AlE425 Intelligent Recommender Systems, Fall Semester 24/25

Assignment #3: Dimensionality Reduction methods

Student ID : A20000375

Full Name: Reem Ahmed Gamal

### ****Summary and Comparison****

#### ****Part 1: Basic Collaborative Filtering****

* **Approach**: Basic collaborative filtering methods rely on computing the mean ratings, differences from the mean, and covariance between items.
* **Results**:
  + Predicted ratings were computed for the missing values using simple similarity measures between items.
  + Accuracy depends on the degree of sparsity in the dataset.
* **Pros**:
  + Straightforward to implement.
  + Requires minimal computational resources for small datasets.
* **Cons**:
  + Sensitive to sparsity in the dataset.
  + Predictions for items with very few or no ratings tend to be less reliable.
  + Assumes linear relationships between items/users.

#### ****Part 2: PCA Method with Maximum Likelihood Estimation****

* **Approach**: PCA was used to decompose the covariance matrix, capturing latent features from the data. Only overlapping ratings were considered for covariance computation, with the assumption that non-overlapping pairs have a covariance of 0.
* **Results**:
  + Covariance matrix was calculated with maximum likelihood estimates, and top-5 and top-10 peers were used to predict ratings for the target items (11 and 12).
  + The predictions improved over Part 1, particularly for items with substantial overlap with other items in the dataset.
* **Pros**:
  + Captures latent structures in the dataset, improving predictive power.
  + Can handle some sparsity effectively using only overlapping ratings for covariance.
* **Cons**:
  + Requires more computational resources than Part 1.
  + Sensitive to dimensionality reduction; using too few or too many features can affect accuracy.

#### ****Part 3: Singular Value Decomposition (SVD)****

* **Approach**: Truncated SVD was applied to the ratings matrix to reduce dimensionality and capture principal components. Missing ratings were replaced with the mean initially, and the reduced matrix was used to predict the ratings.
* **Results**:
  + The reconstructed rating matrix provided more accurate predictions for the missing ratings (items 11 and 12) than both Parts 1 and 2.
  + SVD-based predictions were consistent for both top-5 and top-10 peers.
* **Pros**:
  + Most robust among the three methods due to its ability to capture global patterns in the data.
  + Handles sparsity well and performs better for larger datasets.
  + Provides a principled way of reducing dimensionality.
* **Cons**:
  + Computationally intensive, particularly for large datasets.
  + Requires fine-tuning of the number of latent factors (k) to balance between overfitting and underfitting.

### ****Conclusion****

Matrix factorization methods, particularly **SVD**, have a significant impact on improving the accuracy of predicting missing ratings. While basic collaborative filtering is simpler and interpretable, its reliance on explicit similarities between items makes it less robust for sparse datasets. PCA improves predictions by identifying latent structures but is sensitive to the choice of the number of components.

**SVD emerges as the most effective method** due to its ability to handle sparsity and uncover latent factors that govern user-item interactions. However, its computational complexity can be a limitation for very large datasets.

**Impact of Matrix Factorization**:

* By reducing the dataset to a small number of latent factors, SVD simplifies the rating matrix while preserving its essential structure.
* This enables better generalization and prediction, particularly for users or items with limited historical data.