ECE 219 Project 3: Recommender Systems

Sarah Wilen
UCLA
ECE Department
swilen1@g.ucla.edu
UID: 305070897

Marco Venegas
UCLA
ECE Department
venmarco310@g.ucla.edu
UID: 805461921

Henry Peters
UCLA
CS Department
hpeters@g.ucla.edu
UID: 205704526

1. Question 1

1.1. Part A

Sparsity = 0.0170 Our R matrix is very sparse, which makes sense because if it wasn't, that means most users have watched and rated all of the movies in the set, which is not likely.

1.2. Part B

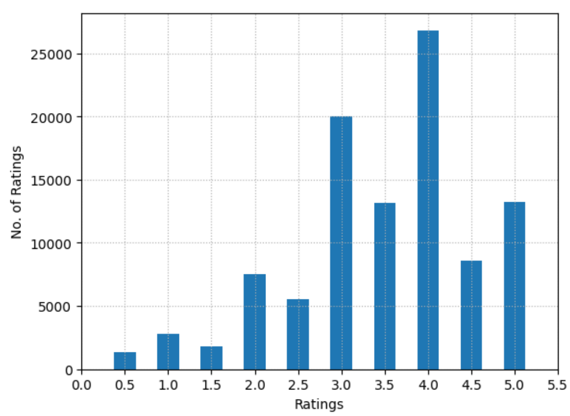

Figure 1: Frequency of the Rating Values

From the histogram shape with an upward trend, we can see that most ratings were higher than 3.0 stars.

1.3. Part C

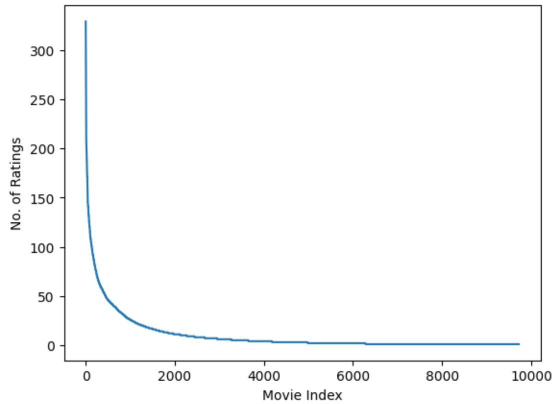

Figure 2: Number of Ratings Received Among Movies

From this distribution, we see that a small minority of the movies have high ratings. The majority have fewer than 40 ratings.

1.4. Part D

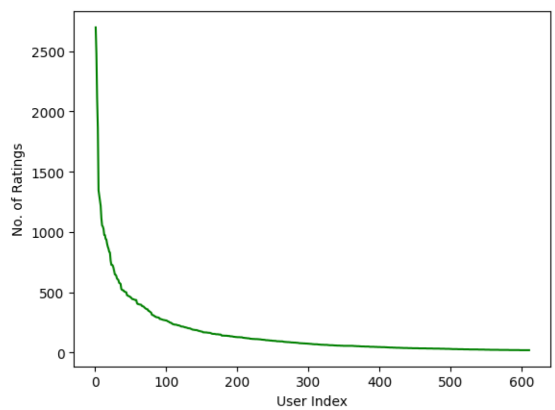

Figure 3: Ratings Among Users

1.5. Part E

We notice that the larger the movie and user index, the smaller the ratings received/given. This may suggest that the majority amount of movies received low amount of ratings. Therefore, as we saw from part a, the rating matrix is sparse. For our recommender system, we should watch out for overfitting by introducing data augmentation (data imputation) and/or regularization. The intuition is that a movie without many ratings won't be recommended as much. Additionally, collaborative filtering will not work as well in this situation because these movie's ratings are not sampled enough.

1.6. Part F

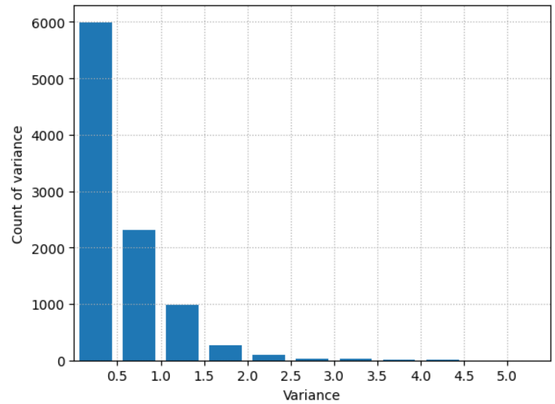

Figure 4: Variance of the Rating Values Received by each movie

Since a majority of the variance lies between 0 and 1.5, we can tell that the ratings for movies are fairly consistent. This means that a movie's ratings do not vary drastically. The intuition is that when moves are rated very high and low, it is difficult for the recommender to recommend these movies, so it's good that we don't have too many controversial movies in our set.

2. Question 2

2.1. Part A

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

2.2. Part B

The intersection of I_u and I_v means the set of movies that are rated by both users u and v. Since the rating matrix is quite sparse, it is possible that users u and v have not rated any of the same movies, therefore, their intersection is 0.

3. Question 3

Mean-centering allows us to obtain a dataset whose mean is zero. This allows different data sets to be comparable. For example, if someone tends to be a critique and rates everything equally low or someone doesn't really care and rates everything similarly high, then we need to subtract the mean to result in a more equitable distribution across users. It acts as a normalization to reduce bias and traits in the ratings, as well as variance caused by polar or critical ratings by some users. It essentially makes the data less noisy

4. Question 4

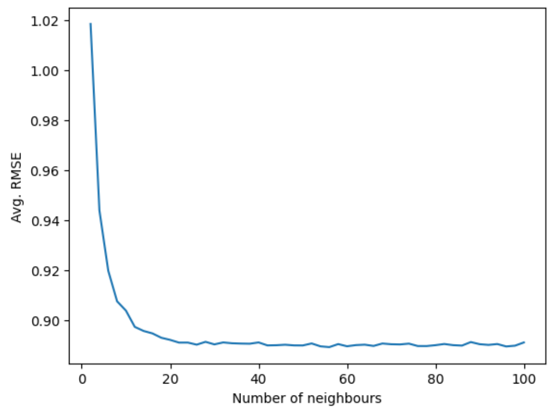

Figure 5: Avg RMSE

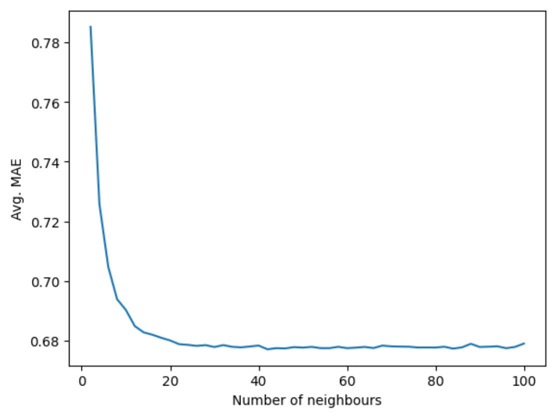

Figure 6: Avg MAE

5. Question 5

The minimum k is 20 with the steady state average RMSE being 0.89 and the average MAE being 0.68

6. Question 6

The average RMSE plot for the unpopular and high variance does not have a clear trend. This is because the number of samples in the unpopular and high-variance subsets is small. For unpopular movies, the movies weren't rated by enough people across a spectrum, so we can't figure out how to recommend them. Additionally, for the high variance set, the movie was too problematic, so it won't likely be recommended. To improve these trimmed sets, we might want to include user and item-specific features. For example, we may want to include a genre feature, which we will do later, or the age and gender of the user.

The popular set still exhibits a decreasing trend as we expect because there is still a large sample population to recommend from.

6.1. No Trim:

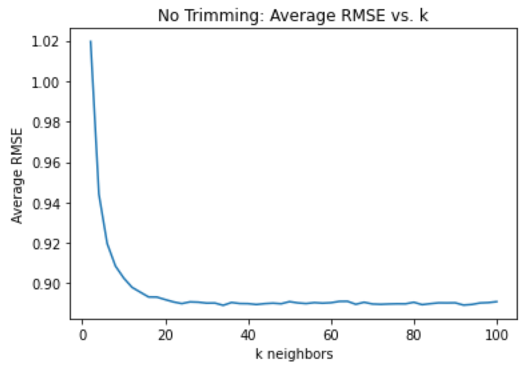

Figure 7: No Trim: Average RMSE

k_min = 20 Average RMSE at k_min = 20: 0.891

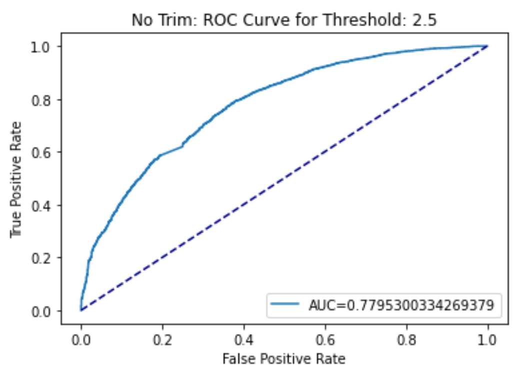

Figure 8: No Trim: ROC

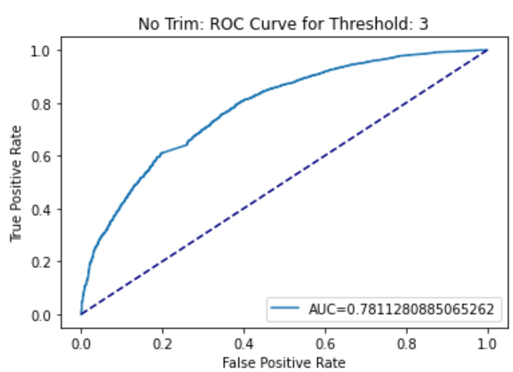

Figure 9: No Trim: ROC

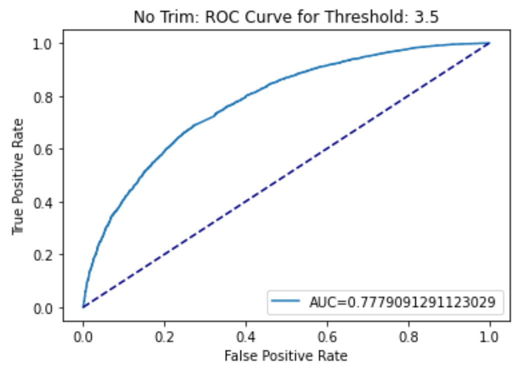

Figure 10: No Trim: ROC

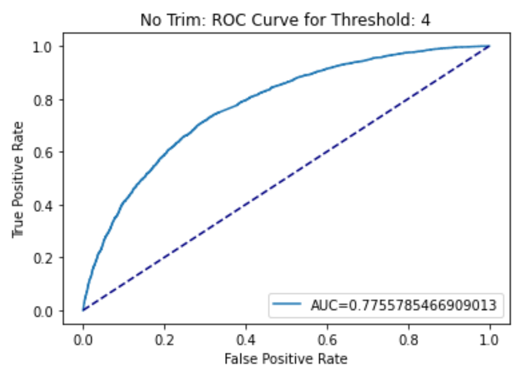

Figure 11: No Trim: ROC

6.2. Popular Trim:

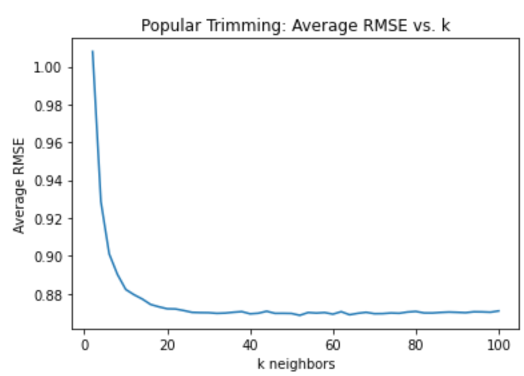

Figure 12: Popular Trim: Average RMSE

k_min = 20 Average RMSE at k_min = 20: 0.877

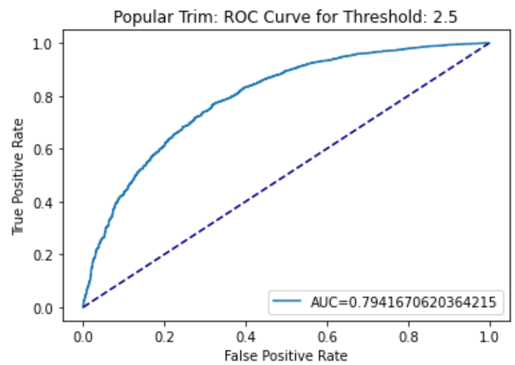

Figure 13: Popular Trim: ROC

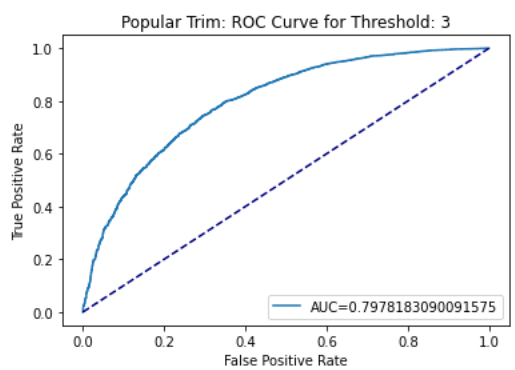

Figure 14: Popular Trim: ROC

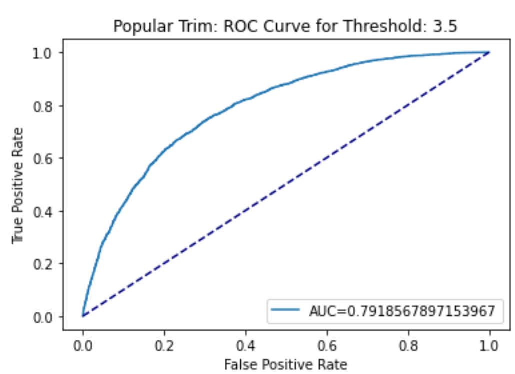

Figure 15: Popular Trim: ROC

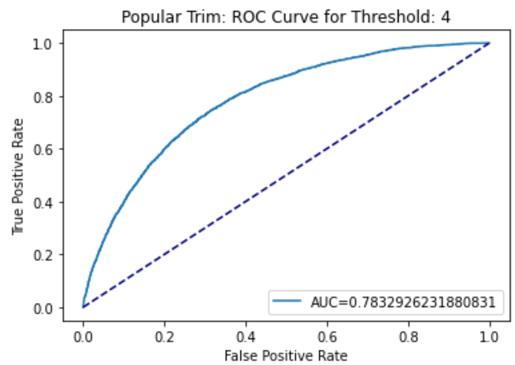

Figure 16: Popular Trim: ROC

6.3. Unpopular Trim:

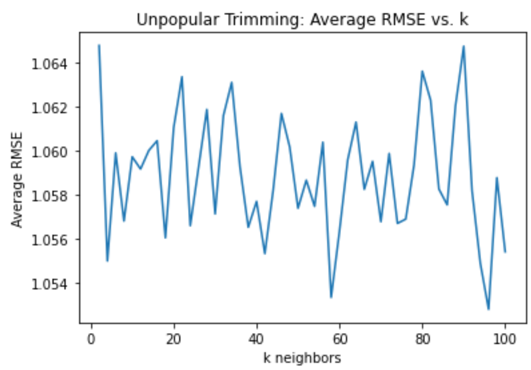

Figure 17: Unpopular Trim: Average RMSE

k_min = 58 Average RMSE at k_min = 20: 1.024

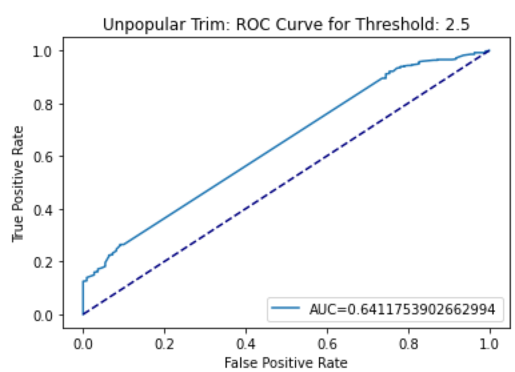

Figure 18: Unpopular Trim: ROC

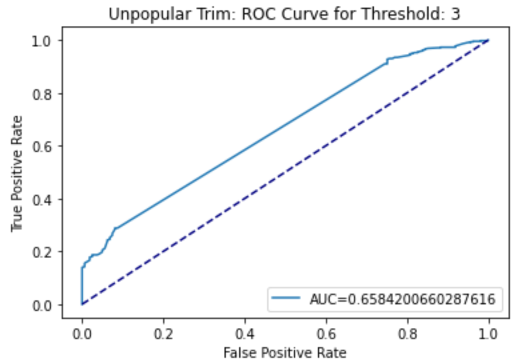

Figure 19: Unpopular Trim: ROC

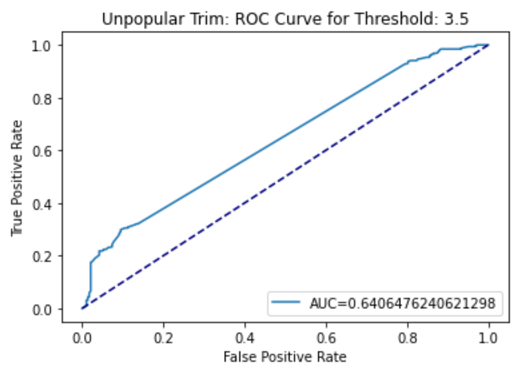

Figure 20: Unpopular Trim: ROC

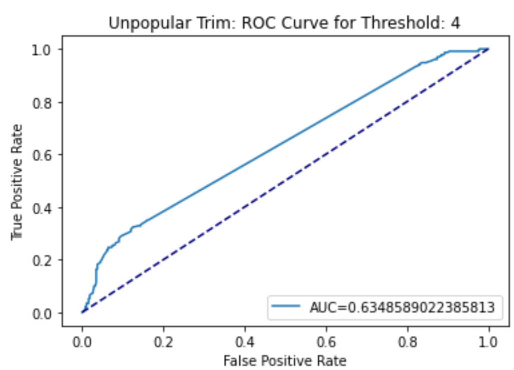

Figure 21: Unpopular Trim: ROC

6.4. High Variance Trim:

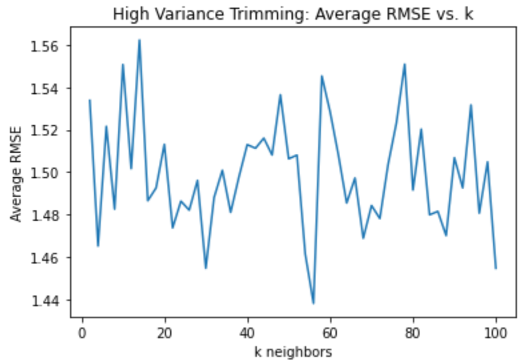

Figure 22: High Variance Trim: Average RMSE

k_min = 58 Average RMSE at k_min = 20: 1.574

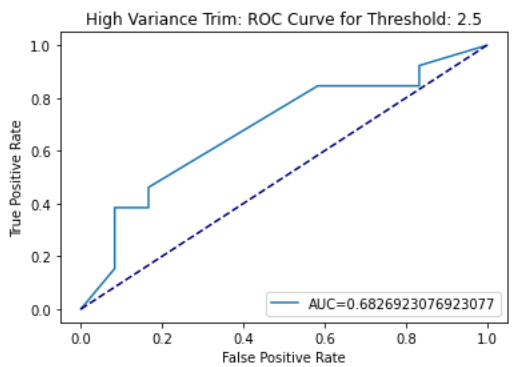

Figure 23: High Variance Trim: ROC

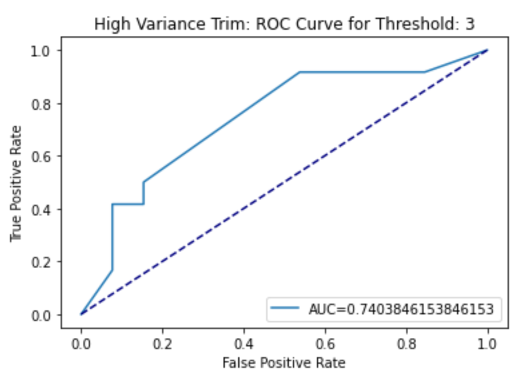

Figure 24: High Variance Trim: ROC

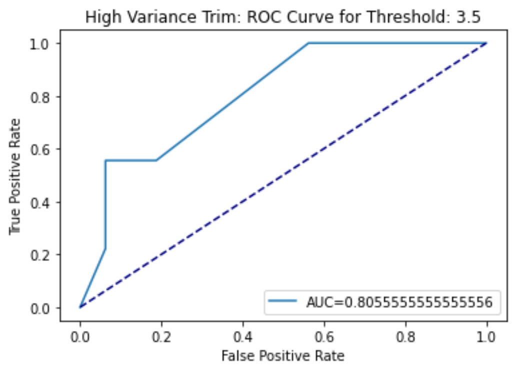

Figure 25: High Variance Trim: ROC

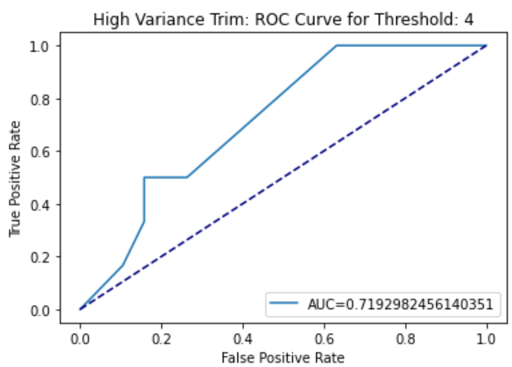

Figure 26: High Variance Trim: ROC

7. Question 7

A function is convex if the Hessian is positive semi-definite. A matrix is positive semi-definite if all of the eigenvalues of the matrix are non-negative.

$$W_{i,j}(r_{i,j} - (UV^T)_{i,j})^2$$

for brevity…

$$w(r - uv)^2 = w(r^2 - 2ruv - u^2v^2)$$

$$H_r(u,v) = \begin{bmatrix} \frac{d^2f}{du^2} & \frac{d^2f}{dudv} \\ \frac{d^2f}{dvdu} & \frac{d^2f}{dv^2} \end{bmatrix}$$

$$H_r(u,v) = \begin{bmatrix} 2v^2 & 4uv - 2r \\ 4uv - 2r & 2u^2 \end{bmatrix}$$

let $v = 2$, $u = 1$, $r = 1$

$$H_r(u,v) = \begin{bmatrix} 8 & 6 \\ 6 & 2 \end{bmatrix}$$

$$\lambda_{1,2} = 5 \pm 3\sqrt{5}$$

$$5 - 3\sqrt{5} < 0$$

Therefore, NMF is NOT convex

Least Squares Set up:

$$\min_{V} \sum_{i=1}^{m} \sum_{j=1}^{n} W_{i,j}(r_{i,j} - (UV^T)_{i,j})^2$$

8. Question 8

8.1. Part A

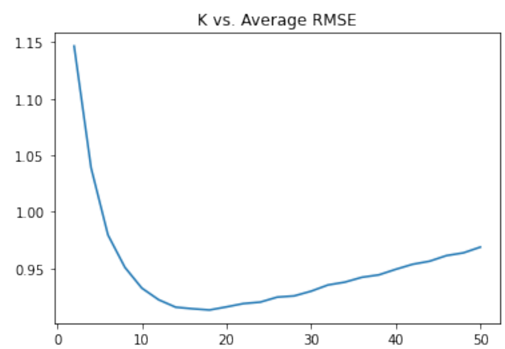

Figure 27: K vs. RMSE

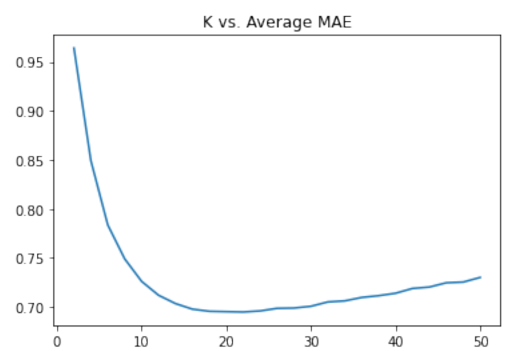

Figure 28: K vs. MAE

8.2. Part B

min k = 18
average RMSE = 0.9134
min k = 20
average MAE = 0.6939
There are 18 defined genres, so the optimal number of latent factors is, in fact, similar to the number of genres (which is 20 if you count no genres as a group).

8.3. Part C

8.3.1 Popular Trim

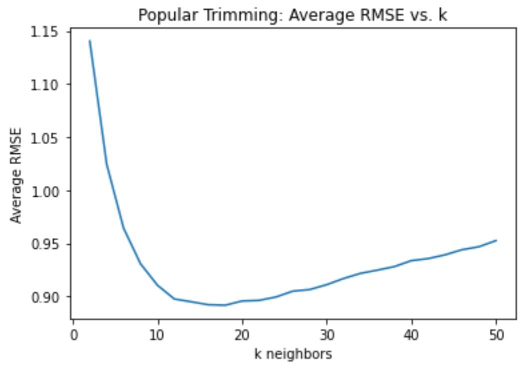

Figure 29: Popular Trim: RMSE

k=17
Average RMSE: 0.9017

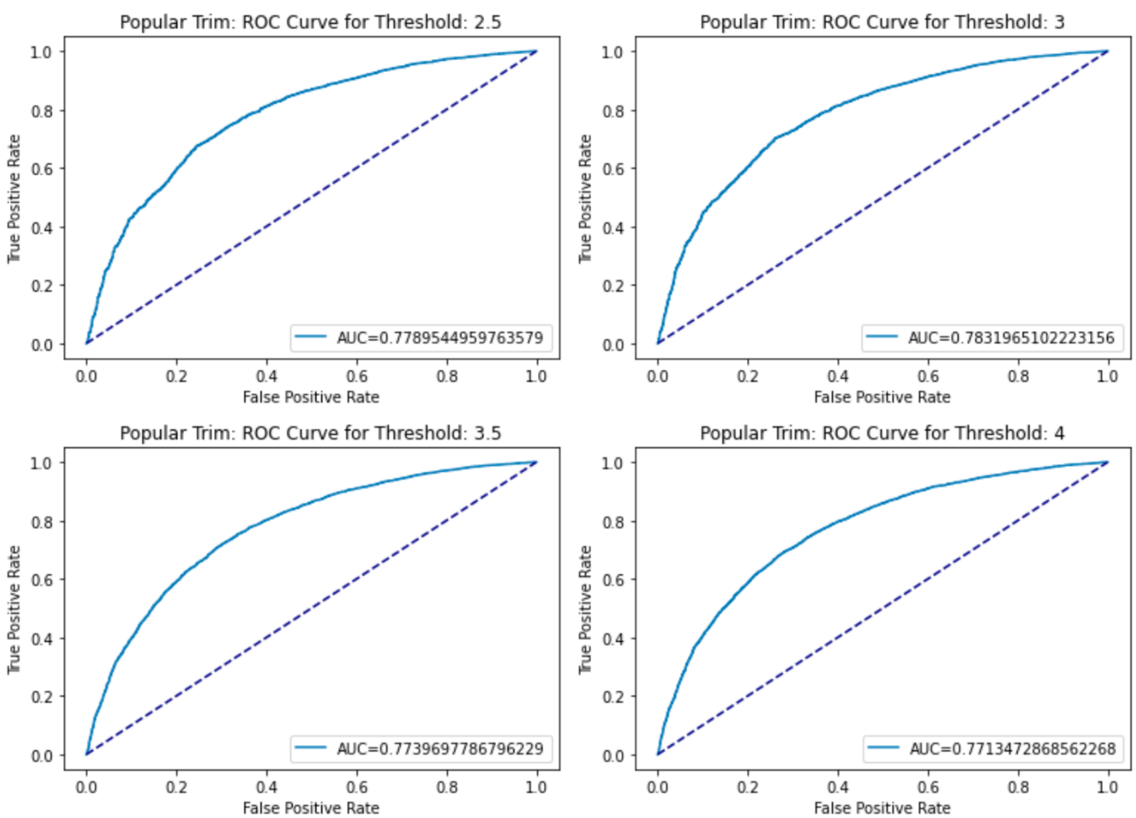

Figure 30: Popular Trim: ROC

8.3.2 Unpopular Trim

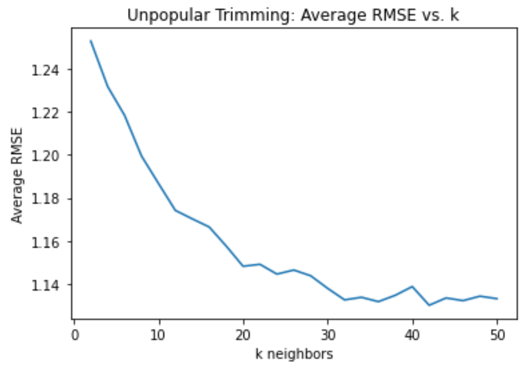

Figure 31: Unpopular Trim: RMSE

k=50 Average RMSE: 1.117

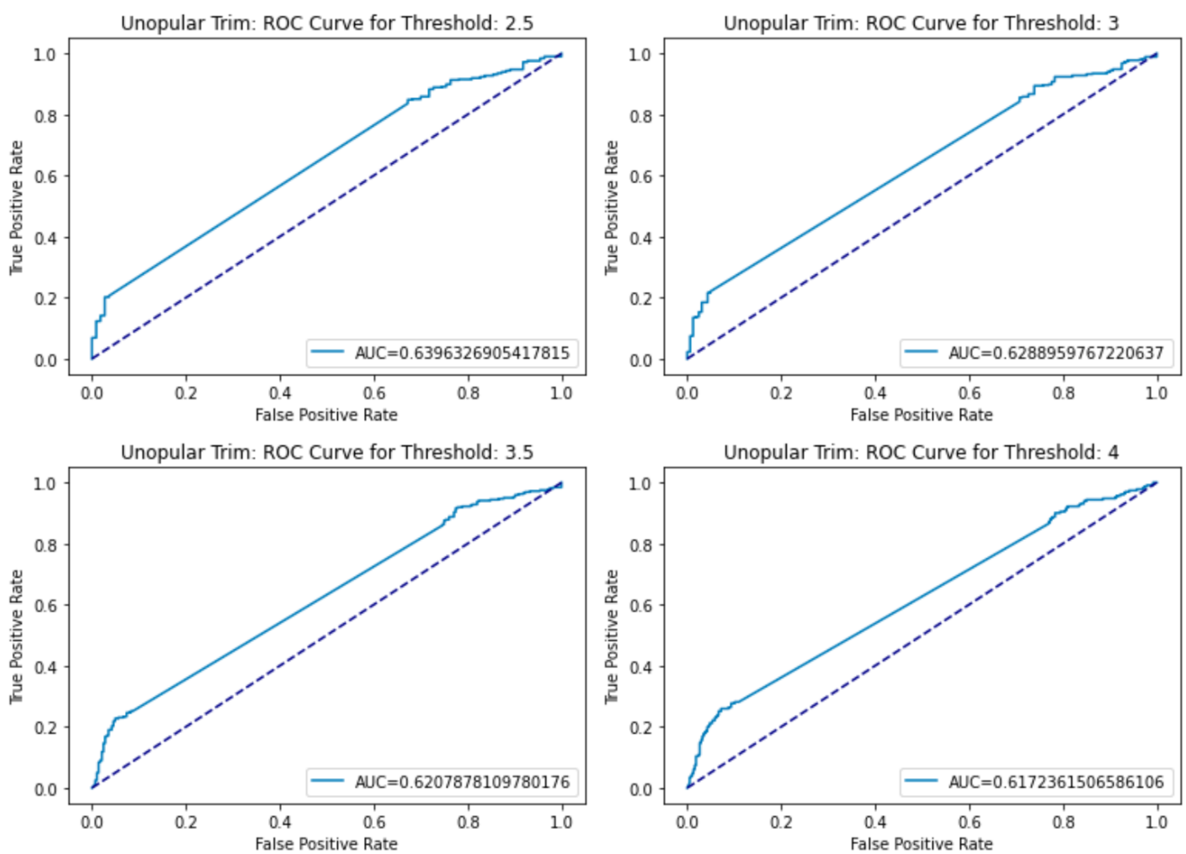

Figure 32: Unpopular Trim: ROC

8.3.3 High Variance Trim

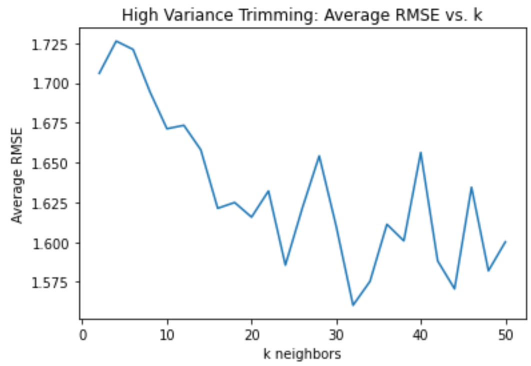

Figure 33: High Variance Trim: RMSE

k=32 Average RMSE: 1.56

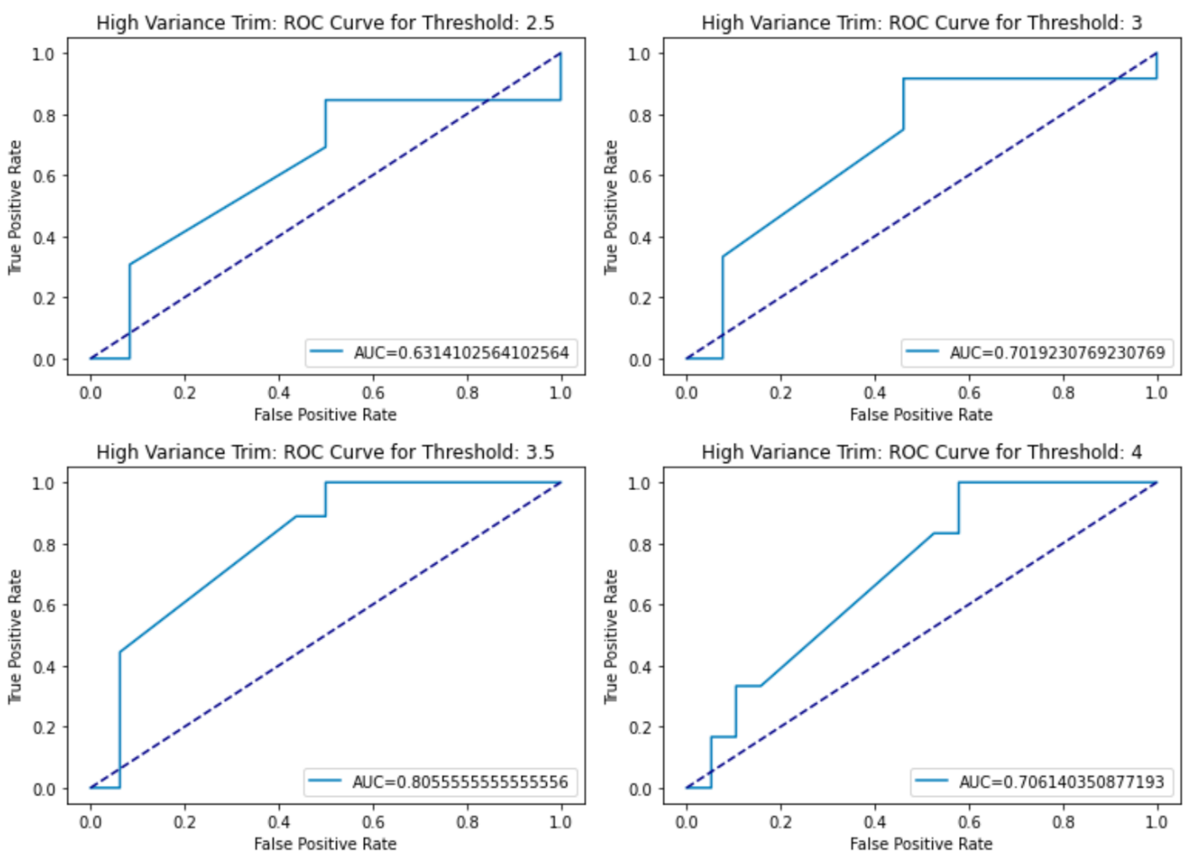

Figure 34: High Variance Trim: ROC

9. Question 9

In previous questions, we noticed that the optimal number of latent factors was around 18, which is the number of genres that the set has. Therefore, in this problem, when we set the latent factors to 20, we sorted the top 10 movies within that column (within that latent factor). We notice that the movies tend to be within the same genre within each column (each latent factor). For example, column 0's top 10 movies are mostly drama and comedy. Column 1's top 10 movies are mostly drama and comedy. Therefore, we can tell that there IS a connection between the latent factors and movie genres, which makes sense. Latent factors are factors that are inferred, and although we didn't do any filtering directly on the genre feature, genres are inherent to the movie's rating and likeability, so it naturally is a factor.

<table>
<tr><td>Column: 0
Drama
Drama|Romance|War
Comedy|Drama|Romance
Horror|Mystery|Thriller
Adventure|Children|Comedy
Drama|Thriller
Comedy
Action|Drama|War
Drama|Romance
Action|Adventure|Thriller</td><td>Column: 1
Drama
Drama|Romance|War
Comedy|Drama|Romance
Horror|Mystery|Thriller
Adventure|Children|Comedy
Drama|Thriller
Comedy
Action|Drama|War
Drama|Romance
Action|Adventure|Thriller</td><td>Column: 2
Drama
Drama|Romance|War
Comedy|Drama|Romance
Horror|Mystery|Thriller
Adventure|Children|Comedy
Drama|Thriller
Comedy
Action|Drama|War
Drama|Romance
Action|Adventure|Thriller</td></tr>
<tr><td>Column: 3
Drama
Drama|Romance|War
Comedy|Drama|Romance
Horror|Mystery|Thriller
Adventure|Children|Comedy
Drama|Thriller
Comedy
Action|Drama|War
Drama|Romance
Action|Adventure|Thriller</td><td>Column: 4
Drama
Drama|Romance|War
Comedy|Drama|Romance
Horror|Mystery|Thriller
Adventure|Children|Comedy
Drama|Thriller
Comedy
Action|Drama|War
Drama|Romance
Action|Adventure|Thriller</td><td>Column: 5
Drama
Drama|Romance|War
Comedy|Drama|Romance
Horror|Mystery|Thriller
Adventure|Children|Comedy
Drama|Thriller
Comedy
Action|Drama|War
Drama|Romance
Action|Adventure|Thriller</td></tr>
<tr><td>…</td><td>…</td><td>…</td></tr>
<tr><td>Column: 7
Drama
Drama|Romance|War
Comedy|Drama|Romance
Horror|Mystery|Thriller
Adventure|Children|Comedy
Drama|Thriller
Comedy
Action|Drama|War
Drama|Romance
Action|Adventure|Thriller</td><td>Column: 10
Drama
Drama|Romance|War
Comedy|Drama|Romance
Horror|Mystery|Thriller
Adventure|Children|Comedy
Drama|Thriller
Comedy
Action|Drama|War
Drama|Romance
Action|Adventure|Thriller</td><td>Column: 15
Drama
Drama|Romance|War
Comedy|Drama|Romance
Horror|Mystery|Thriller
Adventure|Children|Comedy
Drama|Thriller
Comedy
Action|Drama|War
Drama|Romance
Action|Adventure|Thriller</td></tr>
</table>

<table>
<tr><td>Column: 17
Drama
Drama|Romance|War
Comedy|Drama|Romance
Horror|Mystery|Thriller
Adventure|Children|Comedy
Drama|Thriller
Comedy
Action|Drama|War
Drama|Romance
Action|Adventure|Thriller</td><td>Column: 18
Drama
Drama|Romance|War
Comedy|Drama|Romance
Horror|Mystery|Thriller
Adventure|Children|Comedy
Drama|Thriller
Comedy
Action|Drama|War
Drama|Romance
Action|Adventure|Thriller</td><td>Column: 19
Drama
Drama|Romance|War
Comedy|Drama|Romance
Horror|Mystery|Thriller
Adventure|Children|Comedy
Drama|Thriller
Comedy
Action|Drama|War
Drama|Romance
Action|Adventure|Thriller</td></tr>
</table>

Figure 35: Column Genres

10. Question 10

10.1. Part A

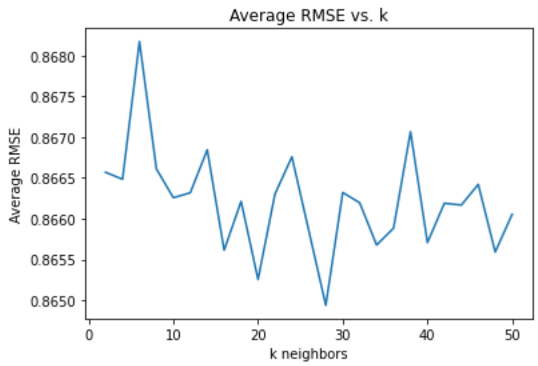

Figure 36: K vs. RMSE

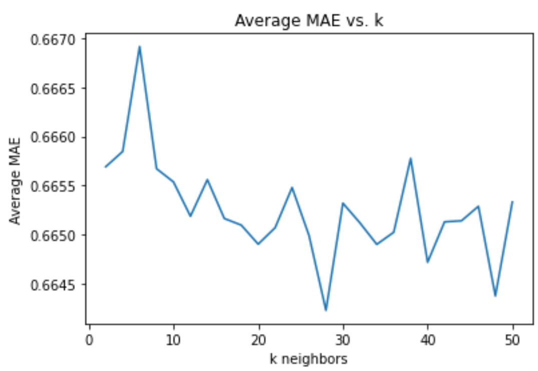

Figure 37: K vs. MAE

10.2. Part B

min k = 28
average RMSE = 0.867
min k = 28
average MAE = 0.6642

There are 20 genres and the optimal latent factors for MF is around 28. however, if we zoom the RMSE and MAE graphs out a bit, we'll notice that the error is pretty consistent around 0.86 and 0.66 respectively, so there is not much of a difference in RMSE and MAE for latent factors of 18 vs. 28, so it is possible that the number of latent factors is representative of the number of movie genres.

10.3. Part C

10.3.1 Popular Trim

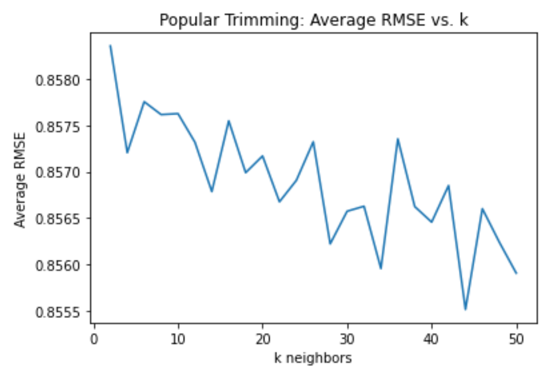

Figure 38: Popular Trim: RMSE

min k = 44 average RMSE = 0.866

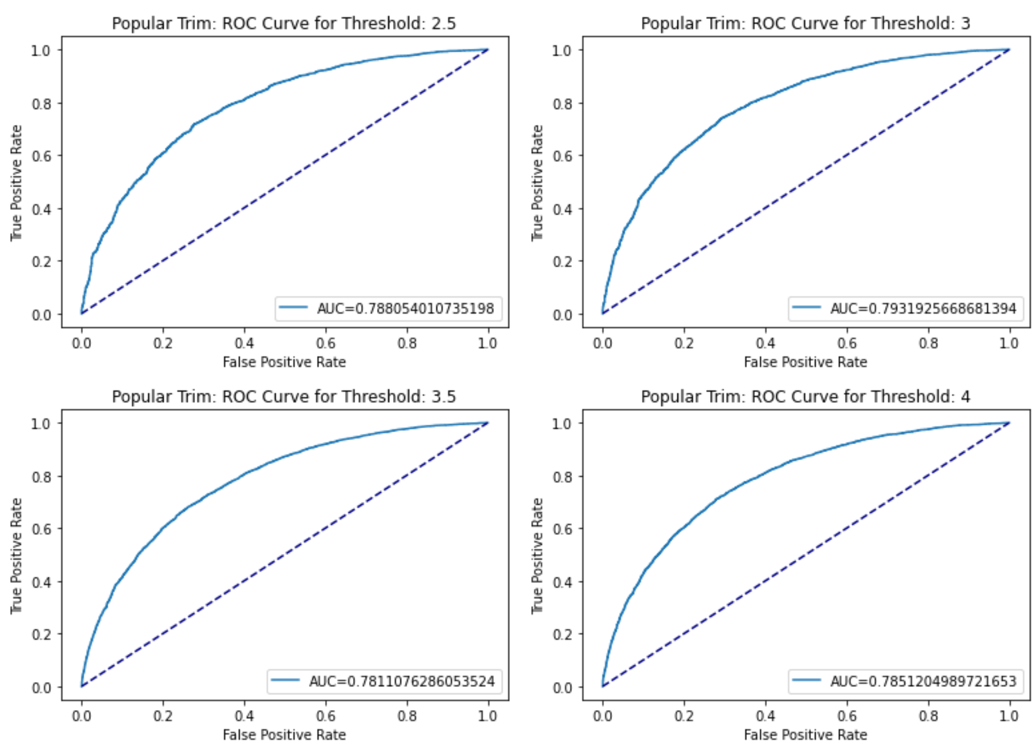

Figure 39: Popular Trim: ROC

10.3.2 Unpopular Trim

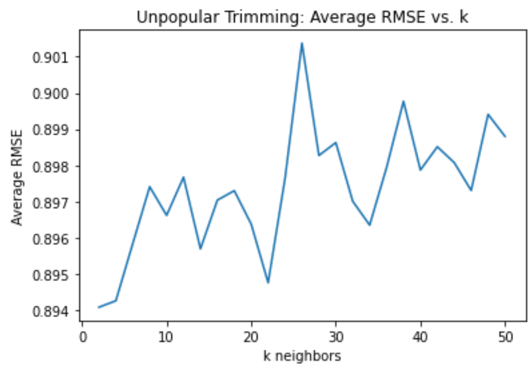

Figure 40: Unpopular Trim: RMSE

min k = 3 average RMSE = 0.8559

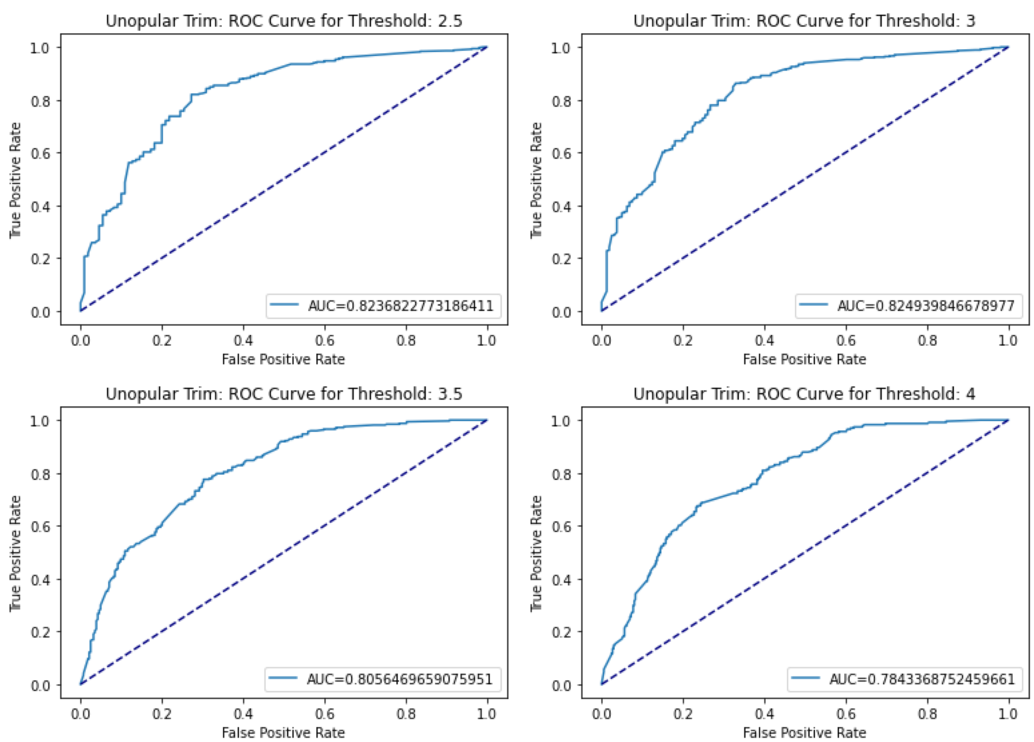

Figure 41: Unpopular Trim: ROC

10.3.3 High Variance Trim

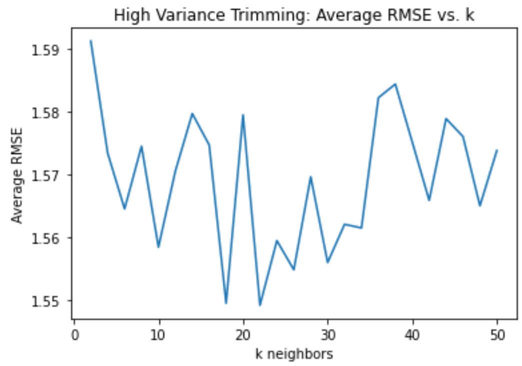

Figure 42: High Variance Trim: RMSE

min k = 22 average RMSE = 1.575

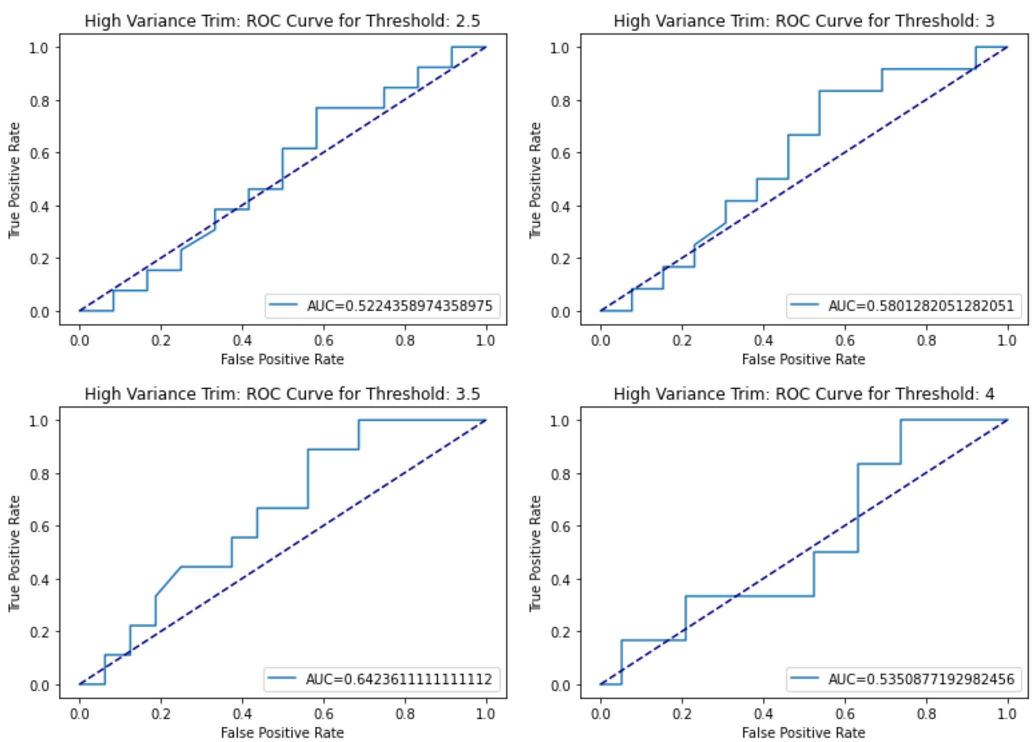

Figure 43: High Variance Trim: ROC

11. Question 11

No Trim - Average RMSE: 0.9410519556984109
Popular Trim- Average RMSE: 0.9376202968550311
Unpopular Trim- Average RMSE: 1.0198618036753762
High Variance Trim- Average RMSE: 1.1113117513991662

12. Question 12

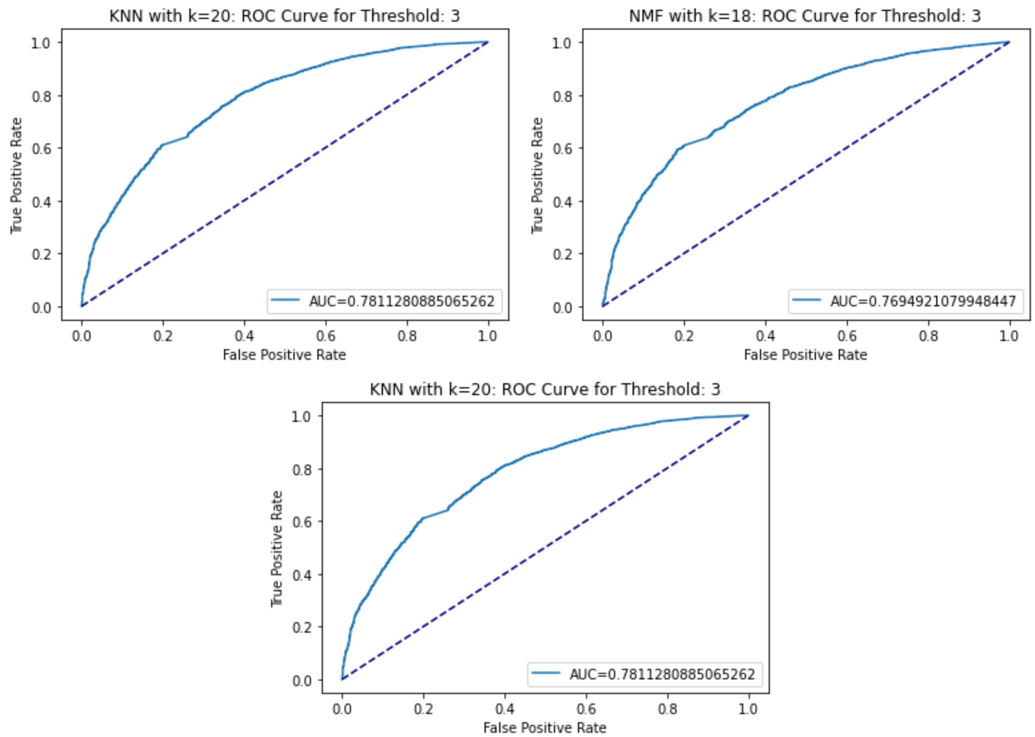

Figure 44: K-NN, NMF, MF ROC Curves

KNN: Average RMSE: 0.8910136771553069
NMF: Average RMSE: 0.9161861277754252
MF with Bias: Average RMSE: 0.8674998266046834

While all three collaborative filters seem to perform with similar average RMSE, the MF with bias has the largest area under the curve. The area under the curve may serve as a good indicator for the best collaborative filter for predicting movie ratings because it uses the whole ROC curve and involves all possible classification thresholds to represent the measure of separability.

13. Question 13

Precision is the items that are both liked by the user and correctly recommended to the user over the number of total items recommended to the user. Essentially, it measures the percentage of the predicted items that are actually liked by the user.

Recall is the items that are both liked by the user and correctly recommended to the user over the number of total items that the user actually likes. Essentially, it measures what percent of true positives (movies liked by the user) were actually identified out of all the movies that the user actually likes.

14. Question 14

14.1. KNN

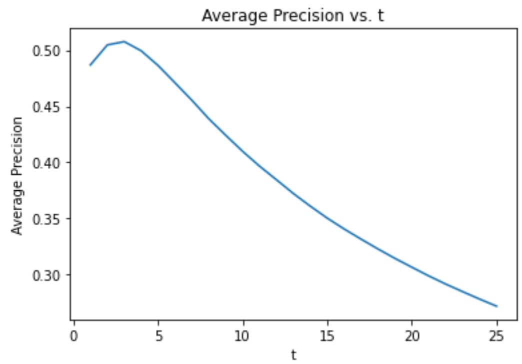

Figure 45: T vs. Precision

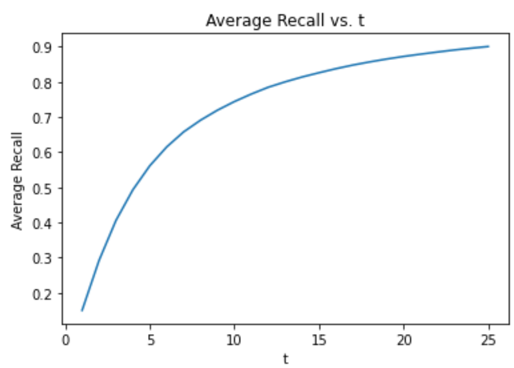

Figure 46: T vs. Recall

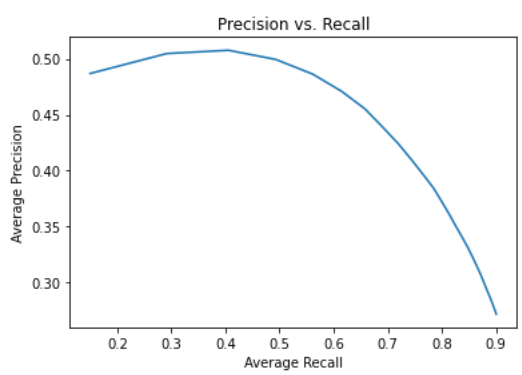

Figure 47: Recall vs. Precision

14.2. NMF

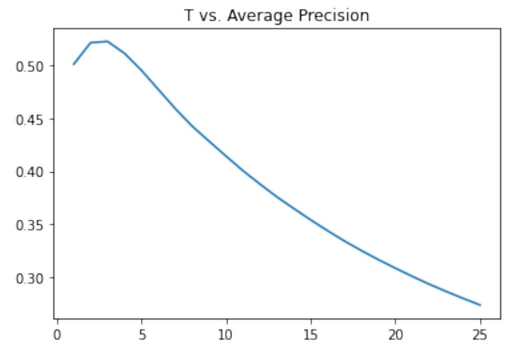

Figure 48: T vs. Precision

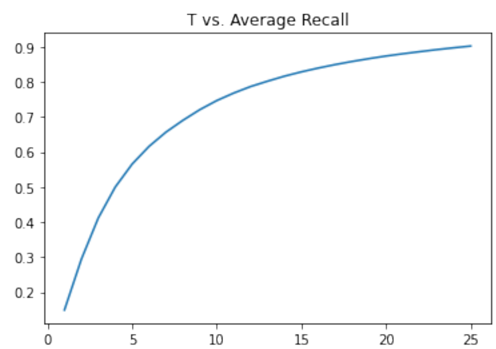

Figure 49: T vs. Recall

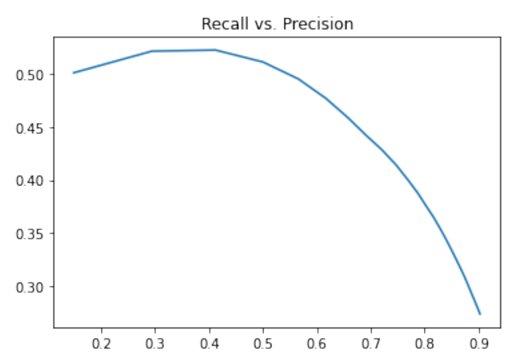

Figure 50: Recall vs. Precision

14.3. MF

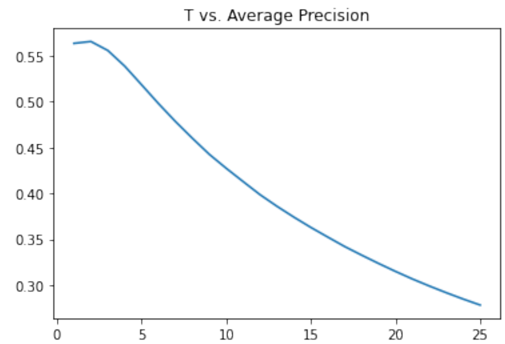

Figure 51: T vs. Precision

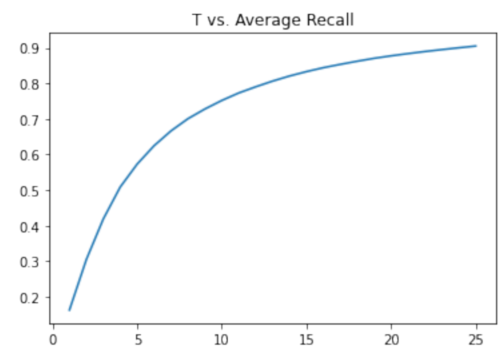

Figure 52: T vs. Recall

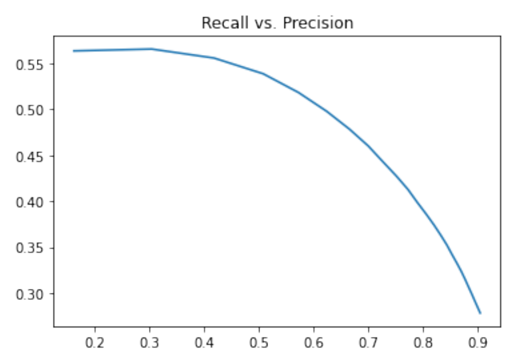

Figure 53: Recall vs. Precision

14.4. Combined Algorithms

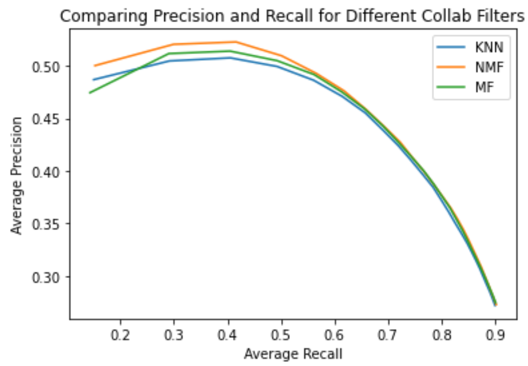

Figure 54: Precision-Recall curves for all 3 Architectures

In general, we notice as the number of items recommended to users (t) increase, the precision decreases and the recall increases. This makes sense because there is a tradeoff between precision and recall. Our models in general have higher recall than precision meaning that our models identified a large proportion of the movies that the user actually liked. It makes sense for precision to decrease as we increase t because as t increases, the set of items recommended to the user increases, so the intersection between what is recommended and what the user actually likes does not increase as much as the S(t) increases, so overall, the precision decreases.