Matrix Completion via SVD in Recommendation Systems

MA614 Mini Project

Ranjeet Singh
2021EEB1203

May 14, 2025

Introduction

- Recommendation systems suggest items based on user preferences.
- MovieLens 100k dataset is used as benchmark.
- Since there are large number of empty entries, it becomes a challenge to recommend users.

Problem Definition

- Goal: Predict missing ratings in a sparse user-item matrix.
- Real-world ratings are often incomplete and sparse.
- Matrix completion helps improve the quality of the recommendation.

Dataset Overview

- MovieLens 100k dataset
- 943 users, 1682 movies
- 100,000 ratings
- Highly sparse user item matrix (943 x 1682)

Truncated SVD

Objective

Given a matrix $A \in \mathbb{R}^{m \times n}$, approximate it as:

$$A \approx U_k \Sigma_k V_k^T$$

where $k \ll \min(m, n)$

Steps

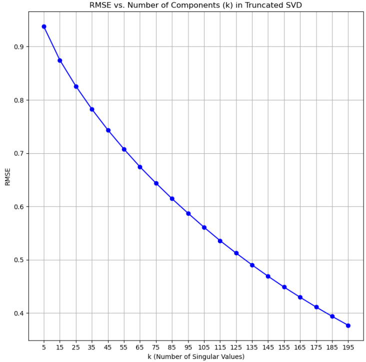
1. Compute $A^T A \in \mathbb{R}^{n \times n}$
2. Perform eigen-decomposition: $A^T A = V \Lambda V^T$
3. Retain top k eigenvectors and eigenvalues
4. Compute singular values: $\Sigma_k = \text{diag}(\sqrt{\lambda_1}, \ldots, \sqrt{\lambda_k})$
5. Compute $U_k = A V_k \Sigma_k^{-1}$
6. Reconstruct: $\hat{A} = U_k \Sigma_k V_k^T$

Algorithm Steps

1. Normalize user ratings (Mean or Z-score)
2. Replace NaNs with 0
3. Apply truncated SVD for some $k << min(m, n)$.
4. Reconstruct matrix and denormalize to obtain the completed matrix.

RMSE Evaluation

- RMSE $= \sqrt{\frac{1}{N}\sum(R_{ij} - \hat{R}_{ij})^2}$ [Only for non-empty values in the original matrix]
- Compare RMSE for different k



Matrix Completion Results: Mean Normalization

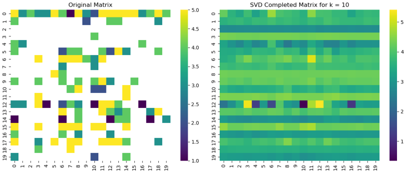


Figure: Heatmap of completed matrix for $k = 10$

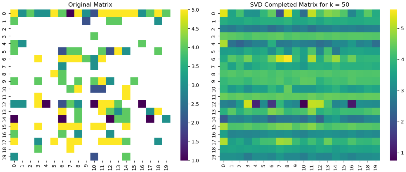


Figure: Heatmap of completed matrix for $k = 50$

Side-by-Side Comparison

Here we compare how the normalization techniques affect the completion matrix

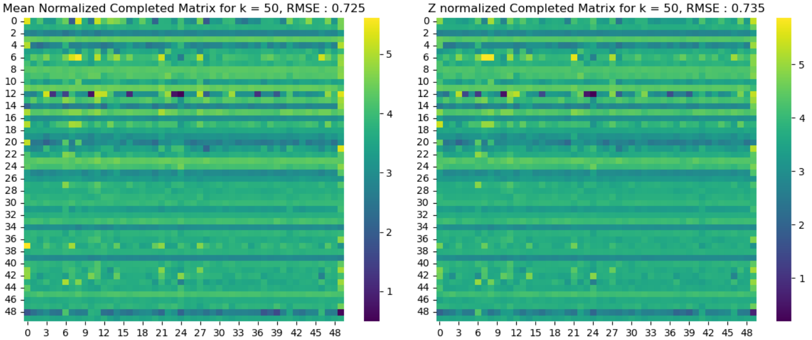


Figure: Comparison of Mean vs Z-score Normalized Matrix

Limitations

- Assumes linearity and low-rank structure
- Sensitive to outliers and noisy ratings
- Computational cost increases with data size

Possible Improvements

- Apply regularization to prevent overfitting
- Use Robust PCA or weighted matrix factorization
- Randomized SVD can be effectively used for larger sparse matrices

Conclusion

- Truncated SVD can effectively predict missing values
- Z-score normalization slightly improves robustness from outliers but might not work against noisy ratings.
- RMSE helps select optimal k

Thank You