**Initial Approach: (Correlation)**

Observations:

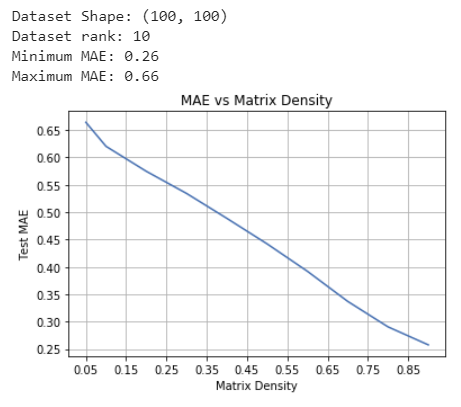
There is a positive correlation between rating counts and average rating, which means a movie that is rated by many people is more likely to have a higher rating

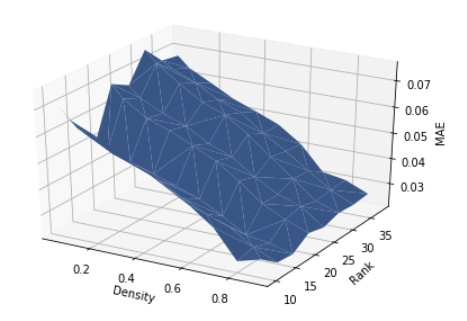
Most movies are highly uncorrelated () due to large sparsity

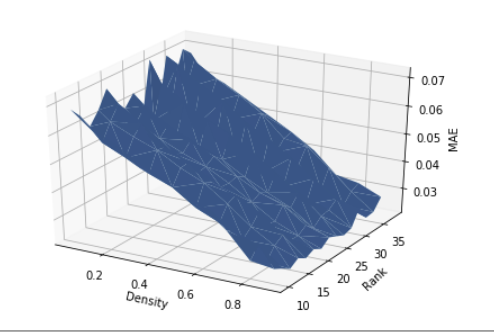
Even if sparsity was not a problem, the results are a combined overview of all users, thus allowing no clear way to incorporate specific user preference

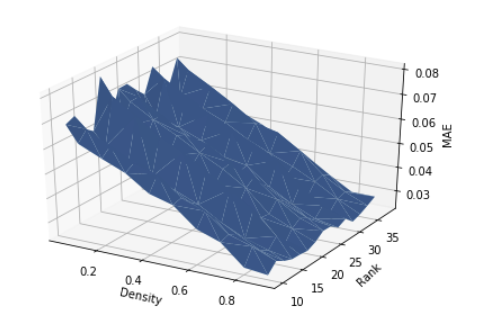
🡪 everyone gets the same recommendations

**Synthetic Data:**









Validity of Iterative SVD:

Density: 0.2

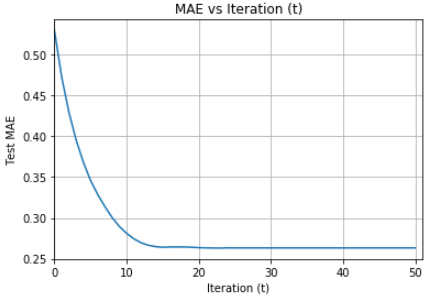
Train Num Count: 2026

Test Num Count: 7974  
Dataset Shape: (100, 100)

Dataset rank: 11

Maximum MAE: 0.53 (t = 1)

Minimum MAE: 0.26 (t = 50)



**SVD**

Part1: Matrix completion

MAE results on different approaches

Filled using zeros: 3.5805

First Iteration: 2.3107

Last Iteration: 1.0338

Filled using user means: 0.8293

First Iteration: 0.7486

Last Iteration: 0.7136

Filled using movie means: 0.7818

First Iteration: 0.7169

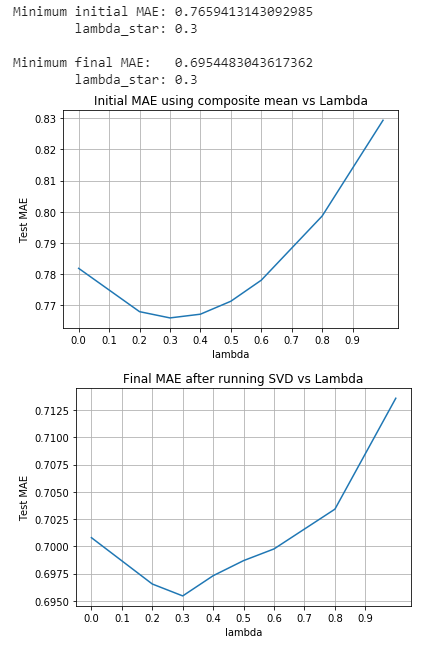
Last Iteration: 0.7008

Filled using average of both: 0.7713

First Iteration: 0.7118

Last Iteration: 0.6987

Part 2: Iterative SVD to tune weights between means



**Regression**

Features: Label: Rating

UserID

MovieID

Age

Occupation

Gender(OneHot)

Genre (MultiHot)

Incorporating only user info:

OLS:

Normal

Train MAE: 0.9272

Test MAE: 0.9272

Rounded

Train MAE: 0.9392

Test MAE: 0.9392

Ridge:

Normal

Train MAE: 0.9278

Test MAE: 0.9278

Rounded

Train MAE: 0.9415

Test MAE: 0.9415

Incorporating user and movie info:

OLS:

Normal

Train MAE: 0.8987

Test MAE: 0.8987

Rounded

Train MAE: 0.8956

Test MAE: 0.8956

Ridge:

Normal

Train MAE: 0.9002

Test MAE: 0.9002

Rounded

Train MAE: 0.8979

Test MAE: 0.8979

Observation:

* Both models are not able to perform well, or learn from the data whatsoever. The MAE values are worse the case of the regular matrix completion using the mean values that can be seen in SVD.
* Incorporating movie data as well as user data performed slightly better than using user data alone.
* Rounding results to nearest 0.5 has barely improved   
  (and in some cases worsened) our results.

**Implementing SVD, then Regression**

OLS:

Normal

Train MAE: 0.4018

Test MAE: 0.9583

Rounded

Train MAE: 0.3750

Test MAE: 0.9470

Ridge:

Normal

Train MAE: 0.4017

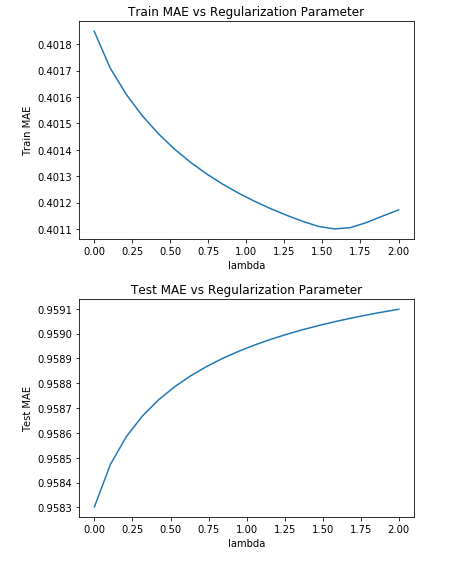
Train MAE: 0.9584

Rounded

Train MAE: 0.3749

Train MAE: 0.9470

Ridge Tuning hyper parameter



Observations:

* Model is overfitting
* Tuning regularization parameter in ridge regression does not impact performance.

🡪 The model is not learning