Contents

[**A.** **COLLABORATIVE FILTERING:** 1](#_Toc77618710)

[**I.** **COLLABORATIVE FILTERING:** 1](#_Toc77618711)

[**II.** **COLLABORATIVE FILTERING ALGORITHM:** 2](#_Toc77618712)

[**III.** **KẾT LUẬN:** 2](#_Toc77618713)

[**B.** **LOW RANK MATRIX FACROTIZATION:** 3](#_Toc77618714)

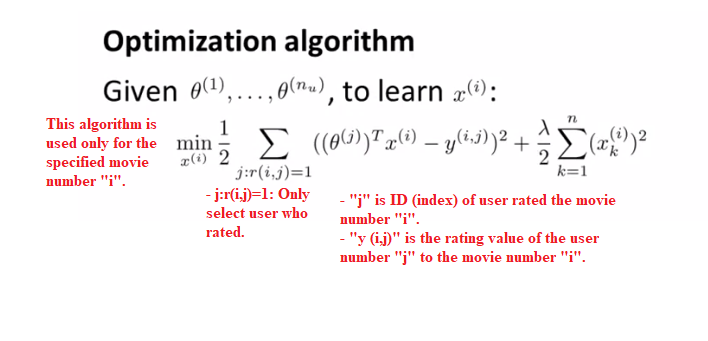
[**I.** **VECTORIZATION: LOW RANK MATRIX FACTORIZATION:** 3](#_Toc77618715)

[**II.** **IMPLEMENTATION DETAIL: MEAN NORMALIZATION:** 3](#_Toc77618716)

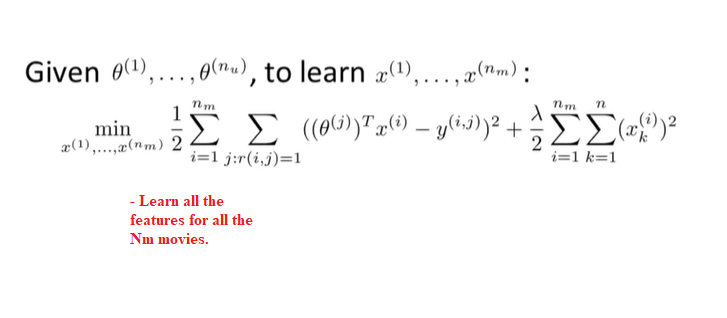
[**C.** **KẾT LUẬN:** 4](#_Toc77618717)

1. **COLLABORATIVE FILTERING:**
2. **COLLABORATIVE FILTERING:**

* Giải thuật tối ưu cho từng phim:



* Giải thuật tối ưu cho toàn bộ các phim:



* Gradient Descent for movie number i:

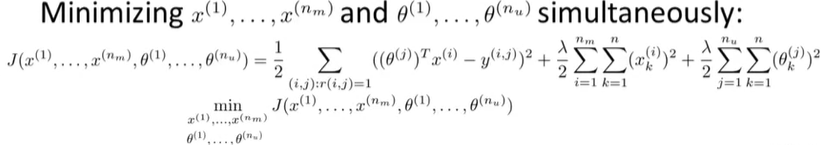


* Giải thuật:

1. **Đoán (Khởi tạo)** các bộ trọng số của từng người dùng. Mỗi bộ trọng số thể hiện **mức độ ưa thích** của người dùng đối với từng **thể loại phim**.
2. Sử dụng **(1)** để tối ưu các bộ **features , mức độ thuộc mỗi thể loại** của từng bộ phim.
3. Sử dụng **(2)** để tối ưu **(1)**.

* Chú ý: Dữ liệu liên tục được nạp vào mô hình.
* Chú ý:
  + Việc dự đoán **features** (Thể loại phim) sẽ dễ hơn là dự đoán **sở thích** của người dùng.
  + Nếu **features** (Thể loại phim) không có ý nghĩa, thì ngẫu nhiên **x &**  trước đều như nhau.

1. **COLLABORATIVE FILTERING ALGORITHM:**

****

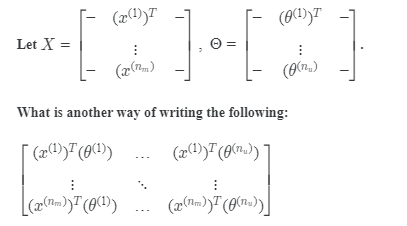
* Step sau tối ưu **x** & để thu được **cost** nhỏ hơn bước trước.
* Có thể hình dung **x** &  đều là **trọng số** của **J**. Như vậy, sử dụng **Gradient Descent** sẽ tối ưu các **trọng số** để giảm **cost** tối đa.
* Chú ý: Không cần đưa thành phần **Bias** & vào mô hình. Trong quá trình học, nếu cần thiết, mô hình sẽ **tự “học”** và tạo ra .

1. **KẾT LUẬN:**

* **x** &  đều là các **trọng số (biến số)** của mô hình.
* Khác với các mô hình **NNs** khác, **dữ liệu** mà mô hình có chỉ là phần **labeling** (Người dùng đã **rating** như thế nào cho từng phim). Mô hình sẽ phải **học** cả **trọng số & features**.
* Đối với các mô hình **NNs** thông thường, việc khởi tạo **trọng số**  **giống nhau KHÔNG** ảnh hưởng đến kết quả huấn luyện như nhiều người vẫn công nhận. Do **features** trong **training set** là khác nhau. Tuy nhiên, đối với mô hình sử dụng cho bài toán **Recommender System**, do mô hình cần học cả **x** &  nên việc khởi tạo ngẫu nhiên là quan trọng, vì nếu không **Gradient Descent** áp dụng cho từng **x** &  sẽ giống hệt nhau.
* Các **x** &  cần đủ nhỏ nên trong khoảng (0;1), xem như một hình thức chuẩn hóa **(VD: Mean normalization)**
* **Chuẩn hóa (Normalization, Standardization)** là đưa dữ liệu từ nhiều khoảng khác nhau về một khoảng tiêu chuẩn.
* Chú ý: Việc chuẩn hóa giúp **Training** nhanh hơn & các **features** có vai trò ngang nhau trong mô hình**.**

1. **LOW RANK MATRIX FACROTIZATION:**
2. **VECTORIZATION: LOW RANK MATRIX FACTORIZATION:**

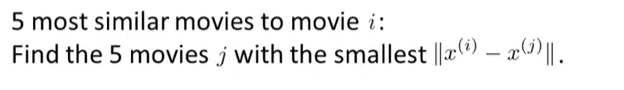
* Predicted ratings:
  + Details:



* + Shorthand:



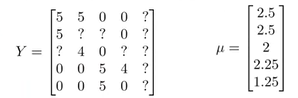
* Find movies j related to movie i:

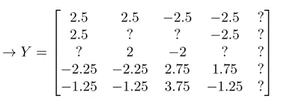


* Note: Essentially, the formula calculates **error** between 2 movies’s features.

1. **IMPLEMENTATION DETAIL: MEAN NORMALIZATION:**

* Mean Normalization:
  + Không cần sử dụng **feature scaling** vì khoảng giá trị của các **features** (Trong trường hợp này là các **labels, ratings**) đều thuộc khoảng [0;5].

****

****

* Chú ý (1): Bản chất, sử dụng **Mean Normalization** để khi dự đoán phim cho người dùng A **chưa từng rating** thì giá trị của các phim sẽ bằng **trung bình cộng** **ratings** của **toàn bộ người dùng còn lại.**
* Chú ý: Có thể hiểu **chú ý (1)** như sau, do không có cơ sở nào để dự đoán cho A, nên lấy giá trị bằng **trung bình cộng của “thị trường”** là tốt nhất.

1. **KẾT LUẬN:**

* Kết thúc **training**, thu được **features** của mỗi **movies** (Bản chất các **featrues** không rõ ràng nhưng có thể hiểu chúng tương đương: **romance, action,...**) & **trọng số (weights)**. Từ đó, **DỰ ĐOÁN RATING CỦA CÁC PHIM CHO CÁC NGƯỜI DÙNG CỤ THỂ**.
* Những người dùng chưa từng **ratings** phim **KHÔNG** làm ảnh hưởng đến **thông tin, dữ liệu** của hệ thống. Như vậy, **không** cần phải đưa vào quá trình **training**, mặc dù trong lý thuyết là có.