

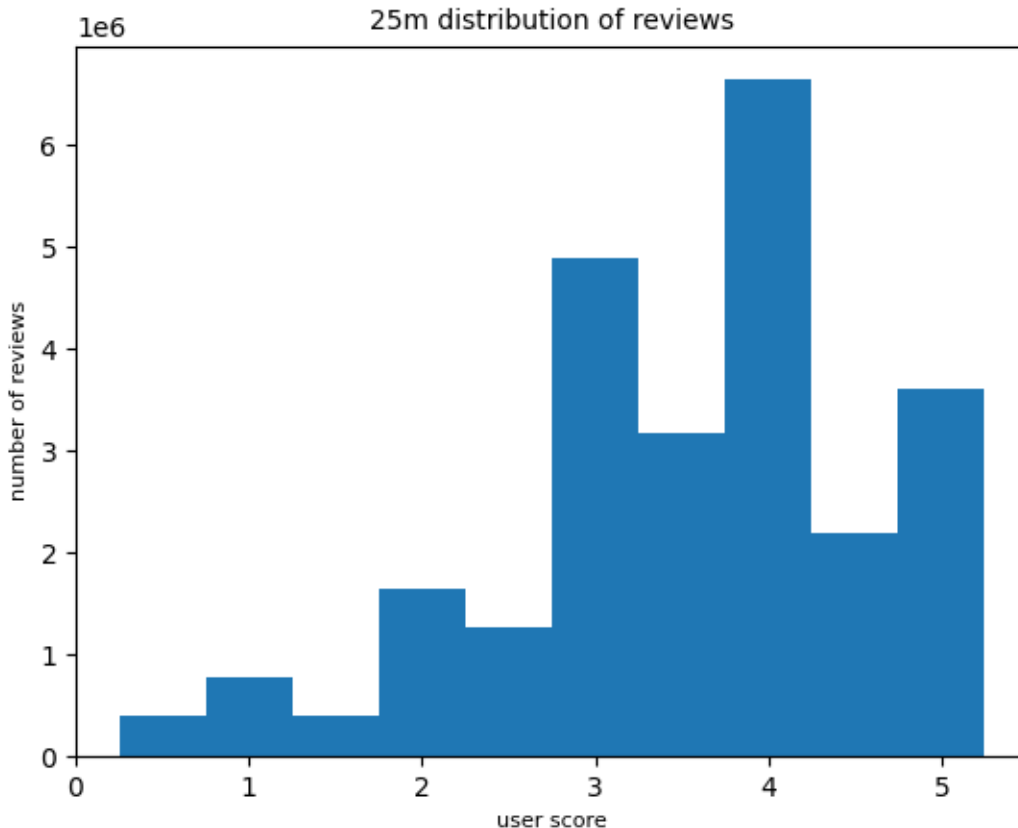
1 Running ASD Matrix Completion on MovieLens 25m dataset

The movielens 25m dataset contains 25 million total ratings by 162,000 users for 62000 movies. The ratings.csv file provided is contiguous for users but not movies, in the interest of saving memory the ratings table was augmented with the movie row number taken from movies.csv. The algorithm still works but it wastes time and memory generating ratings for movies that do not exist. Similarly user ratings were converted into z scores by computing the mean and standard deviation of each users ratings. This process was performed in SQL. The userIds, movie row, and z score were then used to populate two sparse matrices: a training set containing 90% of the entries and a testing set containing the other 10%.

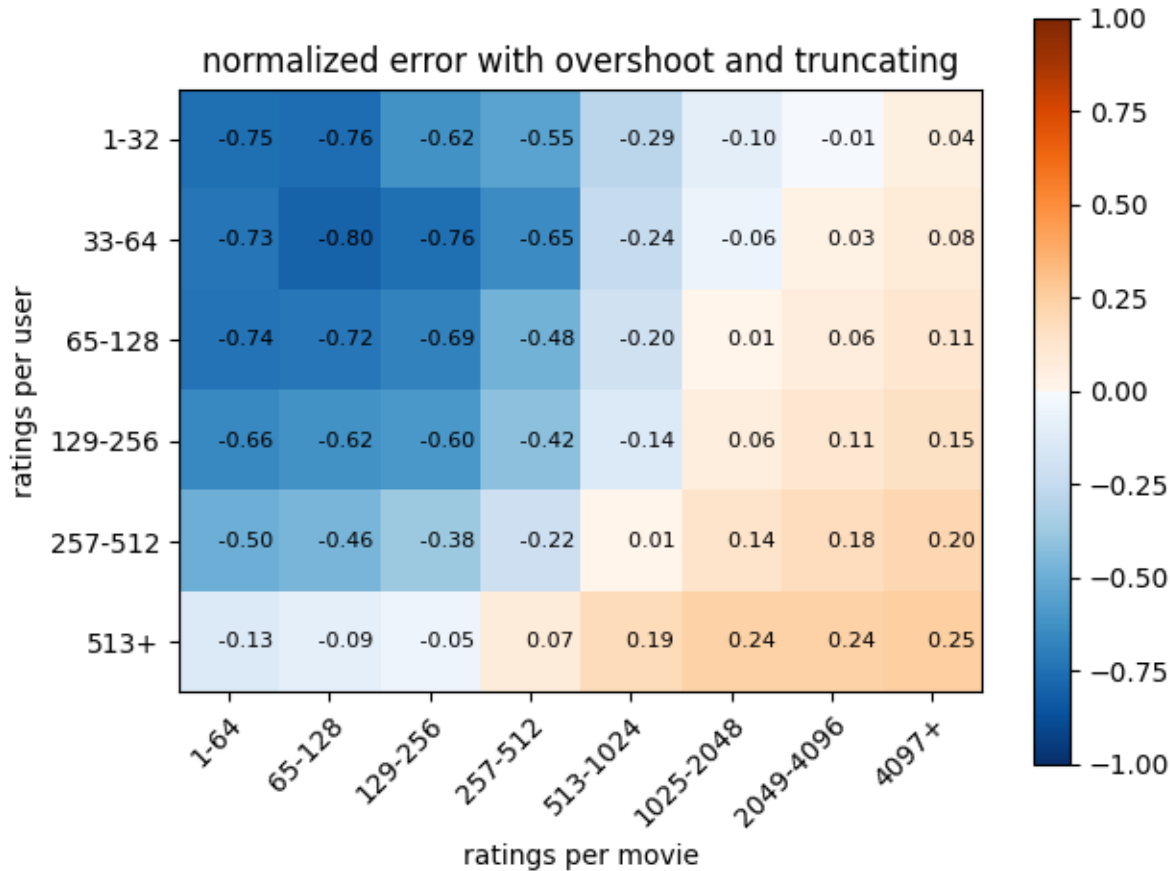
The algorithm was allowed to perform 2000 iterations with rank 32, the result was then used as the starting seed for 3000 iterations at rank 36. This took 18 and 75 minutes respectively. Looking at the singular value decomposition it was clear that the movie features were converging must faster than the user preferences. This suggests that rebalancing the individual iterations of X and Y from one to one to something more proportional with the dimensions of the data would be a minor performance increase. The algorithm began overfitting the data before stabilizing the user preferences which suggests that a rank of 64 is too high. The goal however was only to get a qualitative sense of performance so there was no deep search of the parameter hyperspace.

2 Error Analysis

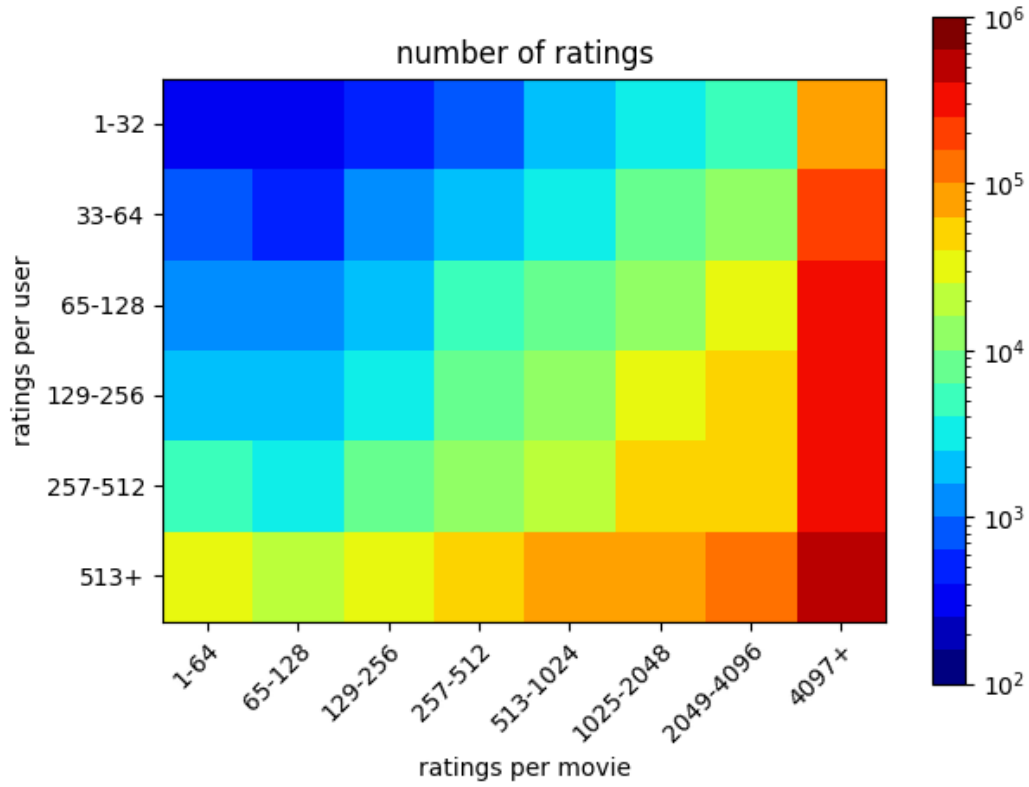
The movielens dataset has a very strong selection bias towards positive ratings. The explanation is fairly straightforward, users for the most part only rate movies they finish watching and they don't even start watching movies that they don't expect to like. As a result the algorithm picks up a heavy positive bias towards any movie recommendation. Despite this, the algorithm was generally able to catch on to the tastes of individual users. The disparity between postive and negative ratings is seen here:



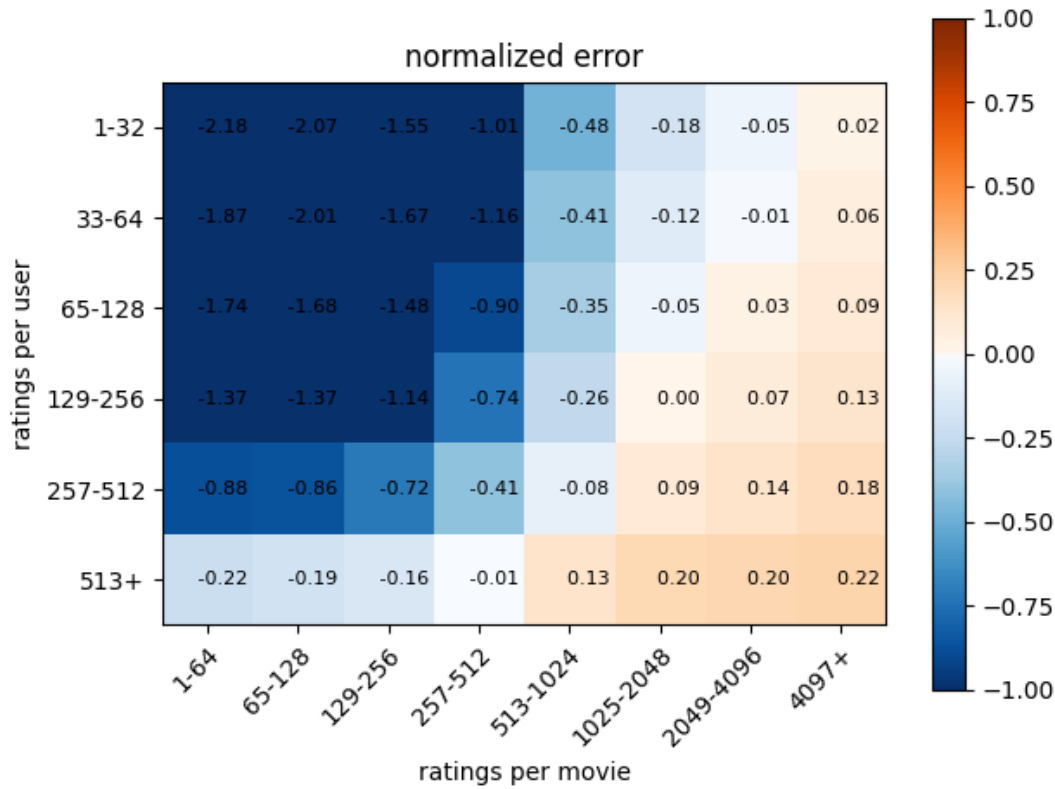
The error between the training and testing set was roughly equal so we will evaluate error for the testing set exclusively. The following heat map breaks each measurement (a rating for a movie) into buckets according to how many times the movie was reviewed (x axis) and how many other movies the user has reviewed (y axis). Essentially the y axis shows confidence in user preferences (ascending from the top) and the x axis shows confidence in movie features (ascending from the left). Ratings with z scores greater than 2 in magnitude were truncated to 2 and a buffer window of half the true z score was allotted for overshooting. Since we are primary concerned with whether or not a user likes or dislikes a movie, overshooting our estimate (assume love when like or hate when dislike) is much less harmful than undershooting. Since our model has rank 64, any user or movie with fewer ratings than that will be underdetermined. Finally the errors are normalized around the rank 1 hypothesis of ‘every user think every movie is average’ which amounts to guessing a value of 0 for every movie after normalization. An error score of 0 means the model is no better than guessing for that cluster, a score of 1 means that the model is perfect, and any negative score means that the model is worse than guessing average and should not be used.



This type of graph helps us determine how many ratings a movie has to have before we can feel confident recommending it to a user. If we were interested in serving recommendations in the low signal zones we would be better off building a mipmap of progressively simpler models e.g ranks 32,16,8,4 and then using the best model for each zone. Note that reducing the rank of our existing model by just removing feature/preference columns does NOT work. We should also note that the distribution of ratings is not uniform. In fact 90% of our ratings fell into the good region of the heat map! The distribution of ratings can be seen in this sister heatmap:



Finally the error map without all truncating or reducing penalty for overshooting is provided here, those factors did surprisingly little in the regions in which we would want to use our model in the first place:



3 Example Outputs

User 57231 - 70 ratings			
Their top picks		Our recommendations (2048+ reviewed)	
Title	Genres	Title	Genres
The Hunger Games: Mockingjay - Part 1 (2014)	Adventure, Sci-Fi, Thriller	Suicide Squad (2016)	Action, Crime, Sci-Fi
Rogue One: A Star Wars Story (2016)	Action, Adventure, Fantasy, Sci-Fi	Avengers: Infinity War - Part II (2019)	Action, Adventure, Sci-Fi
WALL-E (2008)	Adventure, Animation, Children, Romance, Sci-Fi	The Hobbit: The Battle of the Five Armies (2014)	Adventure, Fantasy
Up (2009)	Adventure, Animation, Children, Drama	Transformers: Dark of the Moon (2011)	Action, Adventure, Sci-Fi, War, IMAX
Bohemian Rhapsody (2018)	Drama	Batman v Superman: Dawn of Justice (2016)	Action, Adventure, Fantasy, Sci-Fi
How to Train Your Dragon (2010)	Adventure, Animation, Children, Fantasy, IMAX	Transformers: Revenge of the Fallen (2009)	Action, Adventure, Sci-Fi, IMAX
Coco (2017)	Adventure, Animation, Children	Ready Player One	Action, Sci-Fi, Thriller
The LEGO Ninjago Movie (2017)	Action, Animation, Comedy	Hobbit: The Desolation of Smaug, The (2013)	Adventure, Fantasy, IMAX
Naruto Shippuden the Movie: Blood Prison (2011)	Action, Animation, Comedy, Horror, Mystery, Thriller	Hobbit: An Unexpected Journey, The (2012)	Adventure, Fantasy, IMAX
Logan (2017)	Action, Sci-Fi	Transformers (2007)	Action, Sci-Fi, Thriller, IMAX
Interstellar (2014)	Sci-Fi, IMAX	Dreamcatcher (2003)	Drama, Horror, Sci-Fi, Thriller
Dark Knight, The (2008)	Action, Crime, Drama, IMAX	Avengers: Age of Ultron (2015)	Action, Adventure, Sci-Fi
Moana (2016)	Adventure, Animation, Children, Comedy, Fantasy	Avengers, The (2012)	Action, Adventure, Sci-Fi, IMAX
Guardians of the Galaxy (2014)	Action, Adventure, Sci-Fi	Law Abiding Citizen (2009)	Drama, Thriller
Big Hero 6 (2014)	Action, Animation, Comedy	Wolverine, The (2013)	Action, Adventure, Fantasy, Sci-Fi
Star Wars: Episode IV - A New Hope (1977)	Action, Adventure, Sci-Fi	Green Lantern (2011)	Action, Adventure, Sci-Fi
Avengers: Infinity War - Part I (2018)	Action, Adventure, Sci-Fi	Lucy (2014)	Action, Sci-Fi
Thor: Ragnarok (2017)	Action, Adventure, Sci-Fi	Iron Man 3 (2013)	Action, Sci-Fi, Thriller, IMAX
Arrival (2016)	Sci-Fi	Robin Hood (2010)	Action, Adventure, Drama, Romance, War
Naruto Shippuden the Movie: Road to Ninja (2012)	Action, Adventure, Animation, Fantasy	A-Team, The (2010)	Action, Comedy, Thriller

User 161083 - 256 ratings			
Their top picks		Our recommendations (1024+ reviewed)	
Title	Genres	Title	Genres
Schindler's List (1993)	Drama, War	Reds (1981)	Drama, Romance
Thelma & Louise (1991)	Adventure, Crime, Drama	Straw Dogs (1971)	Drama, Thriller
Airplane! (1980)	Comedy	Call Me by Your Name (2017)	Drama, Romance
Total Recall (1990)	Action, Adventure, Sci-Fi, Thriller	Beauty and the Beast (La belle et la bête) (1946)	Drama, Fantasy
Nightmare Before Christmas, The (1993)	Animation, Children, Fantasy, Musical	Cabin Fever (2002)	Horror, Thriller
Being John Malkovich (1999)	Comedy, Drama, Fantasy	Kicking and Screaming (1995)	Comedy, Drama
Howards End (1992)	Drama	The Hunger (1983)	Horror
Young Frankenstein (1974)	Comedy, Fantasy	Shall We Dance? (2004)	Comedy, Romance
Remains of the Day, The (1993)	Drama, Romance	Life (2017)	Horror, Sci-Fi, Thriller
Iron Giant, The (1999)	Adventure, Animation, Children, Drama, Sci-Fi	Dogtooth (Kynodontas) (2009)	Drama
Philadelphia (1993)	Drama	Peeping Tom (1960)	Drama, Horror, Thriller
Aliens (1986)	Action, Adventure, Horror, Sci-Fi	Carnal Knowledge (1971)	Comedy, Drama
Addams Family Values (1993)	Children, Comedy, Fantasy	Titus (1999)	Drama
12 Angry Men (1957)	Drama	Andalusian Dog, An (Chien andalous, Un) (1929)	Fantasy
Chicken Run (2000)	Animation, Children, Comedy	Millions (2004)	Children, Comedy, Crime, Drama, Fantasy
Pi (1998)	Drama, Sci-Fi, Thriller	Slither (2006)	Comedy, Horror, Sci-Fi
Contact (1997)	Drama, Sci-Fi	Walkabout (1971)	Adventure, Drama
South Park: Bigger, Longer and Uncut (1999)	Animation, Comedy, Musical	Naked (1993)	Drama
Adventures of Priscilla, Queen of the Desert, The (1994)	Comedy, Drama	Everest (2015)	Adventure, Drama, Thriller
Elizabeth (1998)	Drama	Melancholia (2011)	Drama, Sci-Fi
2001: A Space Odyssey (1968)	Adventure, Drama, Sci-Fi	Twilight Saga: Breaking Dawn - Part 1, The (2011)	Adventure, Drama, Fantasy, Romance
Citizen Kane (1941)	Drama, Mystery	Sweetest Thing, The (2002)	Comedy, Romance
Rosemary's Baby (1968)	Drama, Horror, Thriller	Quiet American, The (2002)	Drama, Thriller, War
Dark City (1998)	Adventure, Film-Noir, Sci-Fi, Thriller	Night Watch (Nochnoy dozor) (2004)	Action, Fantasy, Horror, Mystery, Sci-Fi, Thriller

User 162519 - 522 ratings			
Their top picks		Our recommendations (512+ reviewed)	
Title	Genres	Title	Genres
Toy Story (1995)	Adventure, Animation, Children, Comedy, Fantasy	Zero Theorem, The (2013)	Drama, Fantasy, Sci-Fi
This Is Spinal Tap (1984)	Comedy	Spring Breakers (2013)	Comedy, Crime, Drama
Cape Fear (1962)	Crime, Drama, Thriller	Tetsuo, the Ironman (Tetsuo) (1988)	Action, Horror, Sci-Fi, Thriller
Die Hard 2 (1990)	Action, Adventure, Thriller	Greenberg (2010)	Comedy, Drama
Batman Returns (1992)	Action, Crime	Kid Stays in the Picture, The (2002)	Documentary
Jaws (1975)	Action, Horror	Sweet Smell of Success (1957)	Drama, Film-Noir
Chasing Amy (1997)	Comedy, Drama, Romance	Wolf Man, The (1941)	Drama, Fantasy, Horror
Boogie Nights (1997)	Drama	Innocents, The (1961)	Drama, Horror, Thriller
Big Lebowski, The (1998)	Comedy, Crime	Signal, The (2014)	Sci-Fi, Thriller
Out of Sight (1998)	Comedy, Crime, Drama, Romance, Thriller	Tale of Princess Kaguya, The (Kaguyahime no monogatari) (2013)	Animation, Drama, Fantasy
Pi (1998)	Drama, Sci-Fi, Thriller	New World, The (2005)	Adventure, Drama, Romance
Breakfast Club, The (1985)	Comedy, Drama	Boy and His Dog, A (1975)	Sci-Fi
Exorcist, The (1973)	Horror, Mystery	Spy Who Came in from the Cold, The (1965)	Drama, Thriller
Seven Samurai (Shichinin no samurai) (1954)	Action, Adventure, Drama	Won't You Be My Neighbor? (2018)	Documentary
Roger & Me (1989)	Documentary	One I Love, The (2014)	Comedy, Drama, Romance
Lord of the Rings, The (1978)	Adventure, Animation, Children, Fantasy	Land of the Lost (2009)	Action, Adventure, Comedy, Sci-Fi
Beetlejuice (1988)	Comedy, Fantasy	Bread and Tulips (Pane e tulipani) (2000)	Comedy, Drama, Romance
Ronin (1998)	Action, Crime, Thriller	Twilight Samurai, The (Tasogare Seibei) (2002)	Drama, Romance
Thing, The (1982)	Action, Horror, Sci-Fi, Thriller	Your Highness (2011)	Action, Adventure, Comedy, Fantasy
Player, The (1992)	Comedy, Crime, Drama	Lovely & Amazing (2001)	Comedy, Drama, Romance
Producers, The (1968)	Comedy	Ben-Hur: A Tale of the Christ (1925)	Adventure, Drama, Romance
Pleasantville (1998)	Comedy, Drama, Fantasy	Come and See (Idi i smotri) (1985)	Drama, War
Rushmore (1998)	Comedy, Drama	Incredible Shrinking Man, The (1957)	Sci-Fi, Thriller