MATRIX FACTORIZATION

INTERPRETATIBILITY OF LATENT FEATURES

Julien – Matthieu – Mathurin - Laurene

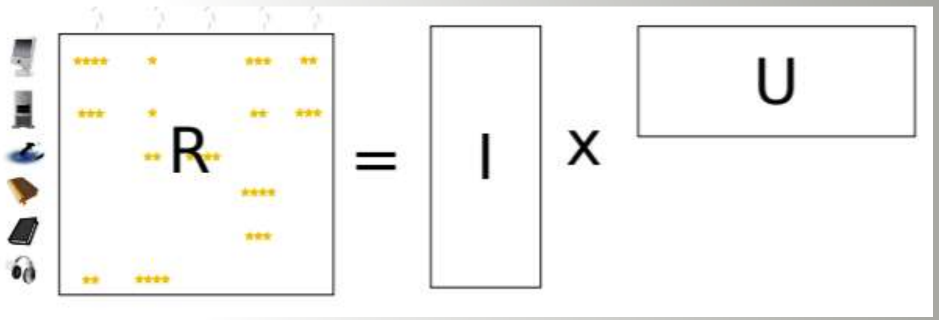
I] Introduction: Matrix Factorization

We chose to work on the **Matrix Factorization** solution for collaborative filtering.

Goal: Find matrices U(mxk) and I(kxn) such that we get the mathematical expression for Low rank matrix factorization:

R = I x UT , where:

-R(mxn) represents the rating matrix / user-item matrix
-U(mxk) is the user-feature matrix
-I(kxn) is the feature-item matrix



Remind the Cost function formula:

$$C(I, U) = \|R - IU^\top\|_{\mathcal{F}}^2 + \lambda\|I\|_{\mathcal{F}}^2 + \mu\|U\|_{\mathcal{F}}^2$$

- **Gradient Descent Method**
  Take small steps in the cost function analysis and progress in the gradient direction until we get to the minimum.

$$I_{t+1} = I_t - \eta_t \frac{\partial C}{\partial I}(I_t, U_t)$$
$$U_{t+1} = U_t - \xi_t \frac{\partial C}{\partial U}(I_t, U_t)$$

- The matrices U and I are initialized with random values and updated with gradient descent method.

II] Investigation – Interpretability of Latent features

Investigation: What can we learn from interpretability of latent features ?

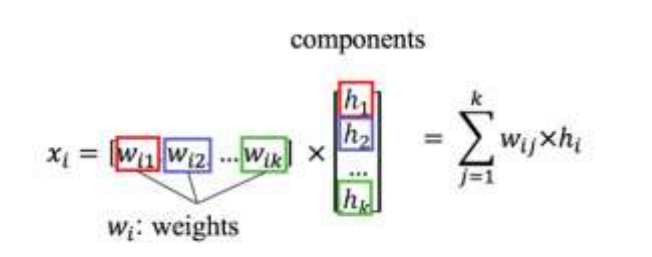
This section concern is to give meaning to the coefficients of lower dimensional matrices: U (user-feature) and I(item-feature). There exist different techniques, we'll compare some of them:

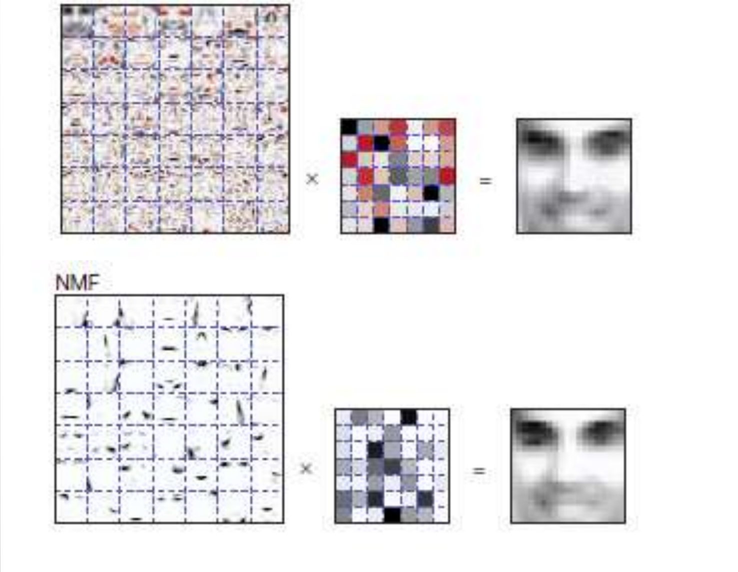
A) Interpretation based on definition – Baseline method

- general definition of MF model to provide coefficients interpretation
- small number of latent features explain the user-item interactions – incomplete picture
- not always possible to derive meaning

B) Interpretation via non-negative MF - NMF

- MF with constraint (all elements must be positive)
- represent original data as a weighted addition of vectors
- associate features with specific preference pattern
- visualization as community of users or imaginary user





C) Interpretation via Probabilistic model

- Latent features represent different groups of users.
- Find relation between users and feature from the rating matrix.
- Rows of I reveal the probability of the fact that user from each group like different items.

Rating matrix → each user provide information about how much he likes an item.

Entry

N : number of user is the rating matrix.
M : number of items in the rating matrix.
$R_{u,k}$: rating of a user u for an item i.

Parameters

K : number of groups user we want to find.
α : probability that a user belongs to one or more groups.
β : amount of evidence needed to deduce that a group likes an item.

Out

Matrix N*K : $a_{u,k}$ probability that a user belongs to a group k ➔ $\sum_{k=1}^{K} a_{u,k} = 1$

Matrix K*M : $b_{k,i}$ probability that a user from u group k like an item i.

Experiments and evaluation metrics

1. Analysis of the tradeoff of interprobability over performance

- Comparaison between Baseline and probabilistic methods
  - The constraints create a difference in the RMSE loss
- As loss increases, interprobability increases (probabilistic method)

2. Creation of a dataset

- characteristic users and items groups creation
- check if we get coherent results

3. NMF and Probabilistic approach application on MovieLens dataset

- What do we learn from the dataset
- Comparaison between the two methods