Machine Learning & Data Mining
Miniproject 2: MovieLens Visualization

Caltech CS/CNS/EE 155
Released February 16st, 2024

---

1 Introduction

- Team name: Butters
- Team member names: Etienne Casanova, Jack Myles, Ishaan Mantripragada
- Work Division: The three of us worked on the basic visualization together to get an understanding of our dataset. After that, Etienne and Ishaan worked on the first Matrix Factorization method, Etienne worked on the second Matrix factorization method, and Ishaan did the third method. Jack was responsible for projections and visualizations. After each step in the process, everything was discussed to make sure that every member understood every part of the project.
- Packages used for the project: Pandas, Numpy, MatplotLib, Seaborn, Surprise, and requests.
- Piazza Post Link

2 Basic Visualization

Code: Basic Visualization Code

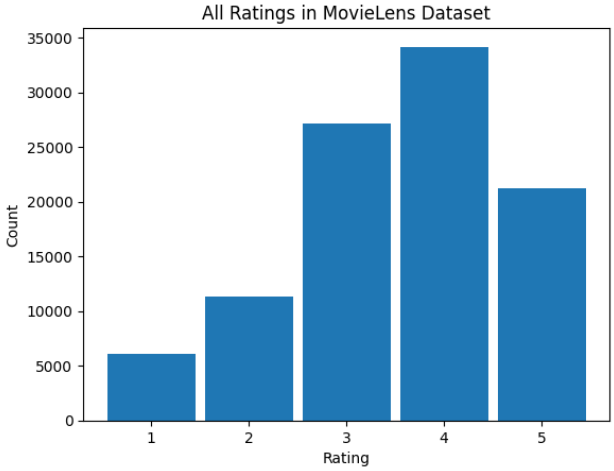


Figure 1: All ratings in the MovieLens Dataset

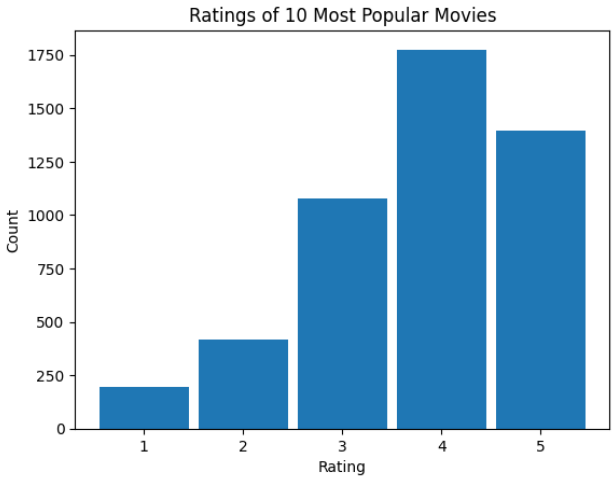


Figure 2: Top Ten Most Popular Movies

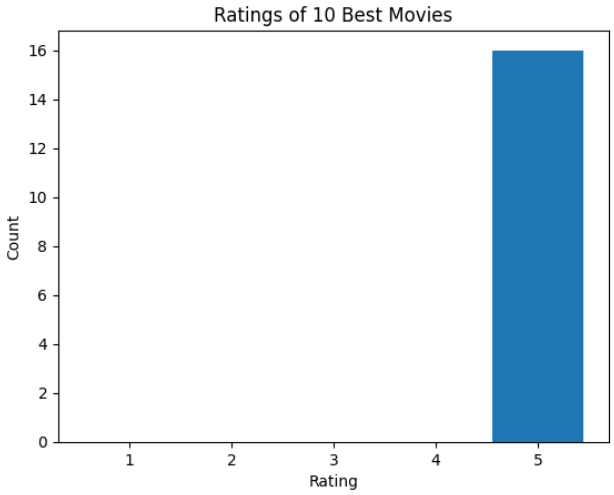


Figure 3: Ratings of the Ten Best Movies (Note: their average is 5.0, so frequency of non-perfect scores is 0)

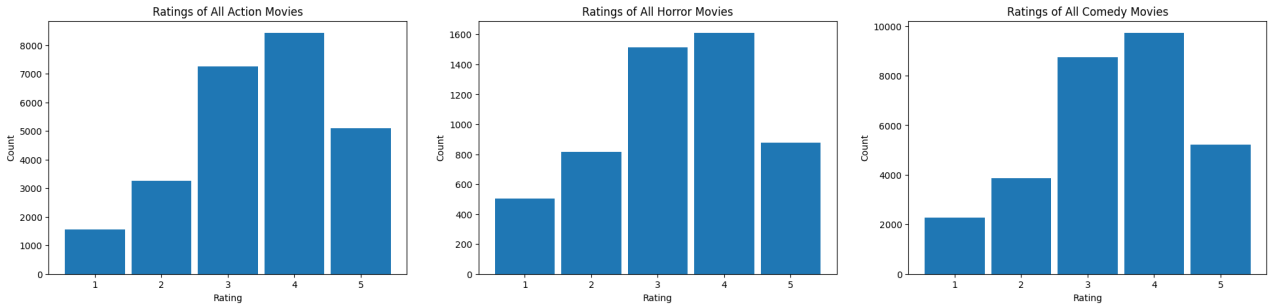


Figure 4: Ratings for Different Genres. Specifically, Action, Horror, and Comedy movies.

We observed that generally the distribution of the movie ratings seems to be skewed left. There is a much higher frequency of higher ratings in the MovieLens dataset than there is of lower ratings (higher being 4-5 and lower 1-2). The most common rating for a movie is 4/5, followed by 3/5, followed by 5/5. This is as expected, as people tend to watch movies that they think they will like, therefore they are more likely to give the movie a higher rating (also people tend to rarely give really ratings). The most popular movies have a higher average score than the average score of all the movies, but the most popular movies are less highly scored than the highest rated movies. This makes sense, as the highest rated movies, by definition, will have the highest ratings. For the highest rated movies they all achieve a perfect average score of 5.0, but it must be noted that they all have a very low amount of ratings, which could mean their perfect scores are not representive of their true scoring distribution. The rating distributions for the three genres selected (Action, Horror, and Comedy) are all pretty similar. Action movies seems to perform slightly better than the other two genres, and Comedy movies seem to slightly outperform Horror movies.

3 Matrix Factorization Visualization

Code: Matrix Visualization Code

Method 1:

For our first method, we used the same SVD method from Homework 5 with a few modifications. As in Homework 5, our learning objective was as follows:

$$\arg\min_{U,V} \frac{\lambda}{2}\left(\|U\|_F^2 + \|V\|_F^2\right) + \frac{1}{2}\sum_{i,j}\left(Y_{ij} - u_i^T v_j\right)^2$$

The value Y_{ij} corresponds to the rating of movie j by user i. As the dataset contains missing rating values for certain movies, collaborative filtering is utilized to develop a latent representation of movies U and users V that accounts for the observed ratings and enables the prediction of all user ratings for all movies. The gradient for row/col of U and V where calculated as follows:

$$\partial u_i = \lambda u_i - v_j(Y_{ij} - u_i^T v_j)$$
$$\partial v_i = \lambda v_i - u_j(Y_{ij} - u_i^T v_j)$$

These values were then multiplied by the learning rate η for each iteration. We then performed a small grid search to determine the optimal number of epochs, learning rate, and regularization term.

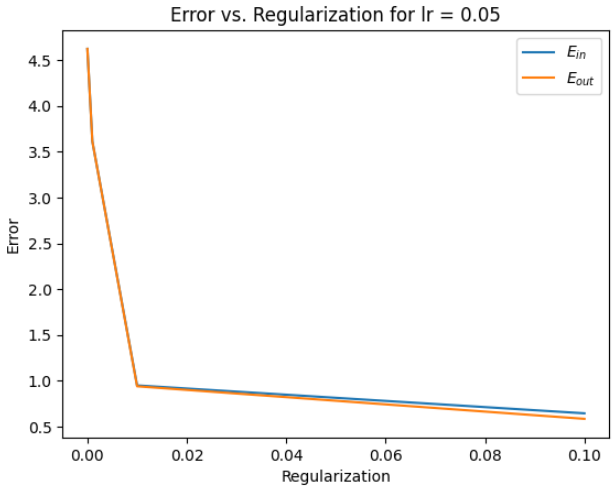


Figure 5: Error vs Regularization Rate

We found that the most optimal regularization term was when $\eta = 0.1$. Keeping this constant, we plotted the learning rate.

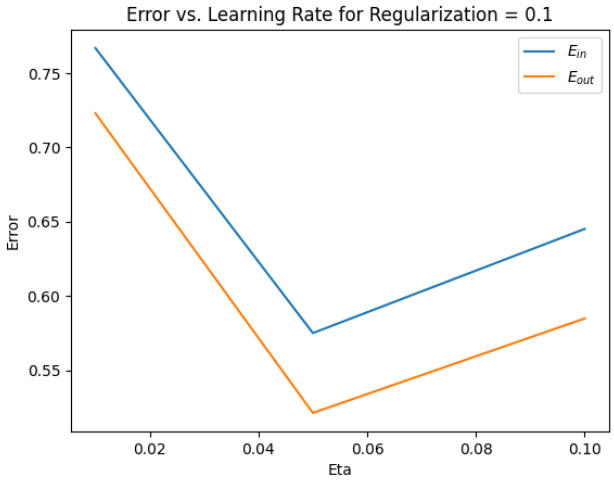
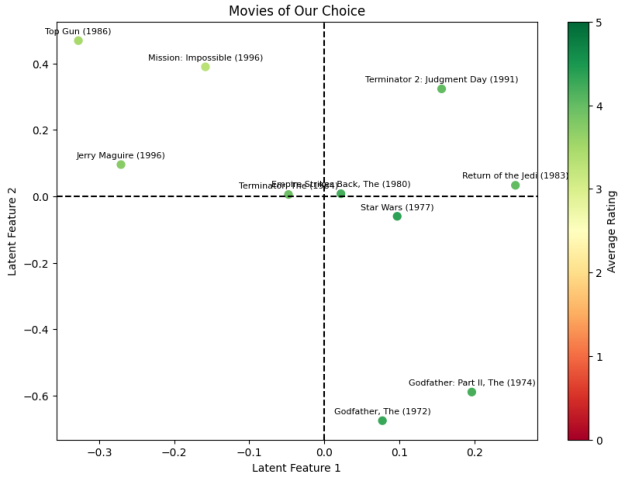
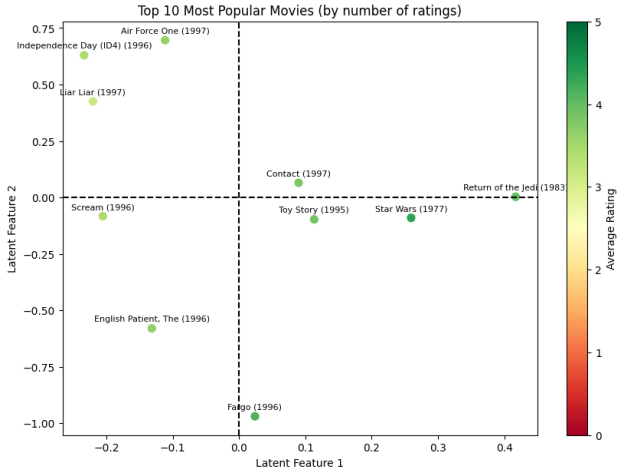


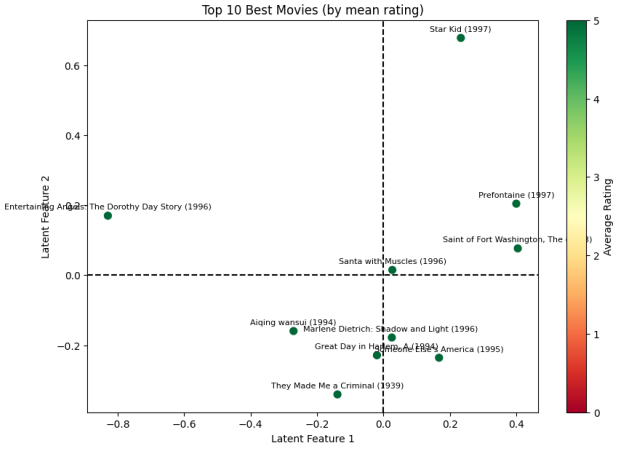
Figure 6: Error vs Learning Rate

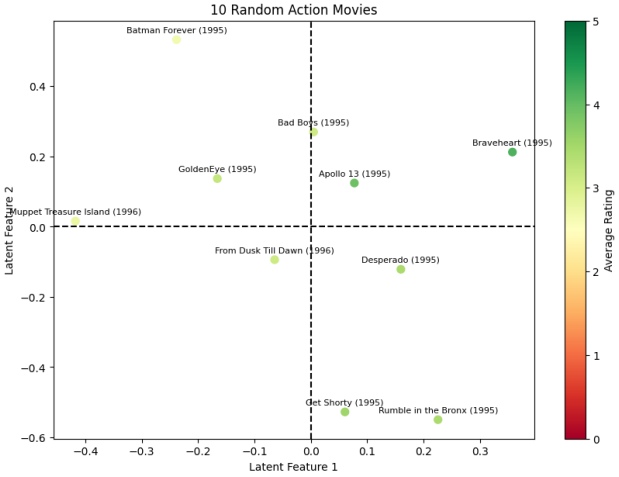
We found that the most optimal learning rate to minimize the error was $\lambda = 0.05$. Our final hyperparameters (which we will keep constant for all three methods) were $\eta = 0.1$, $\lambda = 0.05$, n_epochs = 300, k = 20.

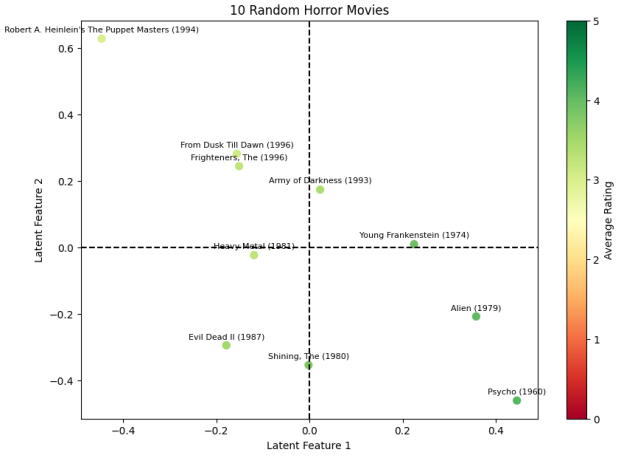
Visualizations:

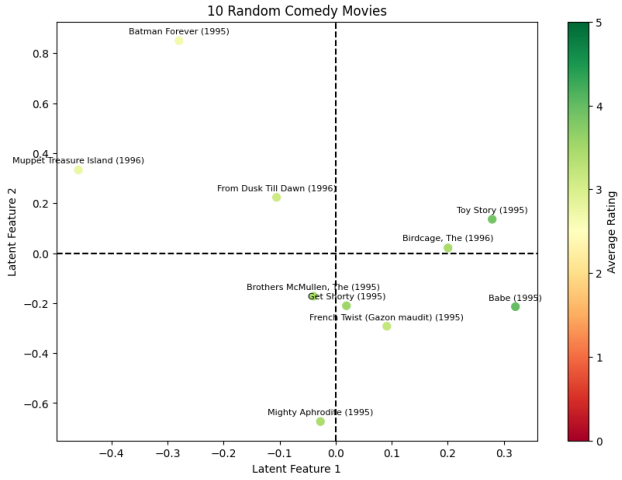





Top 10 Best Movies (by mean rating)
Star Kid (1997)
Prefontaine (1997)
Entertaining Angels: The Dorothy Day Story (1996)
Saint of Fort Washington, The (1993)
Santa with Muscles (1996)
Aiqing wansui (1994)
Marlene Dietrich: Shadow and Light (1996)
Great Day in Harlem, A (1994)
Marlene Dietrich: America (1995)
They Made Me a Criminal (1939)
Latent Feature 1
Latent Feature 2
Average Rating


10 Random Action Movies
Batman Forever (1995)
Bad Boys (1995)
Braveheart (1995)
GoldenEye (1995)
Apollo 13 (1995)
Muppet Treasure Island (1996)
From Dusk Till Dawn (1996)
Desperado (1995)
Get Shorty (1995)
Rumble in the Bronx (1995)
Latent Feature 1
Latent Feature 2
Average Rating


10 Random Horror Movies
Robert A. Heinlein's The Puppet Masters (1994)
From Dusk Till Dawn (1996)
Frighteners, The (1996)
Army of Darkness (1993)
Young Frankenstein (1974)
Heavy Metal (1981)
Alien (1979)
Evil Dead II (1987)
Shining, The (1980)
Psycho (1960)
Latent Feature 1
Latent Feature 2
Average Rating


10 Random Comedy Movies
Batman Forever (1995)
Muppet Treasure Island (1996)
From Dusk Till Dawn (1996)
Toy Story (1995)
Birdcage, The (1996)
Brothers McMullen, The (1995)
Get Shorty (1995)
Babe (1995)
French Twist (Gazon maudit) (1995)
Mighty Aphrodite (1995)
Latent Feature 1
Latent Feature 2
Average Rating

Method 2:

The second method we explored was similar to the first method with additional bias terms incorporated. It used the same code as Method 1 with a couple of adjustments.

The bias terms a and b respectively represent the offset that each user and movie has from the average. Each movie and user has their own bias with this method. This offset allows the model to more accurately represent the data that it is working with, and not factor in the variance of how specific users rate movies and also not factor in the variance in how a movie is received. For example, certain users could be very critical of all movies, giving low ratings, which could become a factor in their recommendations if not dealt with. The term μ also helps clean up the data by subtracting the global bias μ in the learning objective. In theory, this method should perform better than the first method as it allows the model to perform more effectively by cleaning up factors that should not be considered.

The learning objective with the bias terms is as follows (μ is the global bias):

$$\arg\min_{U,V,a,b} \frac{\lambda}{2}\left(\|U\|_F^2 + \|V\|_F^2 + \|a\|_F^2 + \|b\|_F^2\right) + \sum_{i,j}\left((Y_{ij} - \mu) - (u_i^T v_j + a_i + b_j)\right)^2$$

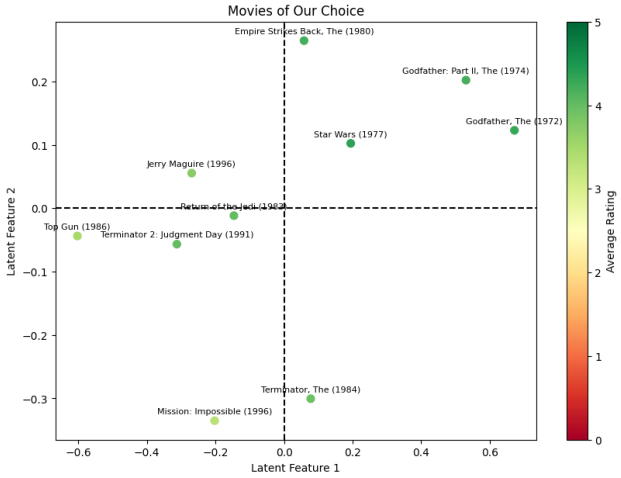
The gradients for the bias terms a and b are as follows:

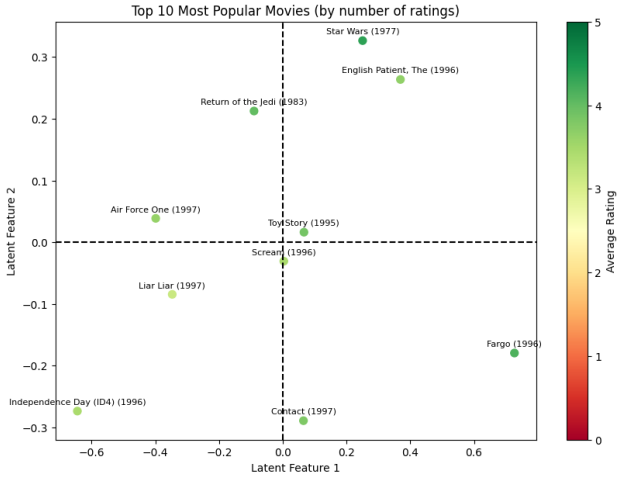
$$\partial a_i = \lambda a_i - (Y_{ij} - \mu - u_i^T v_j - a_i - b_j)$$
$$\partial b_j = \lambda b_j - (Y_{ij} - \mu - u_i^T v_j - a_i - b_j)$$

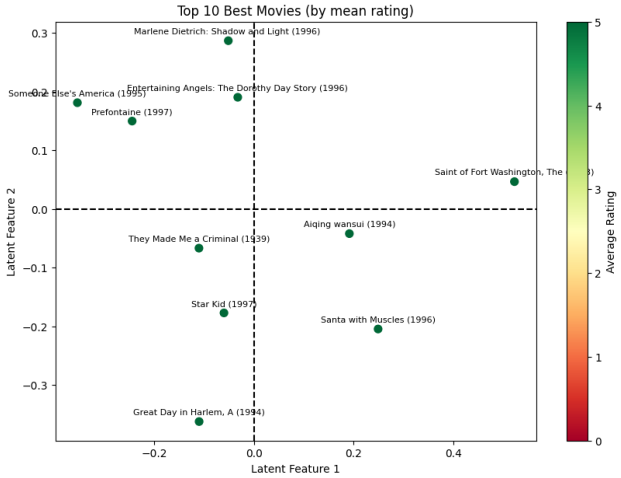
The gradients for U and V also needed to be changed:

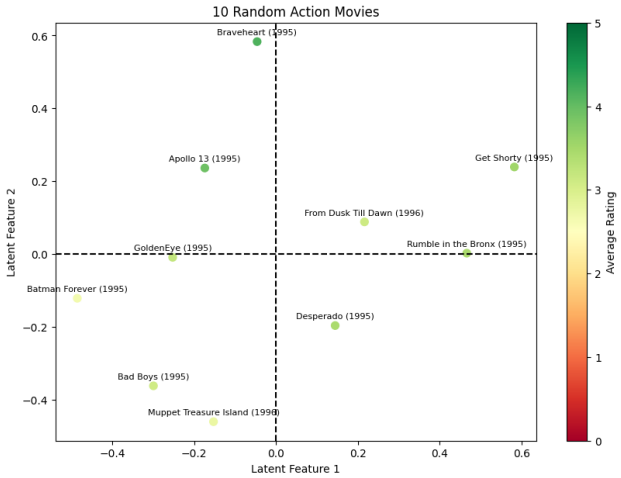
$$\partial u_i = \lambda u_i - v_j(Y_{ij} - u_i^T v_j - a_i - b_j)$$
$$\partial v_i = \lambda v_i - u_j(Y_{ij} - u_i^T v_j - a_i - b_j)$$

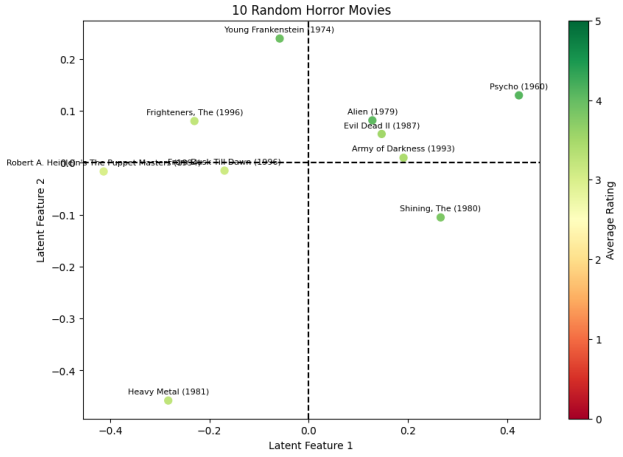
Visualizations:

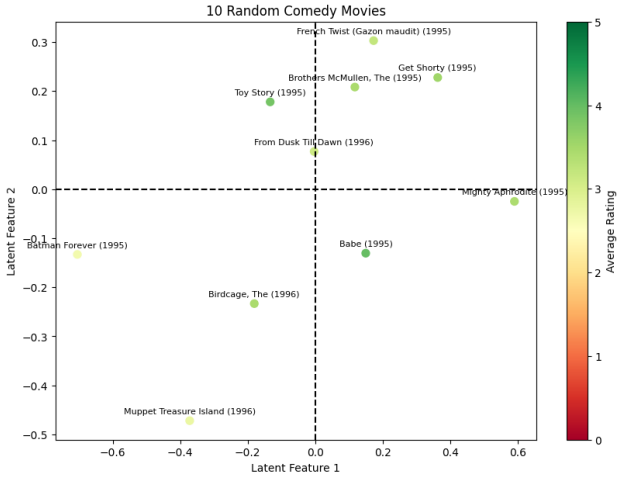





Top 10 Best Movies (by mean rating)
Marlene Dietrich: Shadow and Light (1996)
Entertaining Angels: The Dorothy Day Story (1996)
Someone Else's America (1995)
Prefontaine (1997)
Saint of Fort Washington, The
Aiqing wansui (1994)
They Made Me a Criminal (1939)
Star Kid (1997)
Santa with Muscles (1996)
Great Day in Harlem, A (1994)
Latent Feature 1
Latent Feature 2
Average Rating


10 Random Action Movies
Braveheart (1995)
Apollo 13 (1995)
Get Shorty (1995)
From Dusk Till Dawn (1996)
GoldenEye (1995)
Rumble in the Bronx (1995)
Batman Forever (1995)
Desperado (1995)
Bad Boys (1995)
Muppet Treasure Island (1996)
Latent Feature 1
Latent Feature 2
Average Rating

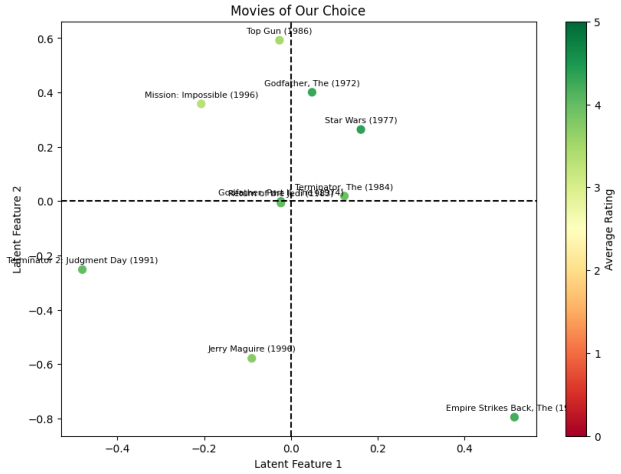

10 Random Horror Movies
Young Frankenstein (1974)
Psycho (1960)
Frighteners, The (1996)
Alien (1979)
Evil Dead II (1987)
Army of Darkness (1993)
Shining, The (1980)
Heavy Metal (1981)
Latent Feature 1
Latent Feature 2
Average Rating

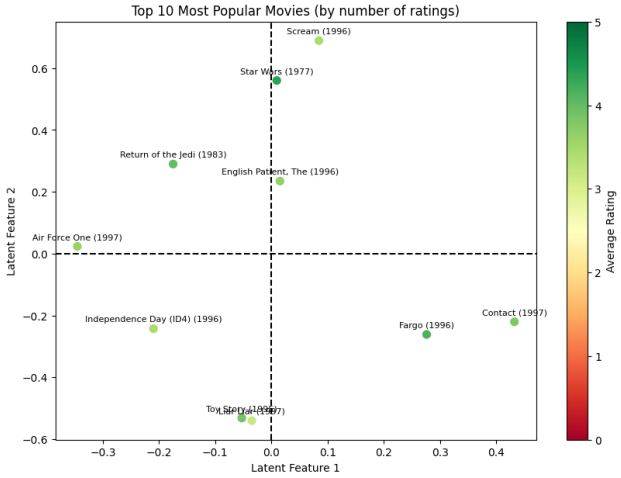

10 Random Comedy Movies
French Twist (Gazon maudit) (1995)
Get Shorty (1995)
Brothers McMullen, The (1995)
Toy Story (1995)
From Dusk Till Dawn (1996)
Batman Forever (1995)
Babe (1995)
Birdcage, The (1996)
Muppet Treasure Island (1996)
Latent Feature 1
Latent Feature 2
Average Rating

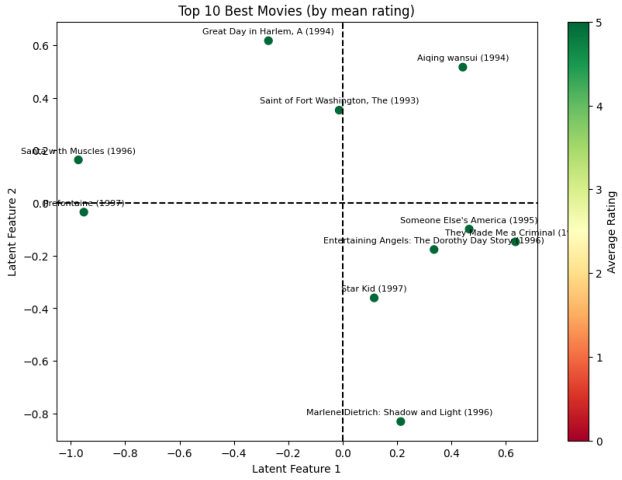
Method 3: Surprise Package

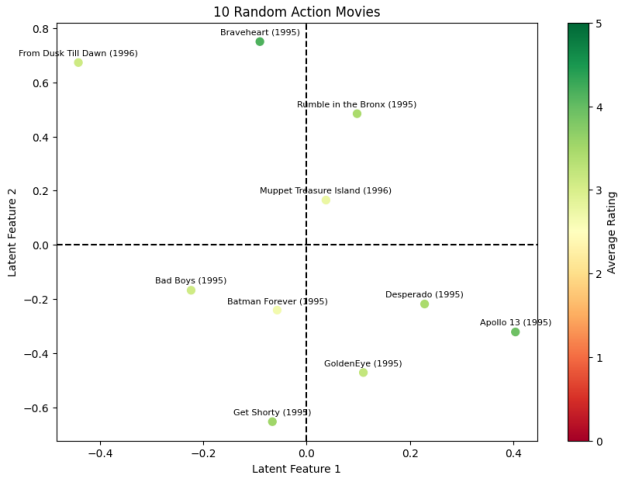
Our final off-the-shelf implementation was to use the Surprise Python Package for Collaborative Filtering. We instantiated a Reader object to parse the dataset from the original pandas dataframe. The dataset was then fully built into a training set. We then created an SVD model to factor a user-item rating matrix into two lower-dimensional matrices whose product approximates the original matrix. We specified n_epochs = 300, n_factors = 20, lr_all (learning rate) = 0.05, and reg_all (regularization term) = 0.1. It is important to note that we chose these values to be consistent with Methods 1 and 2, so that we can better analyze the similarities and differences of each method. After fitting the model to the training data, it was then used to predict ratings for the test set. These predictions were then evaluated using the mean squared error (MSE) between the predicted and actual ratings, providing a measure of the model's accuracy. We extracted the learned user (U) and item latent feature (V) matrices from the fitted SVD model. These matrices represent the underlying user preferences and item characteristics in a reduced-dimensional space, which were then used to visualize our results.

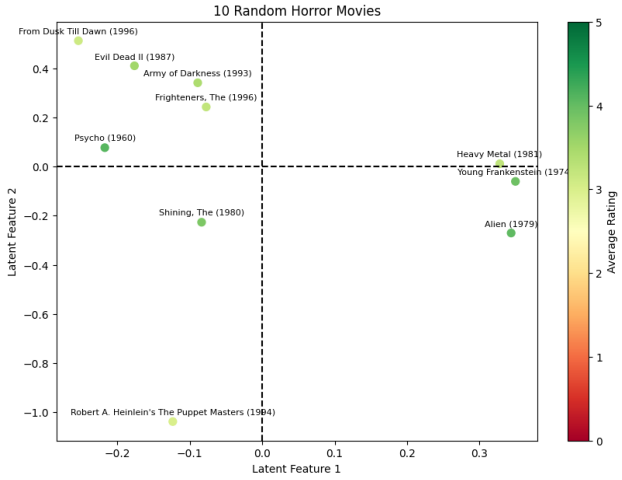
Visualizations:

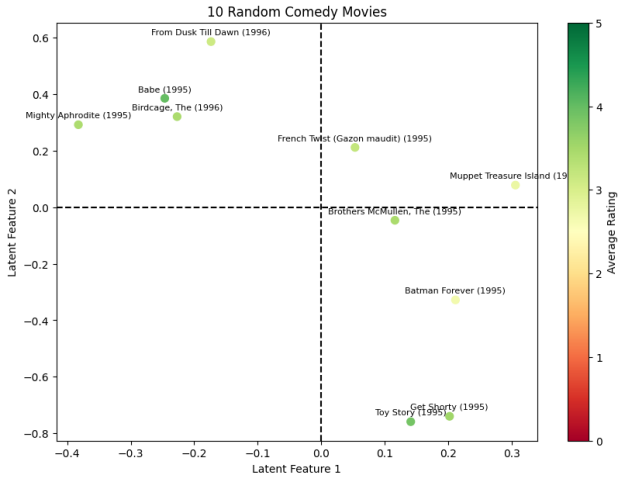





Top 10 Best Movies (by mean rating)
Great Day in Harlem, A (1994)
Aiqing wansui (1994)
Saint of Fort Washington, The (1993)
Someone Else's America (1995)
Entertaining Angels: The Dorothy Day Story (1996)
Star Kid (1997)
Marlene Dietrich: Shadow and Light (1996)
Latent Feature 1
Latent Feature 2
Average Rating


10 Random Action Movies
Braveheart (1995)
From Dusk Till Dawn (1996)
Rumble in the Bronx (1995)
Muppet Treasure Island (1996)
Bad Boys (1995)
Desperado (1995)
Apollo 13 (1995)
GoldenEye (1995)
Get Shorty (1995)
Latent Feature 1
Latent Feature 2
Average Rating


10 Random Horror Movies
From Dusk Till Dawn (1996)
Evil Dead II (1987)
Army of Darkness (1993)
Frighteners, The (1996)
Psycho (1960)
Heavy Metal (1981)
Young Frankenstein (1974
Shining, The (1980)
Alien (1979)
Robert A. Heinlein's The Puppet Masters (1994)
Latent Feature 1
Latent Feature 2
Average Rating


10 Random Comedy Movies
From Dusk Till Dawn (1996)
Babe (1995)
Birdcage, The (1996)
Mighty Aphrodite (1995)
French Twist (Gazon maudit) (1995)
Muppet Treasure Island (19
Brothers McMullen, The (1995)
Batman Forever (1995)
Get Shorty (1995)
Toy Story (1995)
Latent Feature 1
Latent Feature 2
Average Rating

Method Comparisons

Method 1 achieved an MSE loss of 0.53, method 2 achieved an MSE loss of 0.45, and method 3 achieved an MSE loss of 0.96 on the test set. Comparing methods 1 and 2, it makes sense that method 2 performed better than method 1. Incorporating a bias term that takes into account user and rating patterns that continuously gets updated would naturally improve the models performance and be able to fill in empty gaps better than without a bias term. However, it is surprising that method 3 performed significantly worse than methods 1 and 2. The off-the-shelf SVD algorithm works in a similar way to our manual implementations.

General Trends

In general, we noticed that movies with similar trends, such as genre, lead actor, and average rating, tend to group together in the visualizations. This is as we'd expect because the objective of the problem is to use latent factors to group similar movies together and give recommendations to users. So, it is a good sign that the two most prominent latent factors group objectively similar movies together.

Popular Movies vs Best Movies

It is important to note that we normalized the points of each graph we created by subtracting the mean of the points on each axis from each point on each axis (ex. all 10 of the feature 1 values had the feature 1 mean of those 10 points subtracted from them). This made it easier to see the trends within the subset we selected, but made it more difficult to compare across different graphs, since the normalization was specific to the graph. When examining the 10 most popular movies, we must take into account that these movies would have the most reviews and ratings in our input matrix. When examining the 10 best movies, we must take into account that these movies all have very high mean ratings. That being said, we would expect a strong trend in our visualizations. For method 1, we can see that similar movies seem to be grouped near each other (such as the two Star Wars movies) and the mean ratings seem to follow a trend (increasing from left to right). Many of the best movies are grouped together , below feature 2 = 0. A similar, but not as strong, trend is seen in the visualizations for method 2 for the most popular movies. The best movies are noticeably more spread out though. However, for method 3, there seems to be even less of a trend. Even the mean ratings seem to not have much correlation for the popular movies This makes sense given our results, but it still quite surprising. Regardless of popular vs best or which method is used, all movies have relatively good mean ratings, so the rating trend may not be the best indicator of the relationship between movies.

Action vs Horror vs Comedy

For method 1, the action movies are largely spread out, but only one movie is in the bottom left region. This is very different when we look at the action movie graphs for methods 2 and 3. In these graphs, we see the movies more spread out, and there is no correlation for the same movies being in the same region of the graph. For example, method 1 has Apollo 13 in the top right region, method 2 has Apollo 13 in the top left region, and method 3 has Apollo 13 in the bottom right region. For comedy movies, method 1 categorizes them primarily in the top left and bottom right regions. Method 2 does the opposite as it has them mostly in the top right and bottom left regions. Method 3 follows method 1 and categorizes them mostly in the top left and bottom right regions. Again, each movie is shown in a different region of the graph depending on the method used. Finally, for horror movies, method 1 has most of the movies in the top left, bottom

left, and bottom right regions. Method 2 categorizes them primarily in the top half of the graph. Method 3 primarily categorizes them in the left region of the graph with the exception of a few movies. Again, we fail to see any correlation between the methods as to where they categorize particular movies. Additionally, regardless of the genre, we can see many different average ratings which makes sense because people have different tastes.

4 Piazza Post

Link: Piazza Post

References

1. Koren, Y., Bell, R.,& Volinsky, C. (2009). Matrix Factorization Techniques for Recommender Systems Computer, (8), 30-37.

2. Herlocker, J. L., Konstan, J. A., Borchers, A., & Riedl, J. (1999, August). An algorithmic framework for performing collaborative filtering. In Proceedings of the 22nd annual international ACM SIGIR conference on Research and development in information retrieval (pp. 230-237). ACM.