Sparse Subspace K-means

Abdoul Wahab Diallo & Ndèye Niang & Mory Ouattara

AIMS, Mbour, Sénégal
abdoul.w.m.diallo@aims-senegal.org

CEDRIC CNAM Paris, France
ndeye.niang_keita@cnam.fr

Laboratoire de Mathématique Informatique
UFR SFA, Université Nangui Abrogoua,
Abidjan, Côte d'ivoire
ouattaramory.sfa@univ-na.ci

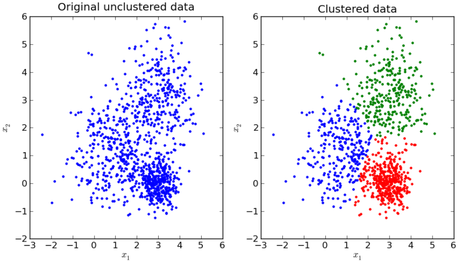
MSDMA December 17, 2021

Overview

Table of contents

Introduction

- Clustering is an unsupervised learning task that seeks to group objects in a dataset into groups (clusters).



Source: https://i.stack.imgur.com/cIDB3.png

- Several clustering methods provide clusters (K-means(Forgy (1965)), or hierarchical clustering methods (CAH) (Gordon(1987)).

Introduction

The complexity of the data has led to many advances in the field of clustering.

Subspace Clustering (Hard, Soft)

Parsons and al (2004), Kriegel and al (2009)

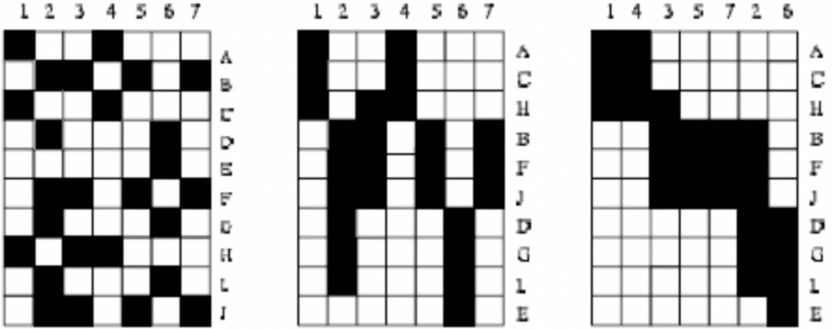
Bi-partionnement or co-clustering

Kriegel and al (2009)

Clustering Ensemble & Multi-block Clustering

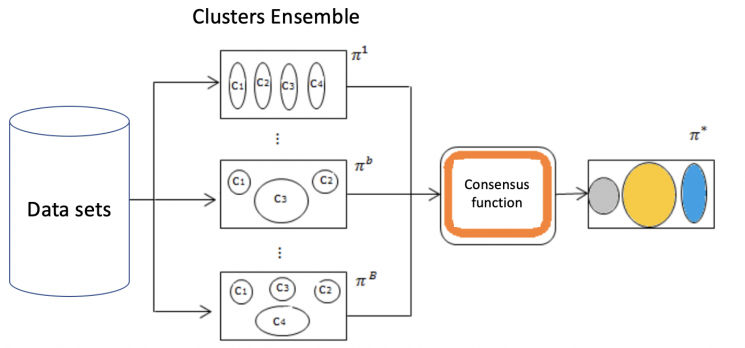
Strehl (2002), Vega-Pons (2010)

They aim to obtain the most homogeneous individual / variable or row / column blocks according to metric or probabilistic criteria



Clustering Ensemble & Multi-block Clustering

These cluster ensemble methods consist of combining multiple partitions of the same dataset to improve the performance of classifications.

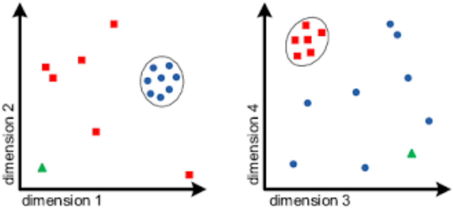


This idea takes up the older concepts of seeking a consensus of scores proposed by [Régnier 1983; Breiman 1996; Gordon et Vichi 1998; Strehl et Ghosh 2002].

Subspace Clustering

Subspace clustering aims at identifying simultaneously

- the clusters
- the specific subspaces of features describing each cluster



Source: researchgate.net/publication

- Bottom-up or Top-down approches
- EWKM (Jing et al) (2007).
- The difficulty of interpretation when the number of features is too large.

Table of contents

K-means clustering algorithm

$E = \{x_1, \ldots, x_n\}$ and $x_i \in \mathbb{R}^p$

The K-means clustering algorithm finds K clusters $C_1, \ldots, C_K$ that minimize the within clusters sum of squares

$$\min_{C_1,\ldots,C_K} \left\{ \sum_{k=1}^{K} \frac{1}{n_k} \sum_{i,i' \in C_k} d_{i,i'} \right\} \quad \text{where} \quad d_{i,i'} = \sum_{j=1}^{p} (x_i^j - x_{i'}^j)^2$$

where $C_1, \ldots, C_K$ are the disjoint cluster, n_k is the number of observations in the k-th cluster.
Minimizing the within-cluster sum of squares is also equivalent to maximizing the between-cluster sum of squares

$$\frac{1}{n} \sum_{i=1}^{n} \sum_{i'=1}^{n} d_{i,i'} - \sum_{k=1}^{K} \frac{1}{n_k} \sum_{i,i' \in C_k} d_{i,i'} = \sum_{j=1}^{p} \left\{ \frac{1}{n} \sum_{i=1}^{n} \sum_{i'=1}^{n} d_{i,i',j} - \sum_{k=1}^{K} \frac{1}{n_k} \sum_{i,i' \in C_k} d_{i,i',j} \right\}$$

Sparse K-means

The SK-means algorithm of Witten and Tibshirani (2010) introduces non-negative weights w_j , $j = 1, ..., p$ for each feature and then solves

$$\max_{C_1,\dots,C_k,w} \left\{ \sum_{j=1}^{p} w_j \Big(\frac{1}{n} \sum_{i=1}^{n} \sum_{i'=1}^{n} d_{i,i',j} \; - \sum_{k=1}^{K} \frac{1}{n_k} \sum_{i,i' \in C_k} d_{i,i',j} \Big) \right\}$$

subject to : $||w||_2 \leq 1$, $||w||_1 \leq s$, $w_j \geq 0$, $\forall j$.

[D.M. Witten and R.Tibshirani, 2010]

Weight optimization

The process of optimizing the weights of the features will be done using the following formula:

$$w = \frac{S((a)_+,\Delta)}{||S((a)_+,\Delta)||_2}$$

where

$$a_j = \frac{1}{n}\sum_{i=1}^{n}\sum_{i'=1}^{n} d_{i,i',j} - \sum_{k=1}^{K}\frac{1}{n_k}\sum_{i,i'\in c_k} d_{i,i',j} \tag{1}$$

- and $\Delta = 0$ if that results in $||w||_1 < s$;
- otherwise, $\Delta > 0$ is chosen so that $||w||_1 = s$.
- S is the soft-thresholding operator, defined as $S(x,c) = sign(x)(|x| - c)_+$.

Selection of the tuning parameter s

1. Obtain permuted data sets $X_1, ...X_B$
2. For each candidate tuning parameter value s:
   - Compute

$$O(s) = \sum_{j=1}^{p} w_j \Big(\frac{1}{n} \sum_{i=1}^{n} \sum_{i'=1}^{n} d_{i,i',j} - \sum_{k=1}^{K} \frac{1}{n_k} \sum_{i,i' \in C_k} d_{i,i',j} \Big)$$

   - For $b = 1, 2, .., B$, compute $O_b(s)$, the objective obtained by performing Sparse K-means with tuning parameter value s on the data X_b.
   - Calculate

$$Gap(s) = log(O(s)) - \frac{1}{B} \sum_{b=1}^{B} log(O_b(s))$$

3. Choose $s^\star$ corresponding to the largest value of Gap(s).

Table of contents

Sparse Subspace K-means

- SSKM is an extension of the Sparse K-means method.
- The optimization criterion associated with the algorithm is given below:

$$\max_{C_1,\ldots,C_k,w} \left\{ \sum_{j=1}^{p} w_j \Big(\frac{1}{n} \sum_{i=1}^{n} \sum_{i'=1}^{n} d_{i,i',j} - \sum_{k=1}^{K} \frac{1}{n_k} \sum_{x_i, x_i' \in C_k} d_{i,i',j} \Big) \right\}$$

$$\max_{c_1,\ldots,c_K,w^k} \left\{ \sum_{j=1}^{p} \sum_{k=1}^{K} w_j^k \left(\frac{1}{nK} \sum_{i=1}^{n} \sum_{i'=1}^{n} d_{i,i',j} - \frac{1}{n_k} \sum_{x_i, x_{i'} \in c_k} d_{i,i',j} \right) \right\}$$

subject to $\| w^k \|_2 \leq 1$, $\| w^k \|_1 \leq s$ et $w_j^k \geq 0, \forall j \; and \; k$.

- The selection of the tuning parameter s is done in the same way as for the Sparse K-means.

Weight optimization

The process of optimizing the weights of the features will be done using the following formula:

$$w^k = \frac{S((a^k)_+, \Delta)}{||S((a^k)_+, \Delta)||_2}$$

où

$$a_j^k = \frac{1}{n} \sum_{i=1}^{n} \sum_{i'=1}^{n} d_{i,i',j} - \frac{1}{n_k} \sum_{i,i' \in c_k} d_{i,i',j}$$

- and $\Delta = 0$ if that results in $||w^k||_1 < s$;
- otherwise, $\Delta > 0$ is chosen so that $||w^k||_1 = s$.
- S is the soft-thresholding operator, defined as $S(x, c) = sign(x)(|x| - c)_+$.

SSKM Algorithm

1. Initialize w^k with $w_1^k = ... = w_p^k = \frac{1}{\sqrt{p}}$, $1 \leq k \leq K$.
2. Iterate until convergence :
   - Fix w^k and optimize the objective function with respect to $C_1, ..., C_K$:

     $$\underset{C_1,\ldots,C_K}{min} \left\{ \sum_{k=1}^{K} \frac{1}{n_k} \sum_{x_i, x_{i'} \in c_k} \sum_{j=1}^{p} w_j^k d_{i,i',j} \right\}$$

     by applying the standard K-means algorithm to the weighted data.
   - Fix $C_1, ..., C_K$ and optimize the objective function with respect to $w^1, ..., w^k$ by applying the formula on weights.

- Step 2 is iterated until the condition below is satisfied.

$$\frac{\sum_{j=1}^{p} |(w_j^k)^r - (w_j^k)^{r-1}|}{\sum_{j=1}^{p} |(w_j^k)^{r-1}|} < 10^{-4}$$

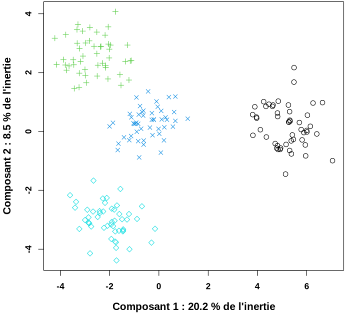
Table of contents

Data

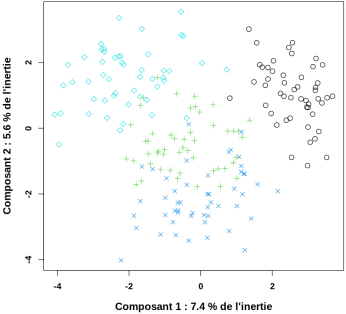
Data	n	p	k	Number of features in each bloc
X_1	200	60	4	60
X_2	200	60	4	15, 15, 15, 15
DMU	2000	649	10	76, 216, 64, 240, 47, 6
IS	2000	18	7	8, 18

Table: Table of simulated and real data

Simulated data



(a) Projection of X_1 clusters

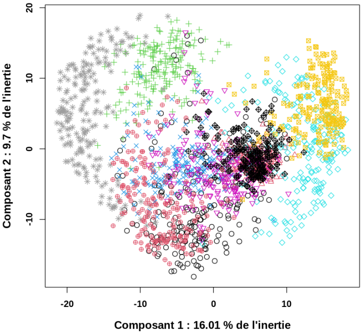


(b) Projection of X_2 clusters

Reals data : DMU

The Multiple Features data set contains 2000 handwritten digits grouped in 10 clusters $c_k k(k = 0, ..., 9)$, each with 200 objects.

Each digit is described by 649 features that are divided into six feature groups:

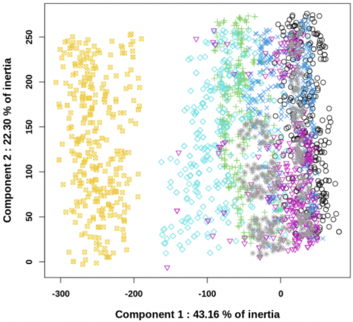


Reals data : Image Segmentation

IS data set consists of 2310 objects grouped in 7 clusters, drawn randomly from a database of out door images.

The data set contains 18 features divided into two feature groups :

$g1 = \{V_i(i = 1, \ldots, 8)\}$: Shape group,
$g2 = \{V_i i(i = 9, \ldots, 18)\}$: RGB group



(a) Projection of IS clusters

Evaluation indices

We have used 2 indices :

- Normalized Mutual Information (NMI) :

$$NMI(\mathcal{C},\mathcal{C}') = \frac{I(\mathcal{C},\mathcal{C}')}{\sqrt{H(\mathcal{C})H(\mathcal{C}')}}$$

où $I(\mathcal{C},\mathcal{C}')$: Mutual Information between two partitions $\mathcal{C}$ et $\mathcal{C}'$ $H(\mathcal{C})$: The entropy associated with the partition $\mathcal{C}$.

- Adjusted Rand Index (ARI) :

$$\mathcal{R}_a(\mathcal{C},\mathcal{C}') = \frac{n^2\sum_{i,j} n_{ij}^2 - \sum_i n_{i.}^2 \sum_j n_{.j}^2}{\frac{1}{2}n^2(\sum_i n_{i.}^2 + \sum_j n_{.j}^2) - \sum_i n_{i.}^2 \sum_j n_{.j}^2}$$

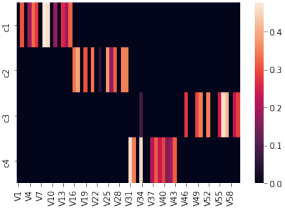
•n_{ij} : number of observations in both c_i and c'_j.
•$n_{i.}$: number of observations in both c_i and c'_k.

Comparison of performance

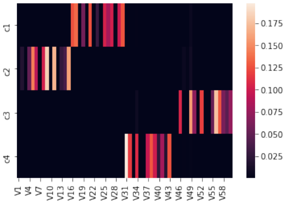
Data	Indices	K-means	EWKM	Sparse K-means	SSKM
X_1	NMI	0.95 (0.1)	0.52 (0.1)	**1** (0)	**1** (0)
	ARI	0.93 (0.13)	0.45 (0.21)	**1** (0)	**1** (0)
X_2	NMI	0.76 (0.08)	0.47 (0.17)	0.89 (0)	**0.92** (0.02)
	ARI	0.77 (0.12)	0.45 (0.17)	**0.92** (0)	**0.92** (0.03)
DMU	NMI	0.73 (0.04)	0.49 (0.05)	0.78 (0)	**0.81** (0.02)
	ARI	0.63 (0.06)	0.38 (0.07)	0.71 (0)	**0.75** (0.05)
IS	NMI	**0.57**(0.02)	0.47(0.08)	0.56(0)	**0.57** (0.01)
	ARI	**0.47**(0.03)	0.34(0.09)	0.46(0)	**0.46**(0.01)

Table: Performances of K-means, EWKM, Sparse K-means and SSKM on X_1, X_2, DMU and IS data sets

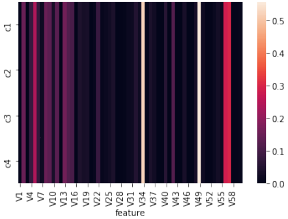
Evaluation of the relevance of the X_2 features



(a) SSKM

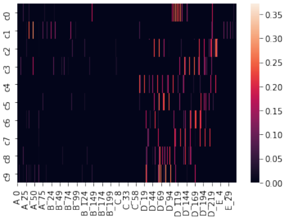


(b) EWKM

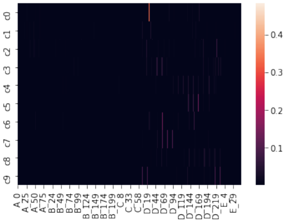


(c) SparseKM

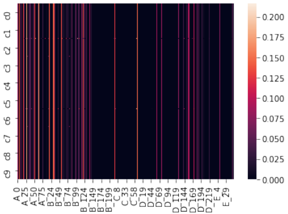
Evaluation of the relevance of features



(a) SSKM

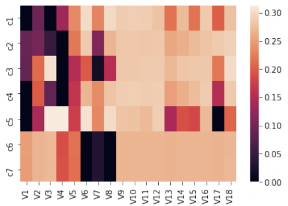


(b) EWKM

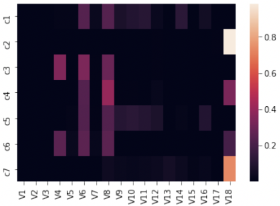


(c) SparseKM

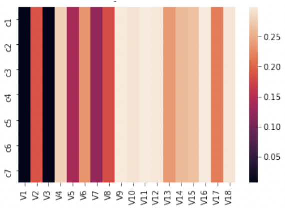
Evaluation of the relevance of features



(a) SSKM



(b) EWKM



(c) SparseKM

Table of contents

Conclusion & Perspectives

- The application of the SSKM method yields very satisfying results in terms of clustering quality and cluster subspaces identification.
- The SSKM method also facilitates the interpretation of the partition.
- Future work, will be dedicated to the extention of SSKM to methods such as FGKM (Chen and al, 2012) taking into account the block structure and getting sparsity both for features and blocks of features.

Thank you for your attention !!!

References

- *Witten, D., Tibshirani, R. (2010). A framework for feature selection in clustering. Journal of the American Statistical Association, 713–726.*
- *Chen, X., Ye, Y., Xu, .X, Huang, J. Z. (2012). A feature group weighting method for subspace clustering of high-dimensional data. Pattern Recognition, 434-446.*
- *Jing, J., Ng, M., Huang, J. (2007). An entropy weighting k-means algorithm for subspace clustering of high-dimensional sparse data. Knowledge and Data Engineering, IEEE Transactions, 1026-1041.*
- *Parsons, L., Haque, E., Lui, H. (2004). Subspace clustering for high dimensional data : a review. SIGKDD Explor, Newsl, 90-105.*

References

- *Weighted Clustering Ensemble: A Review Mimi Zhang School of Computer Science and Statistics, Trinity College Dublin, Dublin 2, Ireland*
- *Banerjee, A., Pujari, A. K., Rani Panigrahi, C., Pati, B., Chandan Nayak, S., & Weng, T. H. (2021). A new method for weighted ensemble clustering and coupled ensemble selection. Connection Science, 1-22.*
- *Zhang, X., Yang, Y., Li, T., Zhang, Y., Wang, H., & Fujita, H. (2021). CMC: A consensus multi-view clustering model for predicting Alzheimers disease progression. Computer Methods and Programs in Biomedicine, 199, 105895.*

References

- *Nazari, A., Dehghan, A., Nejatian, S., Rezaie, V., and Parvin, H. (2017). A comprehensive study of clustering ensemble weighting based on cluster quality and diversity. Pattern Analysis and Applications.*
- *Huang, D.,Wang, C., and Lai, J. (2018). Locally weighted ensemble clustering. IEEE Transactions on Cybernetics, (99):1–14.*
- *Boongoen, T. and Iam-On, N. (2018). Cluster ensembles: A survey of approaches with recent extensions and applications. Computer Science Review, 28:1 – 25.*
- *Antoine Cornuéjols, Cédric Wemmert, Pierre Gançarski, Younès Bennani. Collaborative Clustering: Why, When, What and How. Information Fusion, Elsevier, 2018, 39, pp.81-95. ⟨10.1016/j.inffus.2017.04.008⟩. ⟨hal-02327993⟩*

References

- *Weighted Clustering Ensemble: A Review Mimi Zhang School of Computer Science and Statistics, Trinity College Dublin, Dublin 2, Ireland*
- *Banerjee, A., Pujari, A. K., Rani Panigrahi, C., Pati, B., Chandan Nayak, S., & Weng, T. H. (2021). A new method for weighted ensemble clustering and coupled ensemble selection. Connection Science, 1-22.*
- *Zhang, X., Yang, Y., Li, T., Zhang, Y., Wang, H., & Fujita, H. (2021). CMC: A consensus multi-view clustering model for predicting Alzheimers disease progression. Computer Methods and Programs in Biomedicine, 199, 105895.*

Recent references

- *Nazari, A., Dehghan, A., Nejatian, S., Rezaie, V., and Parvin, H. (2017). A comprehensive study of clustering ensemble weighting based on cluster quality and diversity. Pattern Analysis and Applications.*
- *Huang, D.,Wang, C., and Lai, J. (2018). Locally weighted ensemble clustering. IEEE Transactions on Cybernetics, (99):1–14.*
- *Boongoen, T. and Iam-On, N. (2018). Cluster ensembles: A survey of approaches with recent extensions and applications. Computer Science Review, 28:1 – 25.*
- *Antoine Cornuéjols, Cédric Wemmert, Pierre Gançarski, Younès Bennani. Collaborative Clustering: Why, When, What and How. Information Fusion, Elsevier, 2018, 39, pp.81-95. ⟨10.1016/j.inffus.2017.04.008⟩. ⟨hal-02327993⟩*

Sparse Subspace K-means

Abdoul Wahab Diallo & Ndèye Niang & Mory Ouattara

AIMS, Mbour, Sénégal
abdoul.w.m.diallo@aims-senegal.org

CEDRIC CNAM Paris, France
ndeye.niang_keita@cnam.fr

Laboratoire de Mathématique Informatique
UFR SFA, Université Nangui Abrogoua,
Abidjan, Côte d'ivoire
ouattaramory.sfa@univ-na.ci

MSDMA December 17, 2021