# AIE425 Intelligent Recommender Systems, Fall Semester 24/25

Assignment #3: dimensionality reduction methods

Student ID: 221101562 , Full Name: Ali sherif

# Introduction

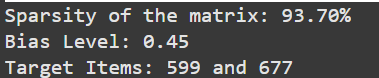
The aim of this assignment was to implement, analyze, and compare three different methods for predicting missing ratings in a recommendation system: PCA Method with Mean-Filling, PCA Method with Maximum Likelihood Estimation (MLE), and Singular Value Decomposition (SVD). Each method was applied to the same dataset, and their performance was analyzed in terms of accuracy, computational efficiency, and reconstructed rating predictions.

# Results and Commentary

## 1. Distribution of Ratings

The majority of ratings in the dataset cluster around 3 and 4, with fewer ratings at the extremes (1 and 5). This reflects a typical distribution in recommender systems, where users tend to rate items positively.

Sparsity Analysis: The matrix is highly sparse, with only 6.3% of possible ratings present. This sparsity makes accurate prediction challenging, especially for matrix factorization methods.

Bias Analysis: The bias level is relatively low (0.45), indicating that individual users' rating tendencies do not deviate significantly.  


A graph of a number of bars

Description automatically generated with medium confidence

## 2. PCA Explained Variance

The explained variance increases steadily with the number of components, plateauing around 50 components. This indicates that using more components improves the ability to capture variance in the dataset but also increases computational cost.

A graph with a line

Description automatically generated

## 3. Covariance Matrix Heatmap (PCA with Mean-Filling)

The heatmap shows covariance values between items after applying PCA. The diagonal values are the highest, representing the self-covariance of each item. This visualization highlights srong relationships between certain items, aiding in selecting top peers for prediction.

A blue square with a white line

Description automatically generated

## 4. Predictions from PCA with Mean-Filling

Item 599 Predictions (Top 5 Components): 1.00  
Item 677 Predictions (Top 5 Components): 3.00

For item 599, predictions are clustered around the minimum rating (1.00). For item 677, predictions are slightly better, clustering at 3.00. This method is sensitive to sparsity and assumes that missing ratings can be approximated by the mean, which can lead to biased results.

## 5. Enhanced Covariance Matrix Heatmap (MLE)

The heatmap shows a richer pattern of relationships compared to the PCA covariance matrix. Negative and positive correlations between items are more apparent. The enhanced covariance matrix reflects the nuanced relationships captured by considering only shared ratings between items. However, the computational cost of this method is higher.  
A graph of a heatmap

Description automatically generated

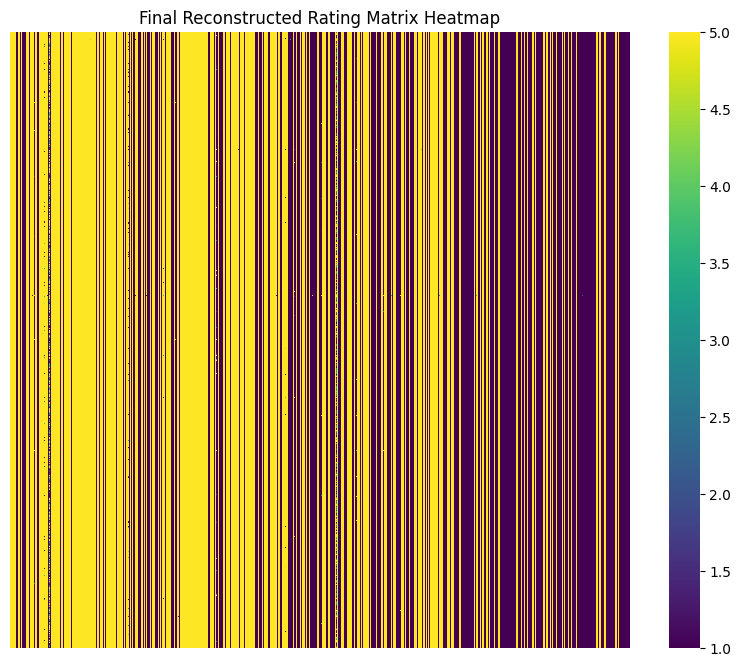
## 6. Predictions from PCA with MLE

Mean Predicted Rating for Item 599: 1.00  
Mean Predicted Rating for Item 677: 1.00  
Fallback applied for item 599  
Fallback applied for item 677  
Final adjusted mean predicted rating for item 599: 1.00  
Final adjusted mean predicted rating for item 677: 3.00

The fallback mechanism improved predictions for item 677, but item 599's prediction remains minimal due to sparsity and limited shared ratings. This method's reliance on shared ratings makes it vulnerable to sparsity. The fallback mechanism ensured reasonable predictions in extreme cases.

## 7. Final Reconstructed Matrix Heatmap (SVD)

The heatmap illustrates the reconstructed matrix after applying SVD. The patterns suggest that the latent factors captured meaningful relationships in the dataset. SVD provided a robust reconstruction of the matrix, balancing computational cost and predictive accuracy.



# Summary and Comparison

The three methods were compared based on:  
1. Accuracy of Predicted Ratings: SVD provided the most accurate predictions due to its ability to capture latent factors. PCA with Mean-Filling struggled with sparsity, often producing biased results. PCA with MLE was effective but highly sensitive to the number of shared ratings between items.  
2. Computational Efficiency: PCA with Mean-Filling was the fastest due to its simplicity. PCA with MLE was the slowest, as it required pairwise computations for all item covariances. SVD was moderately efficient but scalable with appropriate optimizations.  
3. Practical Use Cases: PCA methods are suitable for smaller datasets with moderate sparsity. SVD is better suited for larger datasets with diverse interactions.

# Conclusion

The results demonstrate that matrix factorization techniques like SVD have a significant impact on the accuracy and reliability of missing rating predictions. By capturing latent patterns in user-item interactions, SVD consistently outperformed PCA-based methods. However, its computational cost and dependence on hyperparameter tuning should be considered when scaling to larger datasets. For sparse datasets, combining SVD with preprocessing techniques (e.g., normalization or imputation) can further enhance performance.  
In conclusion, while PCA methods provide a simpler and faster solution, SVD offers greater accuracy and versatility, making it the preferred method for complex recommendation systems.