Final Report

Dataset

MovieLens dataset contains user-generated ratings for movies, information about movies, and information about users.
Our version of the dataset contains:

- 943 users
- 1682 items (movies)
- 100000 ratings

Model

The MatrixFactorization model is a collaborative filtering approach designed for recommendation systems. It comprises the following key components:

1. **Embedding Layers**:
   - Two embedding layers: one for users and one for movies.
   - Each layer learns latent representations (embeddings) for users and movies.
   - Embeddings aim to capture latent factors that influence user-item interactions.
2. **Forward Pass**:
   - The `forward` method performs matrix multiplication between user and movie embeddings.
   - Predicts interactions by computing the dot product of user and movie embeddings.
   - The model output represents the predicted ratings or interactions between users and movies.

Advantages

1. **Scalability**:
   - Efficient for large datasets due to its relatively simple architecture and use of embeddings.
2. **Interpretability**:
   - Embeddings provide interpretable latent factors (e.g., user preferences, item characteristics).

Disadvantages

1. **Cold Start Problem**:
   - Faces challenges in providing recommendations for new users or items without historical data.
2. **Limited Context**:
   - Relies solely on user-item interactions, ignoring contextual information (time, user demographics) that could enhance recommendations.

Evaluation Metric

- **Root Mean Squared Error (RMSE)**:
  - Used to evaluate the model's performance.
  - Lower RMSE values indicate better predictive performance.

This model achieves RMSE of **1.4752**.

Conclusion

The MatrixFactorization model is a foundational approach in collaborative filtering, leveraging user and item embeddings to predict user-item interactions. Its simplicity and interpretability make it a popular choice for recommendation systems. However, it faces challenges related to the cold start problem, and the lack of contextual information. RMSE serves as a useful metric for assessing the model's accuracy in predicting ratings, but other metrics and techniques might be necessary to address its limitations in real-world scenarios.