Notes LLMKL

Kylliann De Santiago

Idées de la méthode :

- L'idée est de combiner un maximum de méthodes afin d'extraire le plus d'information possible.
- La méthode `LLMKL` se base sur : "self-expressivenes graph", Structure globale des données, structure de graphe locale, Noyaux Consensus.
- Objectif : Construire une matrice d'affinité `Z` qui permettra de faire un clustering efficace à partir de ces différents axes.

self-expressivenes graph

- L'idée : Construire une matrice d'affinité Z qui associera de grandes valeurs aux observations du même sous-espace.
- On cherche Z tel que :

$$\min_Z \frac{1}{2}||X - XZ||_F^2 + \alpha R(Z), \text{ où } Z \geq 0, \text{ diag}(Z) = 0$$

On rappelle que :

$$< A, B >_F = \text{Tr}(A^* B) = \text{Tr}(BA^*) \text{ et} ||A||_F^2 = \text{Tr}(A^* A)$$

Avec cette formulation, nous saisissons mal les données non linéaire, on peut reformuler le problème ainsi :

$$\min_Z \frac{1}{2}||\phi(X) - \phi(X)Z||_F^2 + \alpha R(Z)$$

où $Z \geq 0, \text{ diag}(Z) = 0$ et ϕ est une fonction noyau.

$$\min_Z \frac{1}{2}||\phi(X) - \phi(X)Z||_F^2 \ + \alpha R(Z)$$

En développant la norme :

$$\min_Z ||\phi(X)||_F^2 + < \phi(X), \phi(X)Z >_F + ||\phi(X)Z||_F^2 \ + \alpha R(Z)$$

D'où :

$$\min_Z \mathrm{Tr}[\phi(X)^T\phi(X)] + 2\ \mathrm{Tr}(\phi(X)^T\phi(X)Z) + \mathrm{Tr}[(\phi(X)Z)^T\phi(X)Z] + \alpha R(Z$$

Finalement :

$$\min_Z \mathrm{Tr}(H \ + 2\ HZ + \ Z^T HZ) \ + \alpha R(Z)$$

s.c. $\;Z \geq 0,\ \mathrm{diag}(Z) = 0.$

On notera que $\phi(X)^T\phi(X) = H$, $X \in \mathbb{R}^{p\times n}$ et $H \in \mathbb{R}^{n\times n}$

Global structure

- On souhaite que H soit de faible rang. On impose donc en plus un coefficient :
$$\beta||\phi(X)||_*$$
- Puisque $H = \phi(X)^T\phi(X)$ alors rang(H)=rang($\phi(X)$). Donc, minimiser $||\phi(X)||_*$ équivaut à minimiser $||H||_*$
- De plus, H est une matrice symmétrique et semi-définie positive alors $\exists B, H = B^T B$
- Donc, minimiser $||H||_*$ revient à minimiser $||B||_*$.

La matrice B peut capturer la structure globale des données dans l'espace engendré par H.

Notre matrice d'affinité devient donc :

$$\min_{Z,B} \text{Tr}(H \ + 2 \ HZ + \ Z^T HZ) \ + \alpha R(Z) + \beta||B||_*$$

s.c. $\;Z \geq 0, \ \text{diag}(Z) = 0, B^T B = H.$

Local structure

L'idée de cette partie est derajouter une nouvelle pénalisation à partir d'un coefficient directement relié à la structure locale des données.

On commence par définir le graphe complet D tel que $D_{ij} = ||X_i - X_j||_2^2$ où chaque individu est considéré comme un noeud, et chaque lien entre deux noeud correspond à l'affinité entre les deux individus.

On définit :

$$R(Z) = \min_Z \sum_{i,j=1}^{n} ||X_i - X_j||_2^2 Z_{ij}$$

$$R(Z) = \min_Z \sum_{i=1}^{n} \sum_{j=1}^{n} D_{ij} Z_{ij}$$

Donc :

$$R(Z) = \min_Z \operatorname{Tr}(D^T Z)$$

Consensus Kernel

L'idée de cette partie est de terminer un noyaux consensus H proche du mélange de `r`noyaux : $\{H_i\}_1^r$.

Autrement dit, on souhaite trouver H tel que (MKL weighting strategy) :

$$\min_{H,g} ||H - \sum_{i=1}^{r} g_i H_i||_F^2$$

s.c. $g_i \geq 0$ et $\sum_{i=1}^{r} g_i = 1$

Fonction d'ojectif :

Une autre idée a été de décomposer H tel que $H = B^T B + E$ où E est une matrice de bruit.

En aggrégeant les idées, on aboutit à la fonction d'objectif suivante :

$$\min_{Z,B,H,E,g} \ \mathrm{Tr}[(I \ + 2Z + \ Z^T Z)B^T B] \ + \lambda_1 ||B||_*$$
$$+\lambda_3 \mathrm{Tr}(D^T Z) + \frac{\lambda_2}{2}||H - \sum_{i=1}^{r} g_i H_i||_F^2 + \lambda_4 ||E||_1$$

s.c. $Z > 0$, $gi \geq 0$, $\sum_{i=1}^{r} g_i = 1$, $H = B^T B + E$, $\mathrm{diag}(Z) = 0$, $Z^T 1 = 1$ (1 vectoriel).

Résumé :

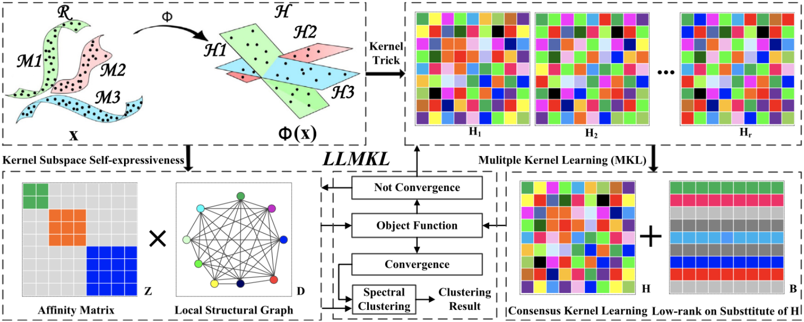

Fig. 1. The block diagram of the proposed LLMKL method.

Algorithme :

Algorithm 1 Solving LLMKL for clustering via ADMM.

Input: Data matrix $\boldsymbol{X}$, r base kernel matrices $\{\boldsymbol{H}_i\}_{i=1}^r$, and tradeoff parameters $\lambda_1, \lambda_2, \lambda_3, \lambda_4$.
Initialize: $\boldsymbol{Z}^1 = \boldsymbol{0}$, $\boldsymbol{E}^1 = \boldsymbol{0}$, $\boldsymbol{H}^1 = \frac{1}{r}\sum_{i=1}^r \boldsymbol{H}_i$, $\{g_i^1\}_{i=1}^r = \frac{1}{r}$, $\varepsilon = 10^{-6}$, $t = 1$, $\mu_{max} = 1e10$, $\mu = 0.1$, $\eta = 20$, and $maxIter = 1e3$.

while convergence criterion is not satisfied, and $t < maxIter$ **do**
    Update $\boldsymbol{Z}^{t+1}$ by using (14).
    Update $\boldsymbol{B}^{t+1}$ by using (20).
    Update $\boldsymbol{H}^{t+1}$ by using (16).
    Update $\boldsymbol{E}^{t+1}$ by using (23).
    Update $\boldsymbol{g}^{t+1}$ by using (25).
    Update $\boldsymbol{Y}_1^{t+1}$ and μ^{t+1} by using (26).
    t = t+1;
end while
Define balanced affinity matrix $\boldsymbol{Z} = \frac{1}{2}(\boldsymbol{Z} + \boldsymbol{Z}^T)$.
Perform spectral clustering by using (27).

Output: The clustering results: ACC, NMI, Purity.