1 Co-clustering by Block Value Decomposition

This is a co-clustering algorithm called Block Value Decomposition (BVD) based on Nonnegative Matrix Factorization (NMF) technique. The goal is to find a factorization for the data matrix $X \in \mathbb{R}_{+}^{N \times M}$, where N is the number of objects, M is the number of features of these objects and the factorization takes the form

$$X \approx USV^T$$

, where $U \in \mathbb{R}_{+}^{N \times L}$ is a matrix of rows factors representing features clusters, $S \in \mathbb{R}_{+}^{L \times K}$ is a block matrix representing how blocks are related, and $V \in \mathbb{R}_{+}^{M \times K}$ is a matrix of columns factors representing rows clusters.
This algorithm solves the following optimization problem:

$$min\,||X - USV^T||^2 \; s.t. \; U \geq 0, S \geq 0, V \geq 0$$

The optimization problem can be solved using Lagrange multipliers (λ), optimizing the following Lagrange function:

$$\mathscr{L} = |X - USV^T||^2 - tr(\lambda_1 U^T) - tr(\lambda_2 S^T) - tr(\lambda_3 V^T)$$

Then $\mathscr{L}$ must satisfy the K.K.T. conditions:

$$\frac{\partial \mathscr{L}}{\partial U} = 0$$

$$\frac{\partial \mathscr{L}}{\partial S} = 0$$

$$\frac{\partial \mathscr{L}}{\partial V} = 0$$

$$\lambda_1 \odot U = 0$$

$$\lambda_2 \odot S = 0$$

$$\lambda_3 \odot V = 0$$

Solving the derivatives and equal them to 0, is possible to solve the optimization problem by applying gradient ascending on $\mathscr{L}$ with the following update rules:

$$U \leftarrow U \odot \frac{XVS^T}{USV^TVS^T}$$

$$V \leftarrow V \odot \frac{U^TXV}{U^TUSV^TV}$$

$$S \leftarrow S \odot \frac{S^TU^TX}{S^TU^TUSV^T}$$

2 Fast Nonnegative Matrix Tri factorization

In this case, the goal is to optimize the following problem:

$$min\,||X - USV^T||^2 \; s.t. \; U \in \Psi^{n\times k}, S \in \mathbb{R}_+^{l\times k}, V \in \Psi^{m\times l}$$

where U and V turns into cluster indicator matrices, with vectors $\vec{u}_i$ and $\vec{v}_j$ that contains $1s$ in only one position, indicating the cluster that that this vector belongs, and $0s$ in the rest. Similar to the other algorithm, it optimizes S with a multiplicative update rule and the following subproblems:

$$S \leftarrow (U^T U)^{-1} U^T X V (V^T V)^{-1}$$

$$v_{ij} \begin{cases} 1 & j = argmin_l ||\vec{x}_i - \vec{\tilde{u}}_l||^2 \\ 0 & otherwise \end{cases}$$

$$u_{ij} \begin{cases} 1 & i = argmin_k ||\vec{x}_j - \vec{\tilde{v}}_k||^2 \\ 0 & otherwise \end{cases}$$

where $\tilde{U} = US$ and $\tilde{V} = SV^T$

3 Overlapping Orthogonal Nonnegative Matrix Tri Factorization

This is a proposal algorithm that aims to solve the following problem:

$$min\,||X - UV'||_F^2$$

$$s.t.\; U^T U = I$$

$$\begin{bmatrix} V^{(1)} \\ V^{(2)} \\ \vdots \\ V^{(k)} \end{bmatrix}^T \begin{bmatrix} V^{(1)} \\ V^{(2)} \\ \vdots \\ V^{(k)} \end{bmatrix} = I$$

$$\begin{bmatrix} V^{(1)} \\ V^{(2)} \\ \vdots \\ V^{(k)} \end{bmatrix} \geq 0$$

$$U, S \geq 0$$

where $V^{(c)} \in \mathbb{R}^{M\times L}$ and $V'_{c\cdot} = S_{c\cdot} V^{(c)^T}, \forall c \in \{1, \ldots, k\}$.
This way the objects (lines) in X can belong to multiple clusters.