James Abundis
jlabundi@calpoly.edu
Thomas Fahrner
tfahrner@calpoly.edu

Lab 6

Methods:

We implement three collaborative filtering methods:

1. Mean utility
2. Weighted Sum
3. KNN Weighted Sum (k = 1000)

For the weighted sum methods, we use the Pearson Correlation similarity metric.

Research Questions:

We compare three memory-based collaborative filtering methods in an attempt to predict the rating a user would give a joke they haven't seen before. For a given user and joke, our algorithms look at how that user's peers have rated that joke. Based on the similarity to their peers, we predict that the rating the user will give the joke.

For each predicted rating, we look at the absolute error between the user's actual rating and the predicted rating. For a group of test cases, we can calculate the mean average error (MAE). We use this metric to determine the relative success of each method.

- For randomly chosen users and jokes, which method tends to give the lowest MAE? The highest?
- What is the standard deviation of the MAE across separate runs?
- How often does the "worst" method give the lowest MAE? How often does the "best" method give the highest MAE?

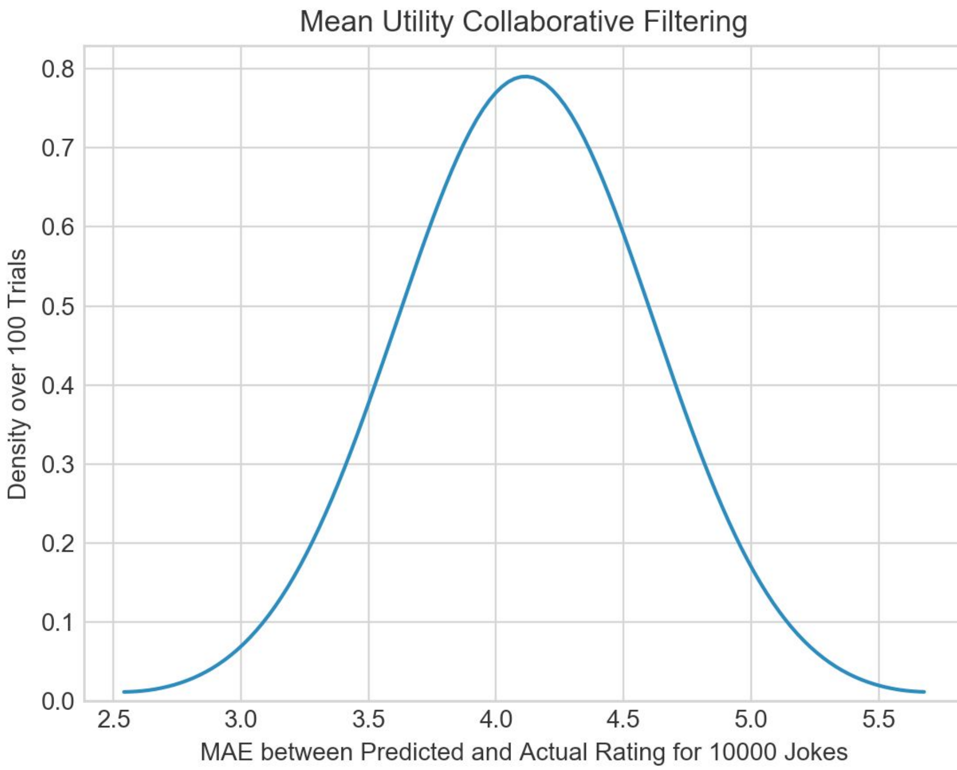
Experiments:

For each collaborative filtering method we ran the following experiment:

1. Randomly select 10000 test cases (100 for KNN): (userId, jokeId).
2. Blind our algorithm from the rating.
3. Predict the rating using the rest of the dataset.
4. Compute the absolute difference between the actual rating and predicted rating.
5. Compute the MAE over the test cases.

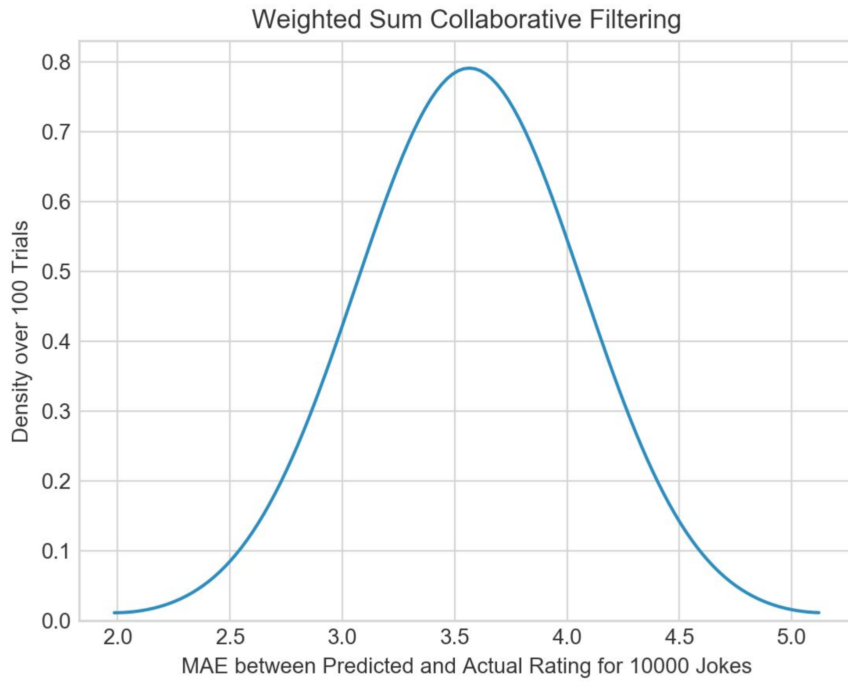
6. Repeat this 100 times. Compute the average MAE and standard deviation of the MAE over the 100 runs.

Results:



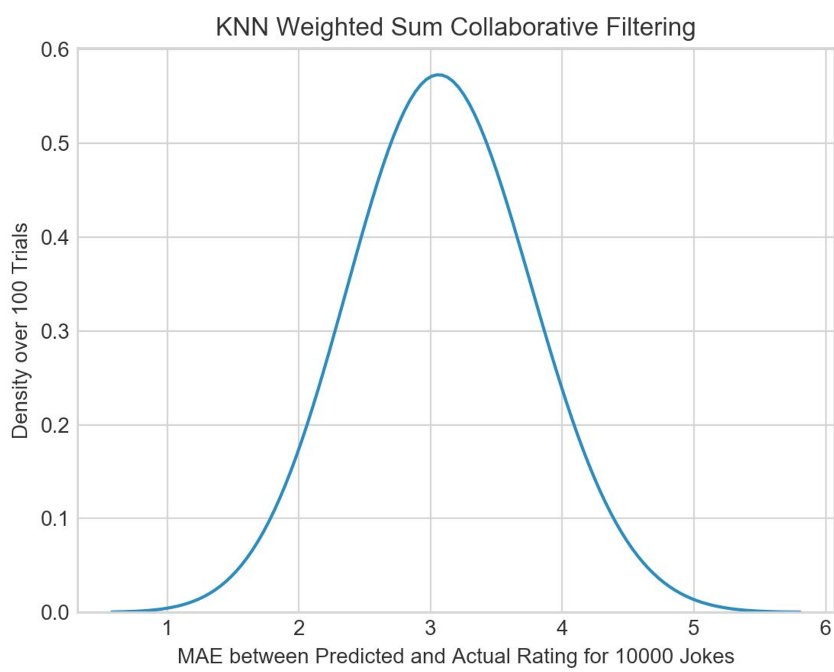
Mean MAE: 4.115

Standard Deviation of the MAE: 0.031



Mean MAE: 3.565

Standard Deviation of the MAE: 0.025



Mean MAE: 3.082

Standard Deviation of the MAE: 0.478

Conclusions:

Through the study of three memory-based collaborative filtering methods, we have determined the relative success of each. In the three graphs shown above, we see that the density of the MAE forms a bell curve for each algorithm. This suggests a normal distribution of the MAE resulting from the 100 trials. We found that Mean Utility resulted in the highest typical MAE (4.115), while KNN Weighted Sum resulted in the lowest (3.082). We suggest that KNN Weighted Sum is most suited for this: the predicted ratings will on average be 3.082 units away from the actual ratings. The standard deviation of the MAE for the KNN Weighted Sum algorithm is significantly higher than the other algorithms because of the smaller number of test cases chosen. Given more time and better computing capabilities, we would be able to take a more representative sampling of the algorithm.

Overall, we were surprised by the relative simplicity and effectiveness of an algorithm like Mean Utility. While Weighted Sum and KNN with Pearson Correlation offers a better MAE, there are tradeoffs with the complexities of the algorithms. We see that all three methods have large spreads in their MAE, and that the wider bell curve corresponds with a lower average MAE. In the future, we would explore different variations of these algorithms and more collaborative filtering techniques. It may be interesting to study different metrics beyond MAE.