Matrix Completion $\longleftrightarrow$ Recommendation System

① Motivation

↓

Movielens Dataset → predict the rating of films

② Definition

1. Utility Matrix $X \in \mathbb{R}^{m \times n}$ $\begin{cases} m: \text{user} \\ n: \text{item} \end{cases}$

↓ (sparse matrix)

$$X_{ij} = \begin{cases} (\text{rating})_{ij}, & \text{user } i \text{ rates movie } j \\ 0, & \text{o/w} \end{cases}$$

(Drawback)

Rmk: The Rating can be biased since it comes from the people who are willing to give response.

2. Recommendation System

↓

to find those high rating entries

Approach:

$\begin{cases} \text{a) Collaborative Filtering} \\ \text{b) Latent Factor Model} \end{cases}$

③ Collaborative Filtering (User- User)

- step 1 : Grouping based on Similarity Measure
- step 2 : Aggregating to achieve prediction

1. Similarity Score

more appropriate for Binary Rating

a) Jaccard Similarity (define on set A, B)

$$Sim(x, y) := J(S_x, S_y) := \frac{|S_x \cap S_y|}{|S_x \cup S_y|}$$

Notation

- x, y : user
- S_x S_y: items rated by user x, user y (set)

more appropriate for Multi-valued Rating

b) Cosine Similarity (vector)

$$sim(x, y) := \cos(r_x, r_y) := \frac{\langle r_x, r_y \rangle}{\|r_x\| \, \|r_y\|} \in [0, 1]$$

Notation :

r_x, r_y : rating vector for user x, user y

c) Normalized Cosine Similarity

$$sim(x, y) := \cos(r_x - \overline{r_x}, r_y - \overline{r_y})$$

2. Collaborating Filter (User - User) Algorithm

→ construct Neighborhood N based on Similarity Score

↳ { threshold
  top-K

→ Aggregating { naive average $\quad r_{xi} := \dfrac{\sum_{y \in N} r_{yi}}{|N|}$

weighted average $\quad r_{xi} := \dfrac{\sum_{y \in N} sim(x,y)\, r_{yi}}{\sum_{y \in N} sim(x,y)}$

④ Item-Item Collaborative Filtering

↓

Similar to User-user CF

⑤ Latent Factor Model (LFM)

Utility Matrix M

↓

define index set $\Omega = \{(i,j) : R_{ij} \text{ is known}\}$

Ground-Truth : $R_{ij} = M_{ij}$

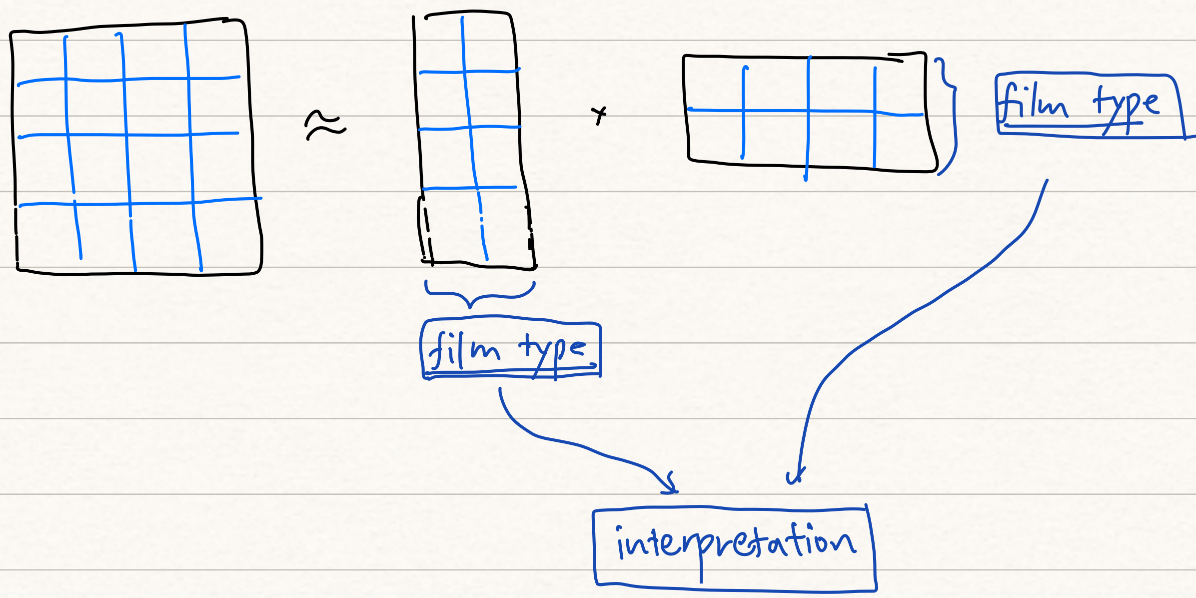
↓

Utility Matrix : M is part of the GT observation

Assumption . R is a low-rank matrix

|

$$R \approx W \cdot H \qquad \begin{cases} W \in \mathbb{R}^{m \times k} \\ H \in \mathbb{R}^{k \times n} \end{cases}$$



Different from here

in HALS of NMF, we focus on

$$\Rightarrow \begin{cases} W_{\cdot i} \\ H_{i \cdot} \end{cases}$$

→ $\boxed{\text{Formulation}}$

1. small reconstruction error

$$\boxed{\sum_{(i,j) \in \Omega} (R - WH)_{ij}^2}$$

ALS

$$= \boxed{\sum_{(i,j) \in \Omega} \left(R_{ij} - W_{i\cdot} H_{\cdot j} \right)^2}$$

→ BCD → Linear Regression Problem

$(W_{i\cdot}, H_{\cdot j})$ → (W, H)

2. $W \geq 0 \quad H \geq 0$

$$\left(\underline{1.} + \underline{2.} \right) \Rightarrow \min_{W, H} \frac{1}{2} \sum_{(i,j) \in \Omega} \left(R_{ij} - W_{i\cdot} H_{\cdot j} \right)^2 + \frac{\lambda}{2} \left(\|W\|_F^2 + \|H\|_F^2 \right)$$

↳ BCD (W, H) + $\boxed{\text{Ridge Regression}}$

ALS

⑥ Rank Perspective of Latent Factor Model

→ Nuclear Norm $\quad \|X\|_* := \sum_{i=1}^{k} \sigma_i(X) \quad$ $k = \text{rank}(X)$

↳ it can be proved (difficult) that it is norm !

↳ convex (triangle - inequality)

→ Rank : $\mathbb{R}^{m \times n} \longrightarrow \mathbb{R}$ $\quad$ (non- convex function)

↳ Count non-zero σ_i number

$\boxed{\text{rank} \longleftrightarrow \text{nuclear norm}}$

connection

$\boxed{\|\cdot\|_0 \longleftrightarrow \|\cdot\|_1}$

→ convex envelope of $f(\cdot)$ (probably non-convex)

↓

the largest convex function $g(\cdot)$ s.t

$$g(x) \leq f(x) \quad \forall x$$

"Ball"

Result : Rank function defined on $\boxed{\{X : \sigma_1(X) \leq 1\}}$

↓ convex envelope

$\boxed{\|\cdot\|_*}$ (nuclear norm)

② Rank function defined on $\boxed{\{X : \sigma_1(X) \le k\}}$

$\downarrow$ convex envelope

$\frac{1}{k}\|\cdot\|_*$ (nuclear norm)

→ $\boxed{\text{Re-formulation}}$ → $\begin{cases} \min\limits_{R} & \text{rank}(R) \\ \text{s.t} & R_{ij} = M_{ij} \end{cases}$ → NP-HARD

$\forall (i,j) \in \Omega$

M_{ij} → observation data (recorded rating)

Surrogation (convex relaxation)

$$\begin{cases} \min\limits_{R} & \|R\|_* \\ \text{s.t} & R_{ij} = M_{ij} \quad \forall (i,j) \in \Omega \end{cases}$$

$\begin{cases} \text{Approach 1 : SDP (Semi-Definite Program)} \\ \text{Approach 2 : PG} \end{cases}$