Report 2 : Partial SVD 1

PSVD-1. (compulsory)
From given partial rating matrices A and B
(given in SVD_interpolation_2021.ipynb),
infer the complete rating matrices by using
SVD_interpolation_2021.ipynb with fixed f.

PSVD-2. (compulsory): leave-one-out validation
In the given partial rating matrices A and B,
there are N non-zero matrix elements.
You may optimize U and M by using N-1 elements,
and evaluate a test error with 1 element.
If you choose N different sets of the N-1 elements
and take a root square mean error of the test errors,
you will obtain an averaged test error.
Obtain the test errors for the given A and B.
You may fix f, k_u, k_m, and μ (given the ipynb file).

Report 2: Partial SVD 2

PSVD-3. (compulsory)
Infer which rating matrix is sampled from a random rating matrix or which rating matrix Is sampled from a low-rank rating matrix.
Show the basis (根拠) for your inference (推定).

PSVD-4. (optional):
Obtain f dependence of the averaged test error for the given A and B.
You may fix k_u, k_m, and μ as in PSVD-2,
or you may optimize k_u, k_m, and μ.

Report Problem 2: Deadline

Deadline: Jaunuary 27, 2022

・A jupyter notebook including matrices *A* and *B* for the Report Problem is uploaded on ITC-LMS and github.

・Optional problems [problems with (optional)] are not mandatory. Even if you will not solve them, you can get the credit.

・Please include your name and student id.

・Please submit your report through ITC-LMS, and
please confirm your submission carefully!
Every single year, there are several students who failed to submit while they did not recognize those failures.
If and only if you have any trouble with ITC-LMS,
please send it to us via email:
YAMAJI.Youhei@nims.go.jp
t-okubo@phys.s.u-tokyo.ac.jp

PSVD: Singular Value Decomposition of Partially Unknown Matrix

As a review,
Chih-Chao Ma, *A Guide to Singular Value Decomposition for Collaborative Filtering*.

Mathematical formulation

Minimize the cost function

$$E = \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{m} I_{ij}[V_{ij} - p(U_i, M_j)]^2$$

$$+ \frac{k_u}{2}\sum_{i=1}^{n}\|U_i\|_2^2 + \frac{k_m}{2}\sum_{j=1}^{m}\|M_j\|_2^2$$

-L_2 regularization

Feature vectors

$$U_i = (U_{1i}, U_{2i}, \ldots, U_{fi})^T$$

$$M_j = (M_{1j}, M_{2j}, \ldots, M_{fj})^T$$

A prediction function for the rating (You need to choose)

$$p(U_i, M_j) = \begin{cases} 1 & \text{if } U_i^T M_j < 0 \\ 1 + U_i^T M_j & \text{if } 0 < U_i^T M_j < 4 \\ 5 & \text{if } 4 < U_i^T M_j \end{cases}$$