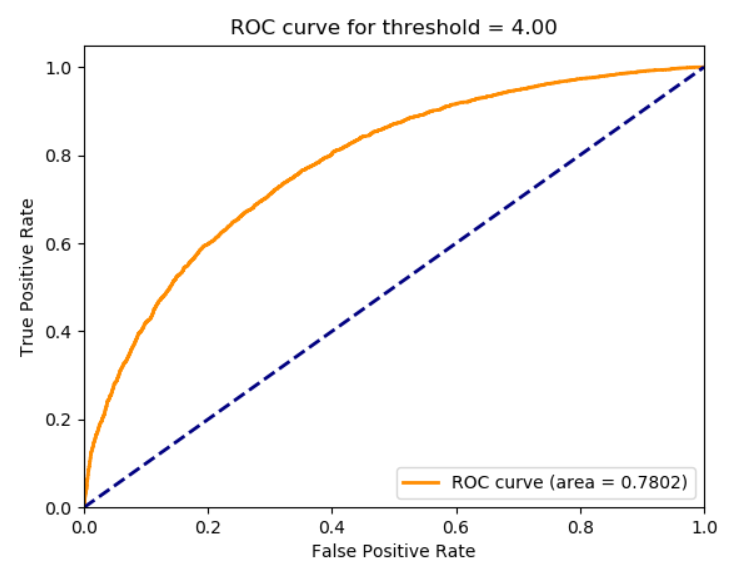


**QUESTION 29:**









**QUESTION 30:**



**QUESTION 31:**



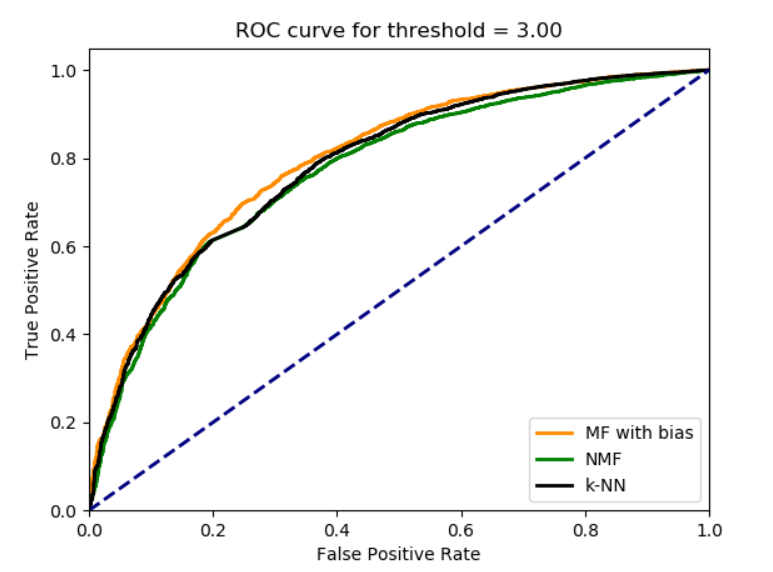
**QUESTION 32:**



**QUESTION 33:**



**QUESTION 34:**



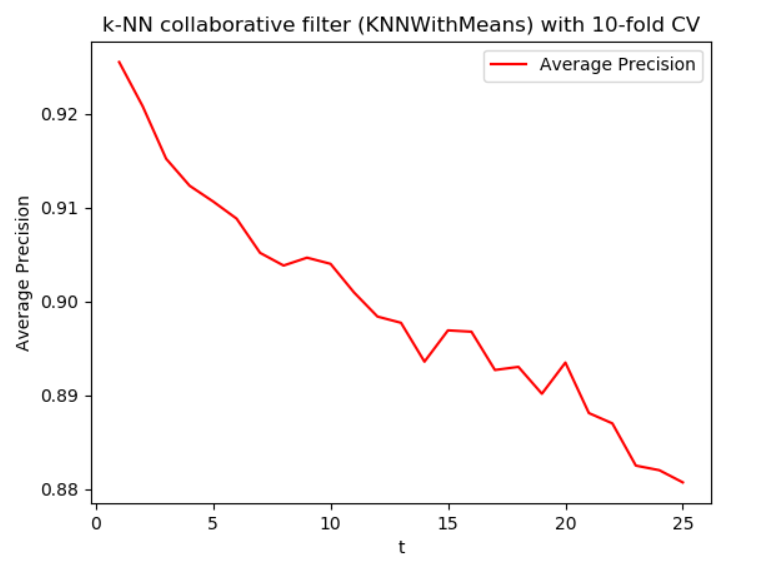
MF with bias seems to have the smoothest curve and has the largest area under the curve when threshold = 3. This means that MF with bias is the best among the three collaborative filters at predicting the ratings of the movies.

**QUESTION 35:**

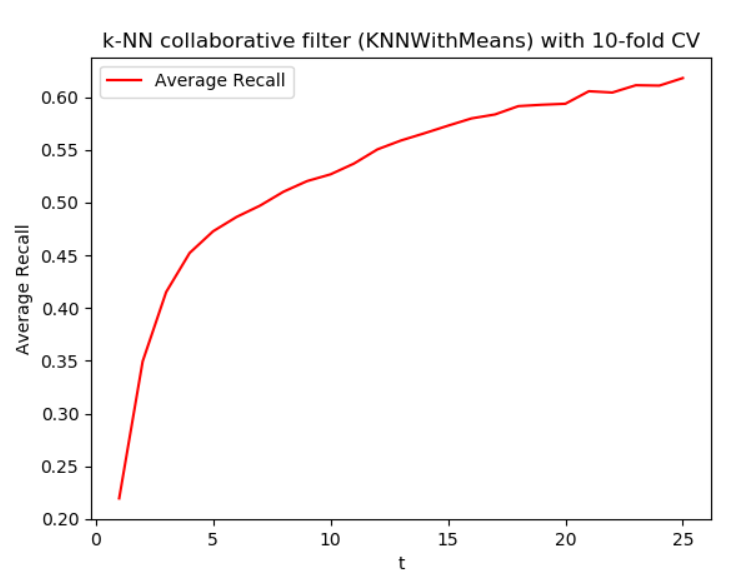
Precision is the fraction of items that the user liked out of the items that were recommended to the user.

Recall is the fraction of recommended items that the user likes out of all the items that the user likes.

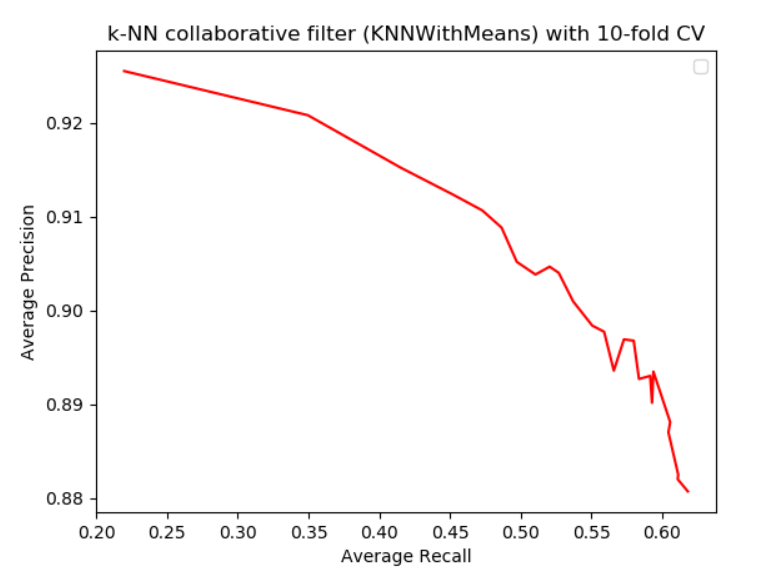
**QUESTION 36:**



As t increases, the average precision decreases generally . There is an inverse relationship between t and average precision.

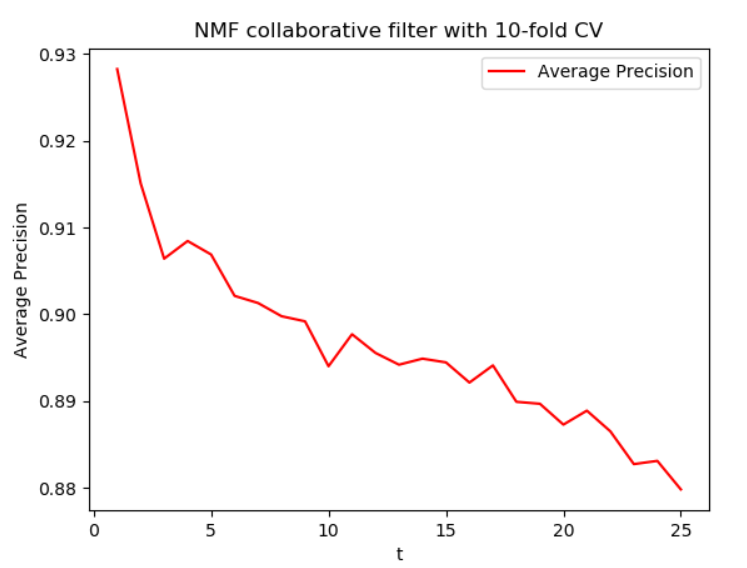


As t increases, the average recall increases. There is a positive, increasing relationship between t and average recall.

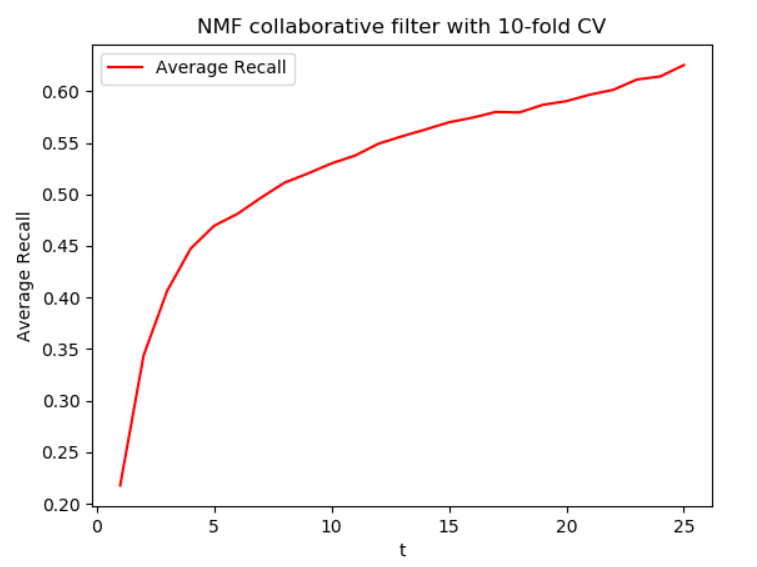


As average recall increases, the average precision decreases. There is an inverse relationship between the two variables. This shows that there is a trade-off. Both variables cannot be maximized simultaneously and there needs to be a compromise.

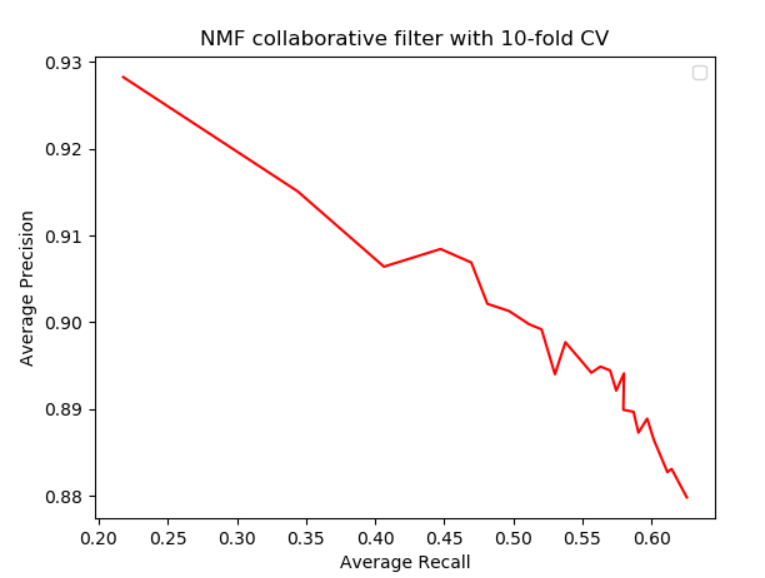
**QUESTION 37:**



As t increases, the average precision decreases generally. There is an inverse relationship between t and average precision.

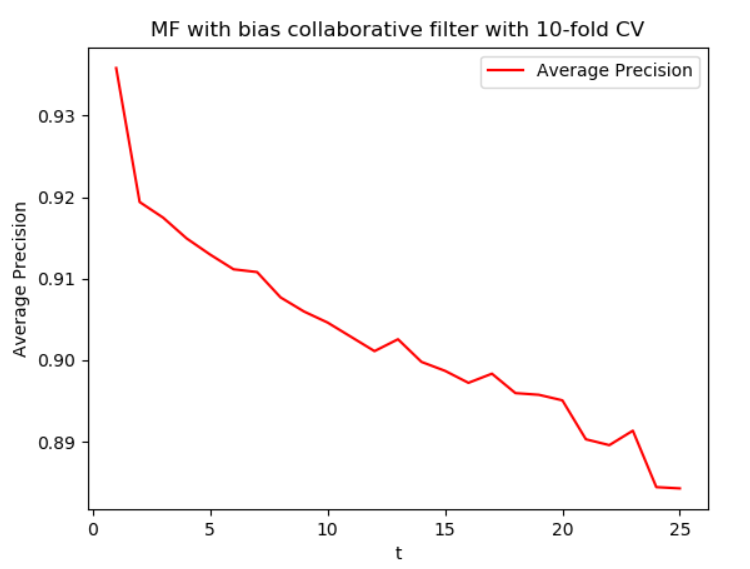


As t increases, the average recall increases. There is a positive, increasing relationship between t and average recall.

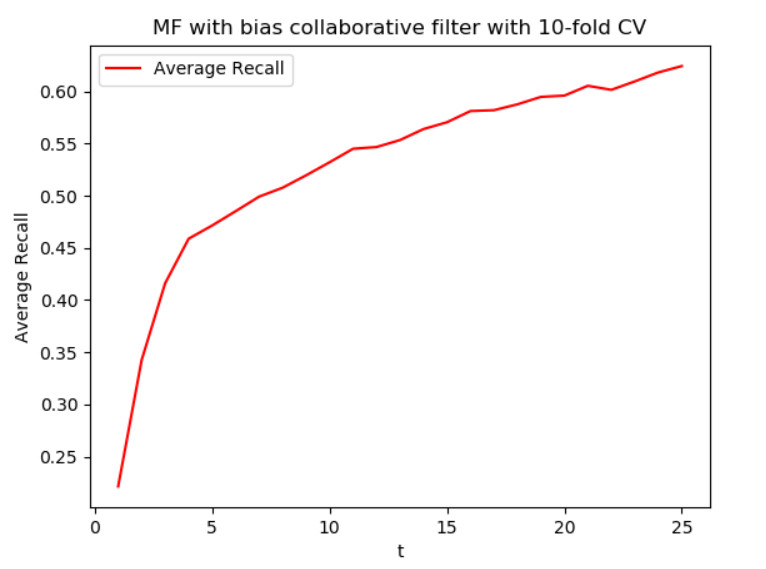


As average recall increases, the average precision decreases. There is an inverse relationship between the two variables. This shows that there is a trade-off. Both variables cannot be maximized simultaneously and there needs to be a compromise.

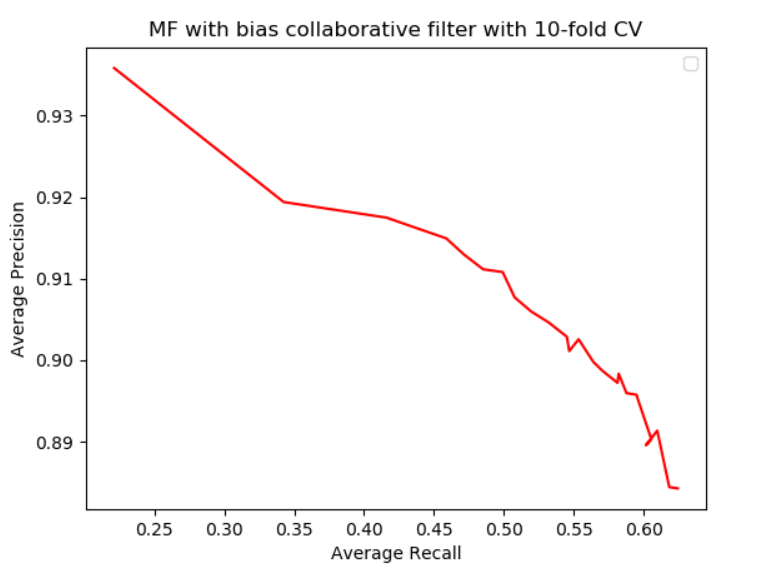
**QUESTION 38:**



As t increases, the average precision decreases generally. There is an inverse relationship between t and average precision.



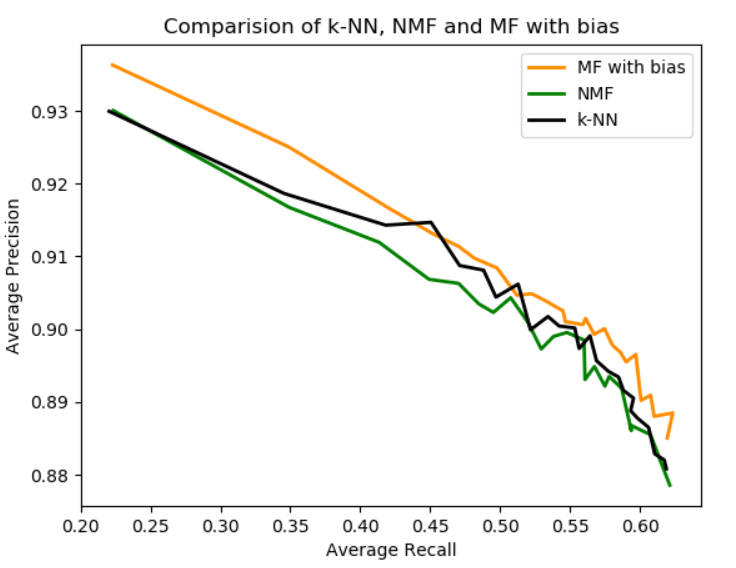
As t increases, the average recall increases. There is a positive, increasing relationship between t and average recall.



As average recall increases, the average precision decreases. There is an inverse relationship between the two variables. This shows that there is a trade-off. Both variables cannot be maximized simultaneously and there needs to be a compromise.

[Grab your reader’s attention with a great quote from the document or use this space to emphasize a key point. To place this text box anywhere on the page, just drag it.]

**QUESTION 39:**



The precision-recall curves for all three prediction algorithms show an inverse relationship between average precision and recall. However, MF with bias shows the best performance since it has the smallest slope. This means that average recall can be increased with only small changes in the average precision.