Report 2 : Partial SVD 1

PSVD-1. (compulsory)
From given sets of partial rating matrices,
infer the complete rating matrices by using PSVD.py.
At least, inset01.in and inset11.in should be used.
In your report, k_u and k_o you used and the inferred matrices should be included.

-We prepared 2 complete rating matrices: A correlated matrix and a random matrix. From these matrices, we randomly sample and make two sets of rating matrices.

> inset01.in, inset02.in, …, inset08.in are sampled from a matrix, and
> inset11.in, inset12.in, …, inset18.in from the other matrix.

-PSVD.py receives an input file, a rank for partial SVD *f*, and maximum iteration *imax* for the steepest gradient.

> python ./PSVD.py *inputfile* *f* *imax*

Report 2: Partial SVD 2

PSVD-2. (optional)
Infer which set of rating matrix is sampled from a random rating matrix.
Show the basis (根拠) for your inference (推定).

-For example, you may compare the inferred rating matrices from (inset01.in, inset02.in, …, inset08.in) or (inset11.in, inset12.in, …, inset18.in) and measure variance of the inferred matrix elements.
-For example, you may infer a rating matrix from inset01.in or inset11.in (training data).
Then, you may compare the inferred rating matrix and inset02.in, …., inset08.in or inset12.in, …., inset18.in (test data).
(Cross validation)
-Variance may depends on the hyperparameters k_u and k_o

Revisit: 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

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

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

-L_2 regularization

Feature vectors

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

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

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

Revisit: Singular Value Decomposition of Partially Unknown Matrix

Original rating matrix

$$V = \begin{bmatrix} 5 & 0 & 0 & 3 & 0 & 0 \\ 0 & 2 & 0 & 0 & 3 & 0 \\ 3 & 0 & 4 & 0 & 0 & 4 \\ 5 & 5 & 0 & 3 & 0 & 0 \\ 0 & 0 & 2 & 0 & 5 & 3 \end{bmatrix}$$

A predicted rating matrix

$$p(U_i, O_j) = \begin{bmatrix} \color{blue}{4.92} & 4.39 & 2.29 & \color{blue}{2.96} & 4.93 & 3.23 \\ 2.57 & \color{blue}{2.02} & 1.07 & 1.78 & \color{blue}{2.94} & 1.60 \\ \color{blue}{3.01} & 3.82 & \color{blue}{3.92} & 2.05 & 2.55 & \color{blue}{3.94} \\ \color{blue}{4.98} & \color{blue}{4.91} & 3.11 & \color{blue}{3.00} & 4.64 & 3.85 \\ 4.55 & 3.85 & \color{blue}{2.01} & 2.78 & \color{blue}{4.92} & \color{blue}{2.98} \end{bmatrix}$$