**Matrix Factorization Collaborative Filtering**

**Phân tích ma trận dựa trên nhân tử**

**Hệ thống gợi ý phim và bộ dữ liệu MovieLens**

* Trong phương pháp này, cần tìm một vector hệ số **w** tương ứng với mỗi user sao cho rating đã biết mà user đó cho item xấp xỉ với:

**y ≈ xw**

Trong đó, mỗi item được mô tả bằng một vector **x** được gọi là *item profile* (trong Content-based RS)

* Utility matrix **Y ≈ XW**

Trong đó, **X** là ma trận của toàn bộ item profiles, mỗi hàng tương ứng với 1 item; **W** là ma trận của toàn bộ user models, mỗi cột tương ứng với 1 user

* Việc xây dựng từng mô hình riêng lẻ cho mỗi user dẫn đến kết quả chưa tốt.
* Giả sử rằng ta không cần xây dựng từ trước các item profile **x** mà vector đặc trưng cho mỗi item này có thể được huấn luyện đồng thời với mô hình của mỗi user.
* Tối ưu cả **X** và **W**
* Y RMxN, X RMxK, W RKxN; K M, N (M, N lần lượt là số items và số users)
* Low-Rank Matrix Factorization
* **Latent features** (tính chất ẩn) mô tả sự liên quan giữa các items và users.
* Lần lượt đi tìm X và W khi thành phần còn lại được cố định.
* NBCF không yêu cầu việc learning quá nhiều, nhưng trong quá trình inference, cần đi tìm độ similarity của user đang xét với toàn bộ các users còn lại rồi suy ra kết quả. Ngược lại, việc learning của MF có thể hơi phức tạp nhưng việc inference đơn giản hơn.

1. **Hàm mất mát**

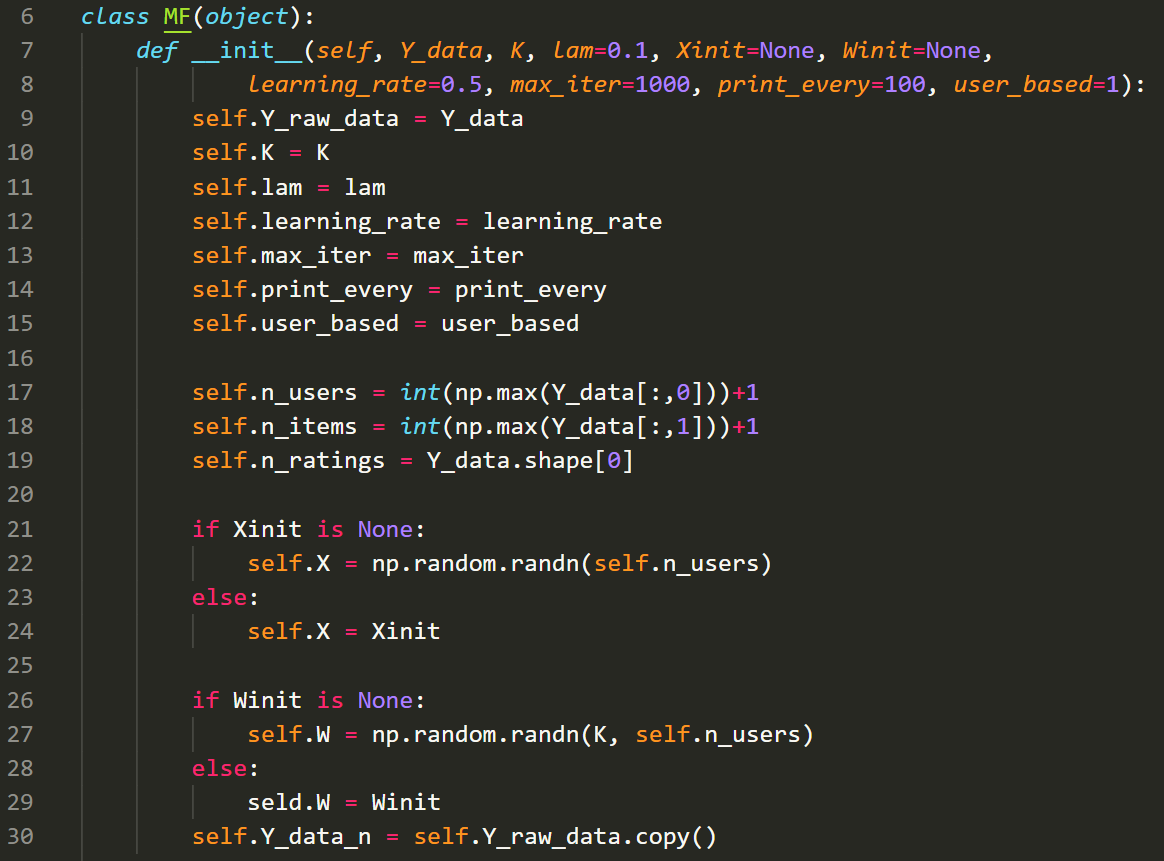
**…**

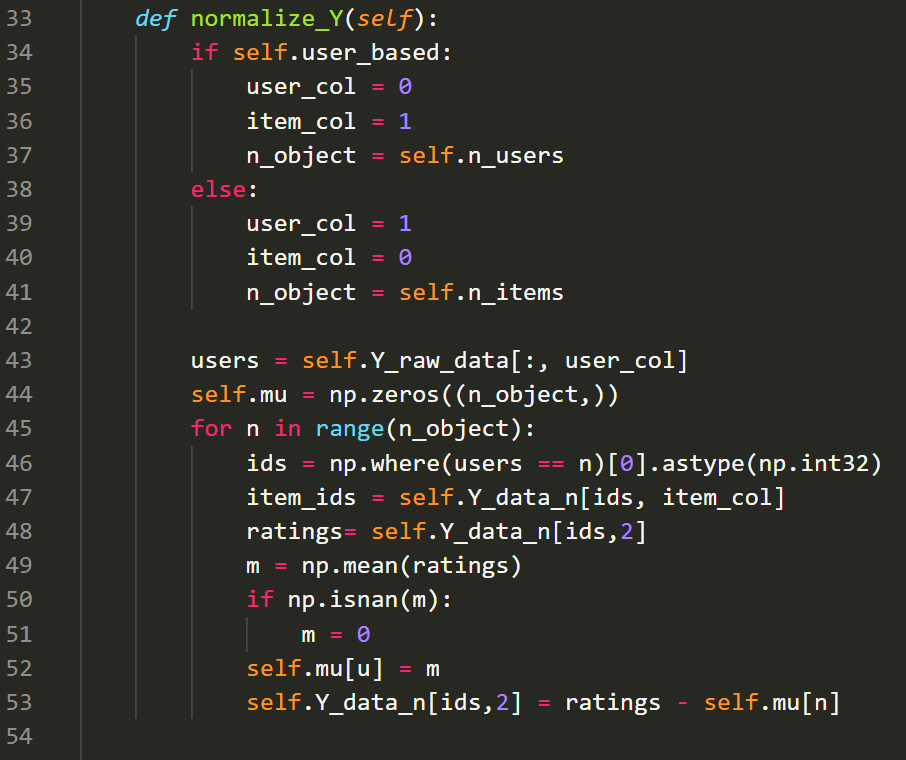
1. **Tối ưu hàm mất mát**

**…**

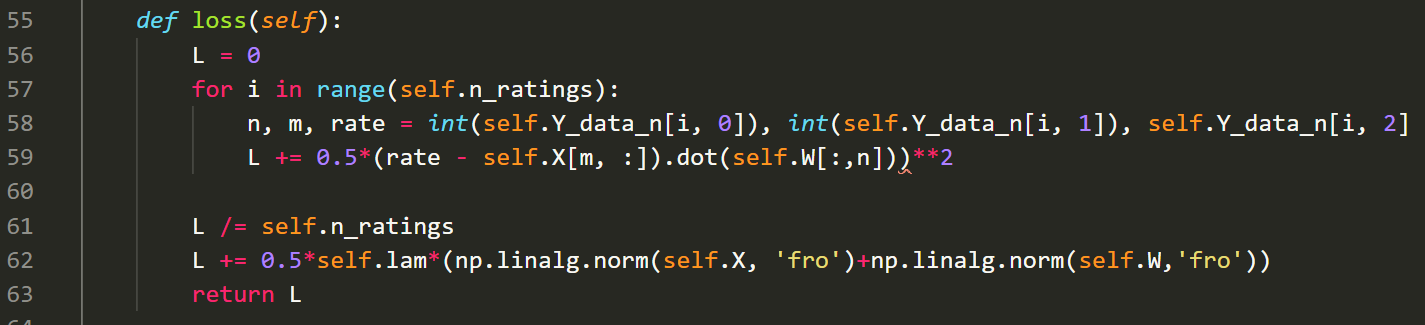
1. **Class MF**

* **Khởi tạo và chuẩn hóa**

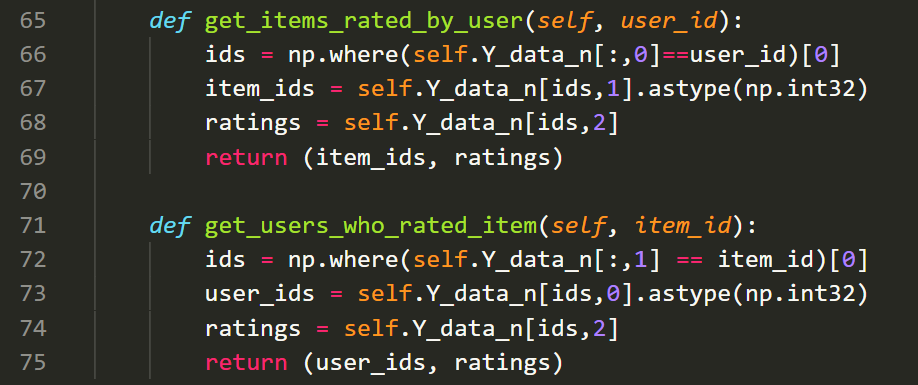
****

****

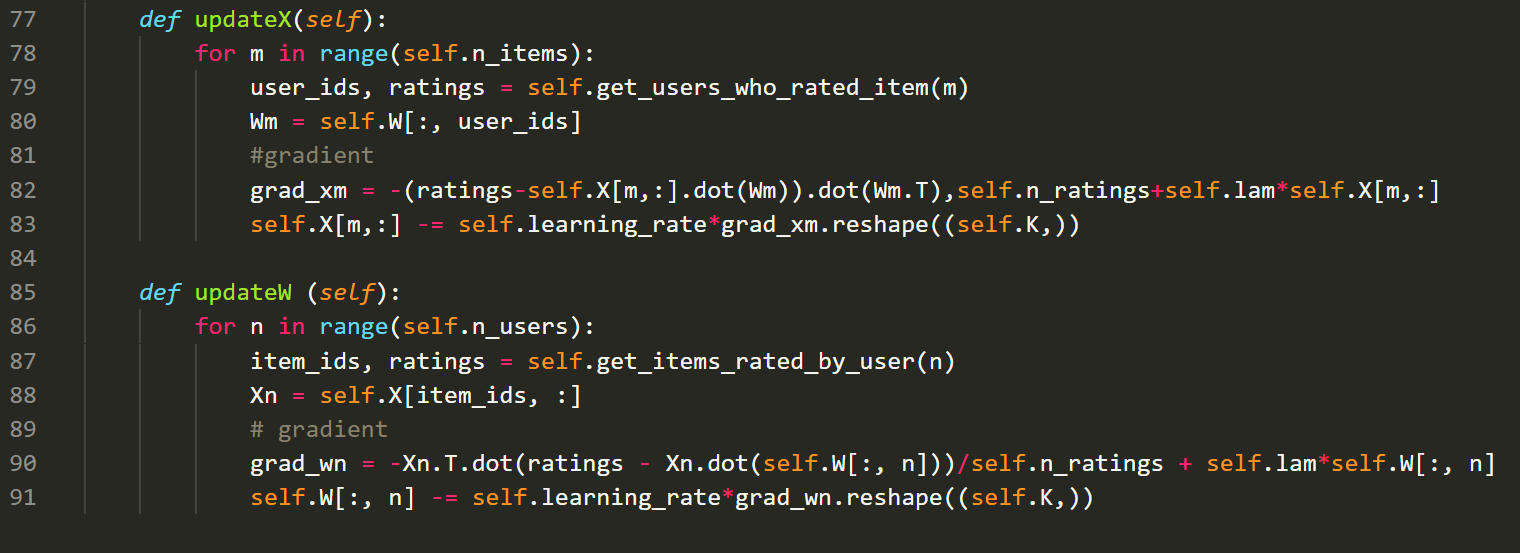
* **Tính giá trị hàm mất mát**

****

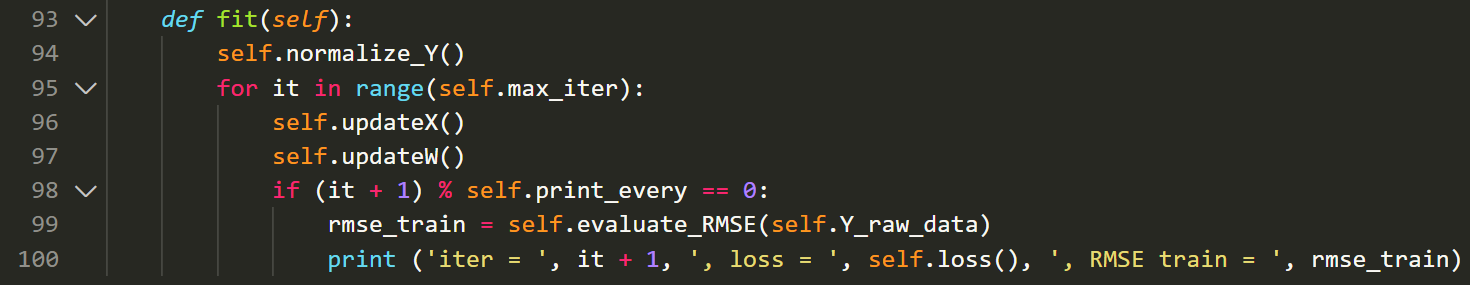
* **Xác định các items được rated bởi 1 user, và các users đã rated 1 item và các ratings tương ứng**

****

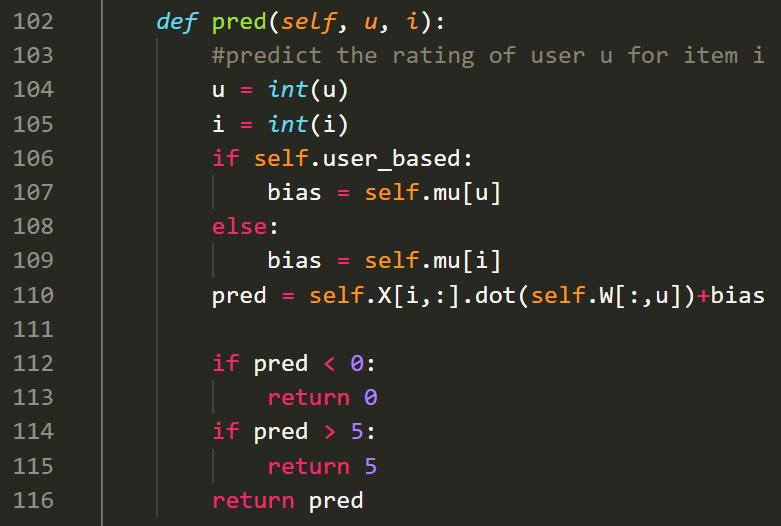
* **Cập nhật X, W: Gradient Descent**

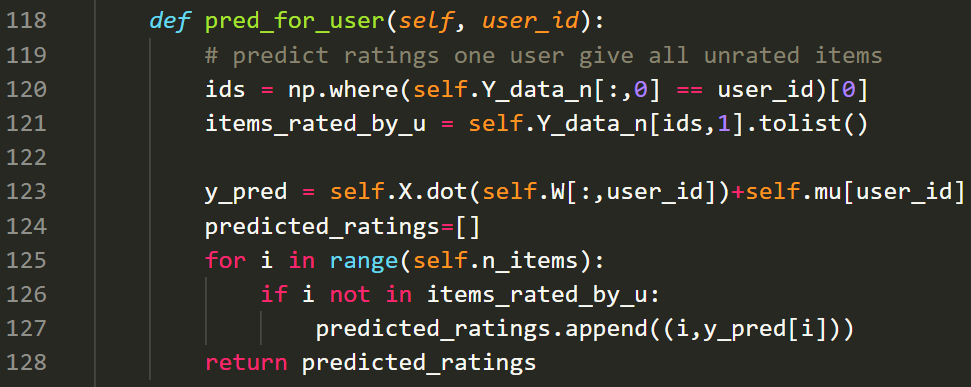
****

* **Thuật toán chính**

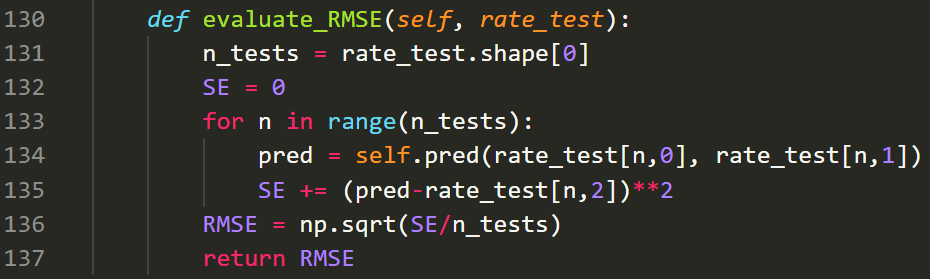
****

* **Dự đoán**

****

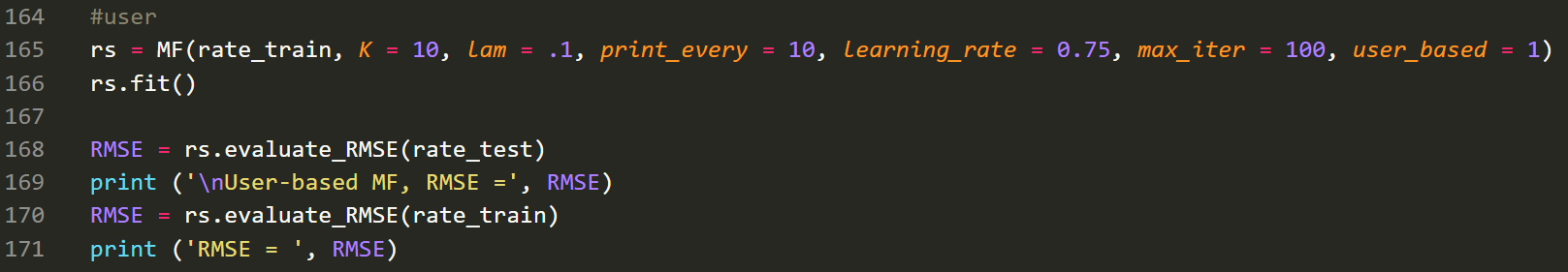
****

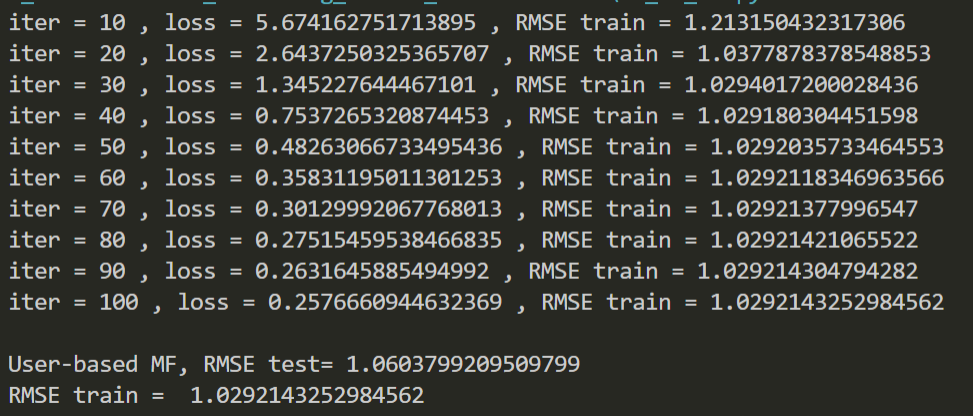
* **Đánh giá kết quả bằng RMSE**

****

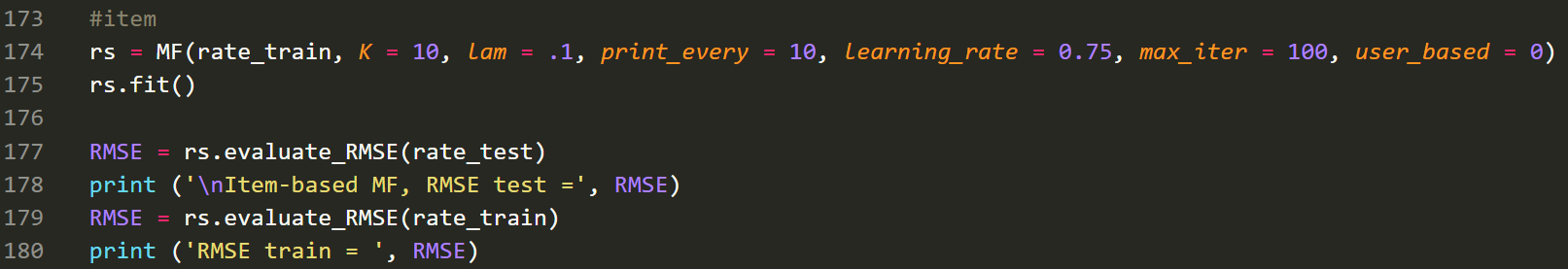
1. **Áp dụng trên MovieLens 100k**

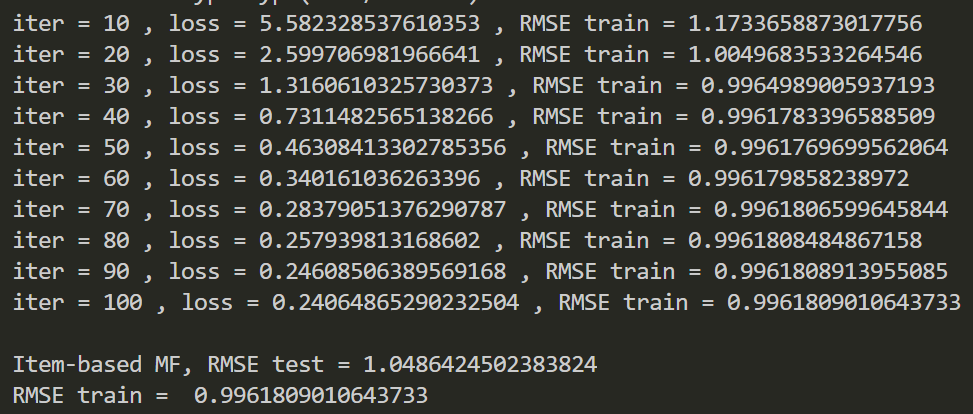
* **Chuẩn hóa dựa trên user**

****

****

* **Chuẩn hóa dựa trên item**

****

****

1. **Thảo luận**

* Khả năng linh hoạt khi có thêm các điều kiện ràng buộc khác
* Nonnegative Matrix Factorization – phân tích ma trận thành c=tích các ma trận có các phần tử không âm
* Incremental Matrix Factorization