FEMCA: A flexible clustering algorithm for noisy data

Frédéric Pascal
Joint work with V. Roizman & M. Jonckheere

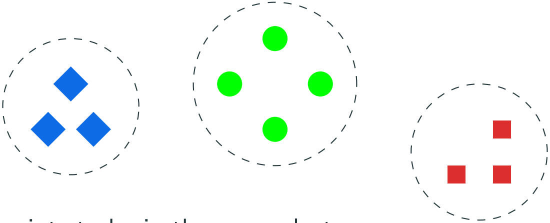
Séminaire de Statistique, CNAM
March 29th, 2024

L2S, CentraleSupélec, University Paris-Saclay, France
http://fredericpascal.blogspot.fr

Motivation

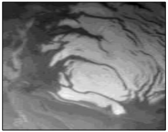
Motivation

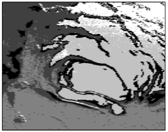
Group data points into clusters to understand the structure of the data.



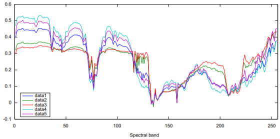
- similar points to be in the same cluster,
- really different points to be in different clusters, and
- well separated clusters.

Image segmentation





Functions/Time-series



Bouveyron et al. (2007)

Some challenges for clustering

Heterogeneous datasets

- Datasets with outliers/noise.
- Heavy tails distributions.
- Different scales/distributions.
- Continuous and discrete data.

Some challenges for clustering

Heterogeneous datasets

- Datasets with outliers/noise.
- Heavy tails distributions.
- Different scales/distributions.
- Continuous and discrete data.

Lots of data ($n \gg$)

- High computational cost.
- Needs of parallelization / batch versions.

Some challenges for clustering

Heterogeneous datasets

- Datasets with outliers/noise.
- Heavy tails distributions.
- Different scales/distributions.
- Continuous and discrete data.

Lots of data ($n \gg$)

- High computational cost.
- Needs of parallelization / batch versions.

High dimensional context ($m \gg$)

- ill-posed problems.
- Data on manifolds.

$\Rightarrow$ Regularization/penalization, dimensionality reduction

Some challenges for clustering

Focus here on:

Heterogeneous datasets

- Datasets with outliers/noise.
- Different scales/distributions.

Some challenges for clustering

Focus here on:

Heterogeneous datasets

- Datasets with outliers/noise.
- Different scales/distributions.

We address "not too high dimensions" regimes (say 30-100).

Reference paper: [Roizman et al., 2023] Roizman, V., Jonckheere, M., & Pascal, F. (2023). A flexible EM-like clustering algorithm for noisy data. *IEEE Transactions on Pattern Analysis and Machine Intelligence*.

Outline

State of the art

K-means

Given $\{\mathbf{x}_i\}_{i=1}^n$, find $\hat{\mathbf{C}} = \{C_1, ..., C_K\}$ with $\boldsymbol{\mu}_k = \frac{1}{\#(C_k)} \sum_{\mathbf{x} \in C_k} \mathbf{x}$ such that

$$\hat{\mathbf{C}} = \underset{\mathbf{C}=\{C_1,\ldots,C_K\}}{argmin} \sum_{k=1}^{K} \sum_{\mathbf{x} \in C_k} \|\mathbf{x} - \boldsymbol{\mu}_k\|_2^2$$

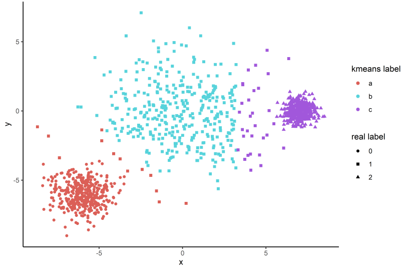
Plain optimization problem.

Simple idea. ✓

Very fast. ✓

Works well only when: ✗

- round-shaped clusters,
- with similar variance, and
- well-separated.

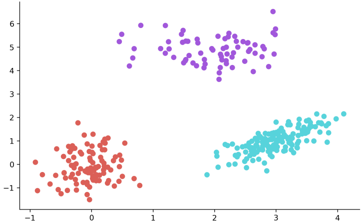


Gaussian Mixture Model (GMM)

We model data as a mixture of Gaussian distributions $\mathcal{N}(\boldsymbol{\mu}_k, \mathbf{M}_k)$:

$$f(\mathbf{x}) = \sum_{k=1}^{K} \pi_k f_k(\mathbf{x}),$$

with π_k the proportion of cluster k and f_k the normal p.d.f.



$$f_k(\mathbf{x}) = \frac{1}{(2\pi)^{m/2}|\mathbf{M}_k|^{1/2}} \exp\left[-\frac{(\mathbf{x}-\boldsymbol{\mu}_k)^T\mathbf{M}_k^{-1}(\mathbf{x}-\boldsymbol{\mu}_k)}{2}\right]$$

Expectation-Maximization (EM) algorithm

Statistical algorithm to estimate parameters based on a **likelihood**.

In the GMM case, we would need the **labels** of the data points to estimate the parameters. **Labels** $\rightarrow$ **Latent variables**

Expectation-Maximization (EM) algorithm

Statistical algorithm to estimate parameters based on a **likelihood**.

In the GMM case, we would need the **labels** of the data points to estimate the parameters. **Labels** $\rightarrow$ **Latent variables**

E-STEP

Computation of the membership a posteriori probabilities

$$p_{ik} = P(Z_i = k | \mathbf{X}_i = \mathbf{x}_i) = \frac{\pi_k f_k(\mathbf{x}_i)}{\sum_{j=1}^{K} \pi_j f_j(\mathbf{x}_i)}$$

with f_k the Gaussian p.d.f.

Expectation-Maximization (EM) algorithm

Statistical algorithm to estimate parameters based on a **likelihood**.

In the GMM case, we would need the **labels** of the data points to estimate the parameters. **Labels** $\rightarrow$ **Latent variables**

E-STEP

Computation of the membership a posteriori probabilities

$$p_{ik} = P(Z_i = k | \mathbf{X}_i = \mathbf{x}_i) = \frac{\pi_k f_k(\mathbf{x}_i)}{\sum_{j=1}^{K} \pi_j f_j(\mathbf{x}_i)}$$

with f_k the Gaussian p.d.f.

M-STEP

Estimation of the parameters

$$\widehat{\pi}_k = \frac{1}{n} \sum_{i=1}^{n} p_{ik}$$

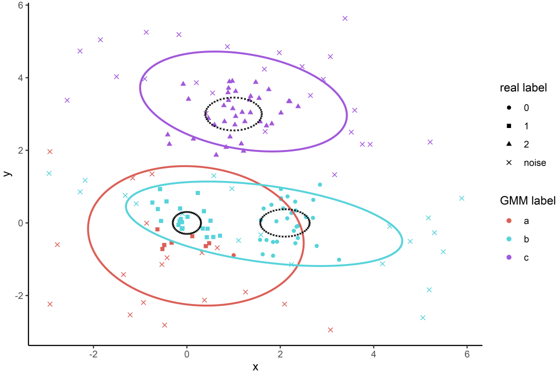
$$\widehat{\boldsymbol{\mu}}_k = \frac{1}{n\widehat{\pi}_k} \sum_{i=1}^{n} p_{ik} \mathbf{x}_i$$

$$\widehat{\mathbf{M}}_k = \frac{1}{n\widehat{\pi}_k} \sum_{i=1}^{n} p_{ik} (\mathbf{x}_i - \widehat{\boldsymbol{\mu}}_k)(\mathbf{x}_i - \widehat{\boldsymbol{\mu}}_k)^T$$

What happens to GMM when the data has some noise or non Gaussian data?

The GMM has problems to cluster and estimate parameters for data with noise, different distribution shapes and outliers.

Result with data contaminated:



What happens to GMM when the data has some noise or is non Gaussian?

Why?

- The **estimators** are **not robust**.
- **Mismatch** between the **model** and the data.
- **No outlier rejection**.

How to address the robustness challenge?

There are mainly two directions to **robustify clustering methods** in the literature:

- model **generalizations**
  - Extra uniform cluster [Banfield and Raftery, 1993]
  - Model low density areas (RIMLE and OTRIMLE) [Coretto and Hennig, 2016]
  - Mixture of $t-$distributions (t-EM) [Peel and McLachlan, 2000]
- models that introduce **classical robust techniques** in the **estimation**
  - Trimming methods (TCLUST) [García-Escudero et al., 2008]
  - k-tau [Gonzalez et al., 2019] and Spatial-EM [Yu et al., 2015]

Some drawbacks

Some **drawbacks** of the state of the art robust clustering methods:

- No closed equations on the M-step, reliance on non-linear optimizers (t-EM).
- Extra parameters difficult to be tuned (RIMLE, TCLUST). *e.g.*, if we misspecify the proportion of noise in the TCLUST algorithm [Gonzalez et al., 2019].
- Models are too specific.

Some drawbacks

Some **drawbacks** of the state of the art robust clustering methods:

- No closed equations on the M-step, reliance on non-linear optimizers (t-EM).
- Extra parameters difficult to be tuned (RIMLE, TCLUST). *e.g.*, if we misspecify the proportion of noise in the TCLUST algorithm [Gonzalez et al., 2019].
- Models are too specific.

Our goal:

- **flexibility to very general models**
- **no extra parameters**

FEMCA: Model, derivation and properties

A very general model

We consider $\mathbf{x}_1, \ldots, \mathbf{x}_n$ independent vectors.

These vectors belong to some clusters $C_1, \ldots, C_K$.

$\mathbf{x}_1, \ldots, \mathbf{x}_n$ **ARE NOT i.i.d. !**

Cluster characterization

$\mathbf{x}_i$ and $\mathbf{x}_j$ belong to C_k if they are drawn from a distribution with the same features

$$\boldsymbol{\mu}_k \quad \text{and} \quad \boldsymbol{\Sigma}_k$$

The **location** and the **scatter matrix** are the **features** that characterize the clusters and not a particular distribution as in GMM or t-EM.

FEMCA: A flexible algorithm relying on a very general model

FEMCA is based on a model where the $\mathbf{x}_1, \ldots, \mathbf{x}_n$ independent vectors are characterized by

Stochastic representation

$$\mathbf{x}_i \in C_k \Rightarrow \mathbf{x}_i \stackrel{d}{=} \boldsymbol{\mu}_k + \sqrt{\mathcal{Q}_{ik}}\ \sqrt{\tau_{ik}}\ \mathbf{A}_k\ \mathbf{u}_i$$

- $\boldsymbol{\mu}_k$ is the mean of the cluster k.
- $\mathcal{Q}_{ik}$ is an independent positive random variable.
- τ_{ik} **are scale (nuisance) parameters that increase the flexibility of the model.**
- $\mathbf{A}_k$ is such that $\mathbf{A}_k^T \mathbf{A}_k = \boldsymbol{\Sigma}_k$ (the scatter matrix of the cluster k).
- $\mathbf{u}_i$ is a uniform vector on the unit hyper-sphere.

$\cdot_{ik}$ represents the possible dependence on k and i.

Elliptical Symmetric family

The stochastic characterization [Cambanis et al., 1981] represents vectors of the Elliptical Symmetric family [Kelker, 1970].

The density can be written as

$$\boxed{f_{\mathbf{x}_i}(\mathbf{x}) = A_m |\tau_{ik}\boldsymbol{\Sigma}_k|^{-1/2} \;\; \mathbf{g}_{ik}\left(\tau_{ik}^{-1}(\mathbf{x}-\boldsymbol{\mu}_k)^T \boldsymbol{\Sigma}_k{}^{-1}(\mathbf{x}-\boldsymbol{\mu}_k)\right)}$$

for some function $\mathbf{g}_{ik}$ called the **density generator**. We denote it as $\mathbf{x} \sim \mathrm{ES}(\boldsymbol{\mu}_k, \tau_{ik}\boldsymbol{\Sigma}_k, \mathbf{g}_{ik})$.

Distributions caracterization

One-to-one relation between $\mathbf{g}_{ik}$ and $\mathcal{Q}_{ik}$

$\Rightarrow$ the **shape** of the distributions

This family includes **Gaussian**, $t-$distribution, Generalized Gaussian distribution. Heavier and lighter (than Gaussian) tails.

Different scenarios

We consider different scenarios based on the nature of the **density generator functions**:

$$g_{ik} = \begin{cases} g_i, & \textbf{each point might come from} \neq \textbf{shaped dist.} \\ & \textbf{BUT shapes do not depend on the cluster} \\ g, & \textbf{the density generator function is} \\ & \textbf{always the same (e.g., Gaussian case)} \\ g_k, & \text{cluster dependent shapes} \\ & \text{extra parameters have to be computed } (\textit{e.g.}, \text{ t-EM}) \end{cases}$$

FEMCA: A flexible algorithm relying on a very general model

Parameter space

Given $\{\mathbf{x}_i\}_{i=1}^n \in \mathbb{R}^m$ we have to estimate the usual parameters

$$\boldsymbol{\Theta} = \{(\pi_k, \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)\}_{k=1,..,K}$$

AND we now have a lot of (nuisance) parameters τ

$$\widetilde{\boldsymbol{\Theta}} = \{\tau_{ik}\}_{\substack{k=1,..,K \\ i=1,..,n}}$$

MLE

We derive the two-step (E-M) algorithm based on the likelihood of the model (using the trick of [Ollila and Tyler, 2012]).

E-step: First Miracle

Proposition

Assume $g_{ik} = g_i$, then the membership probabilities MLE are

$$\widehat{p}_{ik} = \frac{\widehat{\pi}_k \left((\mathbf{x}_i - \widehat{\boldsymbol{\mu}}_k)^T \widehat{\boldsymbol{\Sigma}}_k^{-1} (\mathbf{x}_i - \widehat{\boldsymbol{\mu}}_k) \right)^{-m/2} |\widehat{\boldsymbol{\Sigma}}_k|^{-1/2}}{\sum_{j=1}^{K} \widehat{\pi}_j \left((\mathbf{x}_i - \widehat{\boldsymbol{\mu}}_j)^T \widehat{\boldsymbol{\Sigma}}_j^{-1} (\mathbf{x}_i - \widehat{\boldsymbol{\mu}}_j) \right)^{-m/2} |\widehat{\boldsymbol{\Sigma}}_j|^{-1/2}}.$$

Insensitivity: the expression of the membership **does not** depend on the particular density g_i that generates each data point, neither on the τ_{ik}

Proof details: see [Roizman et al., 2023]

M-step: Second Miracle

Proposition (Location and scatter matrix estimators)

We almost obtain Tyler's estimators.

$$\widehat{\boldsymbol{\mu}}_k = \frac{\displaystyle\sum_{i=1}^{n} \frac{\hat{p}_{ik}\mathbf{x}_i}{(\mathbf{x}_i - \widehat{\boldsymbol{\mu}}_k)^T \widehat{\boldsymbol{\Sigma}}_k{}^{-1}(\mathbf{x}_i - \widehat{\boldsymbol{\mu}}_k)}}{\displaystyle\sum_{i=1}^{n} \frac{\hat{p}_{ik}}{(\mathbf{x}_i - \widehat{\boldsymbol{\mu}}_k)^T \widehat{\boldsymbol{\Sigma}}_k^{-1}(\mathbf{x}_i - \widehat{\boldsymbol{\mu}}_k)}}$$

$$\widehat{\boldsymbol{\Sigma}}_k = m\sum_{i=1}^{n} \frac{w_{ik}(\mathbf{x}_i - \widehat{\boldsymbol{\mu}}_k)(\mathbf{x}_i - \widehat{\boldsymbol{\mu}}_k)^T}{(\mathbf{x}_i - \widehat{\boldsymbol{\mu}}_k)^T \widehat{\boldsymbol{\Sigma}}_k{}^{-1}(\mathbf{x}_i - \widehat{\boldsymbol{\mu}}_k)}, \quad \textit{with} \quad w_{ik} = \hat{p}_{ik} \Big/ \sum_i \hat{p}_{ik}$$

And the τ_{ik}'s?

Furthermore,

Proposition (τ_{ik} estimator)

$$\widehat{\tau}_{ik} = \frac{(\mathbf{x}_i - \widehat{\boldsymbol{\mu}}_k)^T \widehat{\boldsymbol{\Sigma}}_k^{-1} (\mathbf{x}_i - \widehat{\boldsymbol{\mu}}_k)}{a_{ik}},$$

where a_{ik} depends only on g_{ik}

$$a_{ik} = \arg\sup_t \{t^{m/2} g_{i,k}(t)\}$$

.

e.g., for the Gaussian case $a_{ik} = m$.

Estimators intuitively

$\widehat{\boldsymbol{\mu}}_k$ and $\widehat{\boldsymbol{\Sigma}}_k$ are like usual sample estimators with small weights for outlying points

$$\frac{1}{n}\sum_{i=1}^{n}\mathbf{x}_i \Longrightarrow \frac{1}{n}\sum_{i=1}^{n} w_i\mathbf{x}_i$$

$$\frac{1}{n}\sum_{i=1}^{n}(\mathbf{x}_i - \widehat{\boldsymbol{\mu}})(\mathbf{x}_i - \widehat{\boldsymbol{\mu}})^T \Longrightarrow \frac{1}{n}\sum_{i=1}^{n} w_{ik}(\mathbf{x}_i - \widehat{\boldsymbol{\mu}})(\mathbf{x}_i - \widehat{\boldsymbol{\mu}})^T$$

with $w_{ik} \approx \frac{\hat{p}_{ik}}{(\mathbf{x}_i - \widehat{\boldsymbol{\mu}}_k)^T \widehat{\boldsymbol{\Sigma}}_k^{-1}(\mathbf{x}_i - \widehat{\boldsymbol{\mu}}_k)}$

Tyler estimators [Tyler, 1987] (classical robust estimator [Maronna, 1976]) fulfill very similar equations. **HINT** about robustness of the model.

About FEMCA

Properties

- The random vectors that represent the data points are independent but not necessarily i.i.d.
- Generalizes GMM. (Gaussian $\in$ ES)
- If $g_{ik} = g_i$, the membership probabilities do not depend on the shape of the distributions!
- If $g_{ik} = g_k$, we can derive extra estimators to be computed on the M-Step.
- The model leads to estimators that are similar to classical robust estimators (Tyler) [Ollila and Tyler, 2012].

If the dimension grows... Some hints

When the dimension grows we can better estimate the parameters τ_{ik}.

Convergence of $\hat{\tau}$ when g is the Gaussian density generator

Let $\mathbf{x} \stackrel{d}{=} \boldsymbol{\mu} + \sqrt{\tau}\mathbf{A}\mathbf{q}$, with $\mathbf{q}$ a standard Gaussian. Under some assumptions, for any $a \in \mathbb{R}$, $\forall \varepsilon > 0$ and $\mathbf{y} \sim \mathcal{N}(\tau, 2\tau^2/m)$, then

$$|\mathbb{P}(\{\widehat{\tau} \leq a\}) - \mathbb{P}(\mathbf{y} \leq a)| < \varepsilon, \text{ if } n \text{ and } m \text{ are large enough}$$

This is in agreement with previous RMT results [Couillet et al., 2014].

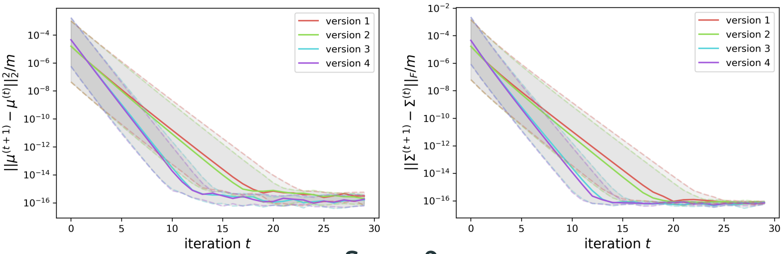
We can combine this result with parsimonious restrictions on the covariance matrix to avoid issues in the case of **very large** m.

Implementation details

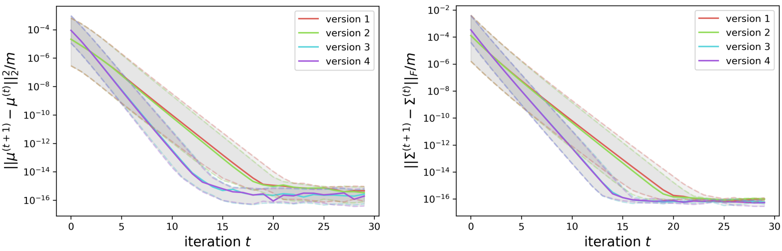
- The trace of the scatter matrix estimator is fixed.
- Four slightly different versions were propsed:
  - **Version 1:** the parameter $\boldsymbol{\mu}$ used to compute the estimator $\boldsymbol{\Sigma}$ is the one obtained in the **same iteration** of the fixed-point loop.
  - **Version 2:** the $\boldsymbol{\mu}$-parameter is the one obtained in the **previous iteration**.
  - **Version 3:** $(\mathbf{x}_i - \widehat{\boldsymbol{\mu}}_k)^T \widehat{\boldsymbol{\Sigma}}_k^{-1} (\mathbf{x}_i - \widehat{\boldsymbol{\mu}}_k)$ for $\boldsymbol{\mu}$ are replaced by their square root (original Tyler M-estimators).
  - **Version 4:** Version 3 on top of the algorithm of Version 2.
- Center initialization: quick run of k-means.
- Code available: `github.com/violetr/fem`

Convergence of the fixed-point loops

Setup 1



Setup 2



Outlier rejection

It is possible to use some heuristic outlier rejection methods based on

$$\Delta(\mathbf{x}_i; \widehat{\boldsymbol{\mu}}_k, \widehat{\mathbf{M}}_k) = (\mathbf{x}_i - \widehat{\boldsymbol{\mu}}_k)^T \widehat{\mathbf{M}_k}^{-1} (\mathbf{x}_i - \widehat{\boldsymbol{\mu}}_k) \sim ?.$$

Threshold to reject $= 1 - \alpha$ quantile of the distribution.

We developed some alternatives [Roizman et al., 2020] based on a scaled Fisher distribution [Drašković and Pascal, 2018].

Rejection block

It has been implemented and plugged in at the end of the FEMCA (OPTIONAL).

FEMCA: Experimental results

Measuring the performance

We compare our algorithm to

- k-means
- GMM-EM
- Spectral Clustering
- Mixture of Student's t (t-EM or EMMIX)
- TClust
- RIMLE

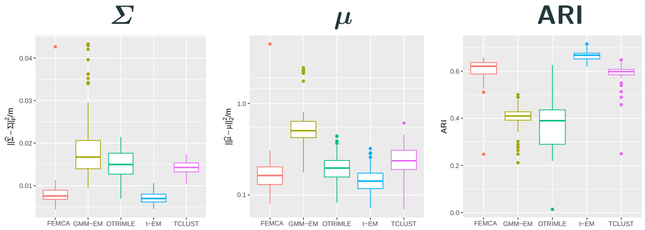
Metrics

- Adjusted Mutual Information (AMI),
- Adjusted Rand Index (AR).
- Estimation error of the parameters (only for simulations).

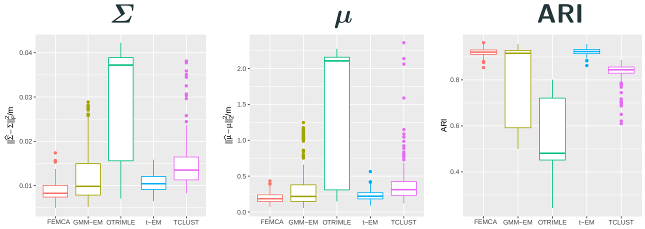
Some simulation results

Mixtures of t-distributions with different degrees of freedom and covariance matrix classes, mixtures of more general distributions, clusters with different g_i.

Setup 1:
$t-$distributions
ν small

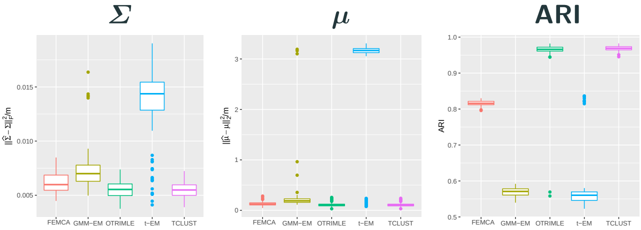


Setup 2:
$t-$distributions
$\nu = 10$

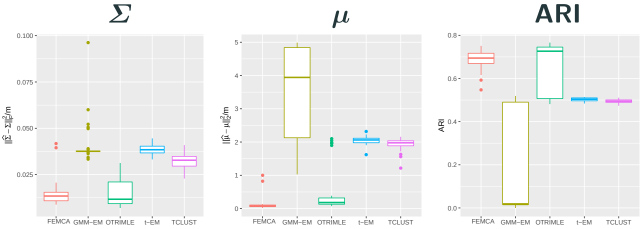


Some simulation results

Setup 3: Gaussian + uniform noise

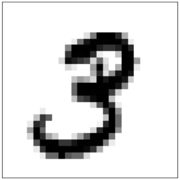
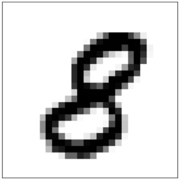
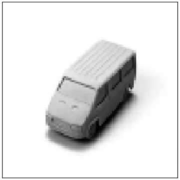


Setup 4: Elliptical different g_i



FEMCA performs well even in the situations that do not match the model.

Real data clustering results

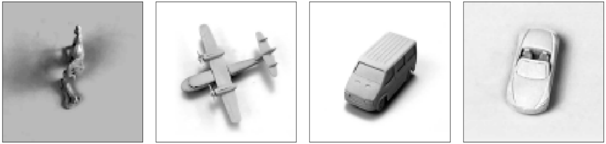

MNIST (LeCun, 1998) **NORB** (LeCun, 2004)

Set	k-means	GMM	t-EM	FEMCA	spectral	TCLUST	RIMLE
MNIST38	0.2884	0.5716	0.6397	**0.6887**	0.6866	0.6847	0.2494
MNIST71	0.8486	0.8905	**0.9432**	0.9360	0.9384	0.6885	0.2493
MNIST386	0.6338	0.7332	0.8262	0.8306	**0.8542**	0.8366	0.4274
MNIST386+n	0.4475	0.4909	0.5296	0.5548	0.3115	**0.6908**	0.1498
smallNORB	0.0015	0.0468	0.4223	**0.5067**	~ 0	0.1330	0.1472
20news	0.1883	0.2739	0.4426	**0.5114**	0.0987	0.2664	0.0026

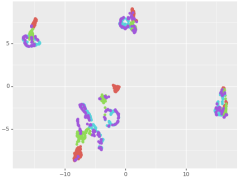
Table 1: Median AMI

Real data clustering results - The NORB case

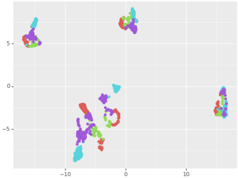
Dataset	kmeans	GMM-EM	t-EM	FEMCA	spectral	TCLUST	RIMLE
small NORB	**0.0015**	**0.0468**	0.4223	**0.5067**	**~ 0**	**0.1330**	**0.1472**



t-SNE embedding of the dataset colored with labels:



real labels



FEMCA labels

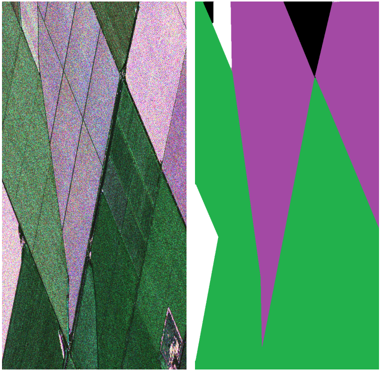
GMM-EM labels

Applications to PolSAR Images

Applications to PolSAR Images

Land use segmentation

Joint work with V. Roizman and G. Draskovic. [Roizman et al., 2019]



Change detection

Joint work with V. Roizman, G. Ginolhac and M. Jonckheere.

two times

$t_0 \rightarrow t_1$

one region, two images

$I_{t_0} \rightarrow I_{t_1}$

CD map

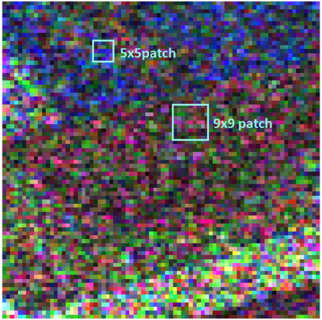
$\mathrm{CD}(x_l) \in \{0, 1\}$

Extension for PolSAR images segmentation

Segment PolSAR images [Conte et al., 2002, Gini et al., 2000] with a clustering algorithm to detect land use.

Keep **flexibility** but also take advantage of **spatial structure**.

Compute by **patches** $\rightarrow$ R-EM.



R-EM: Modifying FEMCA

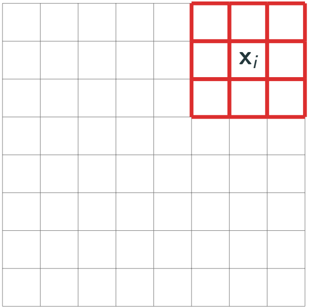
We propose a modification to include spacial information (and to deal with small dimension...). We estimate the membership probabilities p_{ij} based on $\Delta_{ik}^{(l)}$, computed over all the neighbors.

For each pixel $\mathbf{x}_i$:

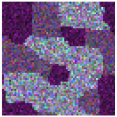
For each pixel $\mathbf{x}_t$ in the patch of $\mathbf{x}_i$:

$$\Delta_{tk}^{(l)} = (\mathbf{x}_t - \boldsymbol{\mu}_k^{(l)})^T (\boldsymbol{\Sigma}_k^{(l)})^{-1} (\mathbf{x}_t - \boldsymbol{\mu}_k^{(l)})$$

Set $\Delta_{ik}^{(l)} = h(\{\Delta_{tk}^{(l)}\}_t)$



For different patch sizes and different $h(x)$ summary functions as mean, median and trimmed mean.

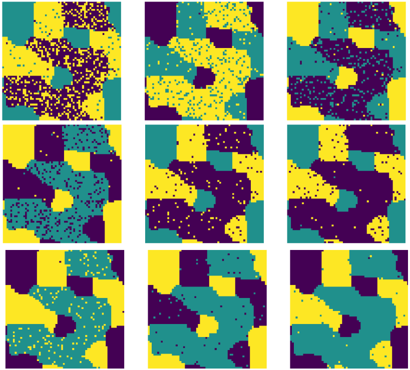
Image example  **Classes** 

From left to right: k-means, GMM and R-EM

6-looked

9-looked

12-looked



Clustering accuracy

	n-looked		
	6	**9**	**12**
k-means	0.85	0.82	0.96
GMM	**0.92**	0.88	0.98
R-EM	**0.92**	**0.91**	**0.99**

Conclusions and Perspectives

Conclusions

- We developed a **very general flexible** clustering algorithm based on Elliptical Symmetric distributions.
- We proved some **interesting properties**.
- We showed a **good performance of FEMCA on experiments**.
- We applied the flexible clustering algorithm to **PolSAR image problems**.

Perspectives

- Implementation of **regularizations** methods when $m > n$.
- Design a **model selection** method more specific than AIC/BIC.

- Study of the **consistency of the estimators** and the **behaviour of the $\boldsymbol{\tau}_{ik}$** estimation.
- Apply the parameter τ_{ik} addition to **other similar Machine Learning problems** that include covariance matrices.

Thanks for your attention!

Questions ?

References i

Banfield, J. D. and Raftery, A. E. (1993).
Model-based gaussian and non-gaussian clustering.
Biometrics, 49(3):803–821.

Cambanis, S., Huang, S., and Simons, G. (1981).
On the theory of elliptically contoured distributions.
Journal of Multivariate Analysis, 11(3):368–385.

Conte, E., De Maio, A., and Ricci, G. (2002).
Covariance matrix estimation for adaptive CFAR detection in compound-gaussian clutter.
IEEE Transactions on Aerospace and Electronic Systems, 38(2):415–426.

Coretto, P. and Hennig, C. (2016).
Robust improper maximum likelihood: Tuning, computation, and a comparison with other methods for robust gaussian clustering.
Journal of the American Statistical Association, 111(516):1648–1659.

Couillet, R., Pascal, F., and Silverstein, J. W. (2014).
Robust estimates of covariance matrices in the large dimensional regime.
IEEE Transactions on Information Theory, 60(11):7269–7278.

Drašković, G. and Pascal, F. (2018).
New insights into the statistical properties of M-estimators.
Signal Processing, IEEE Transactions on, 66(16):4253–4263.

García-Escudero, L. A., Gordaliza, A., Matrán, C., and Mayo-Iscar, A. (2008).
A general trimming approach to robust cluster analysis.
Ann. Statist., 36(3):1324–1345.

Gini, F., Greco, M. V., Diani, M., and Verrazzani, L. (2000).
Performance analysis of two adaptive radar detectors against non-gaussian real sea clutter data.
IEEE Transactions on Aerospace and Electronic Systems, 36(4):1429–1439.

Gonzalez, J. D., Yohai, V. J., and Zamar, R. H. (2019).
Robust Clustering Using Tau-Scales.
arXiv e-prints, page arXiv:1906.08198.

Kelker, D. (1970).
Distribution theory of spherical distributions and a location-scale parameter generalization.
32(4):419–430.

References iii

Maronna, R. A. (1976).
Robust M-Estimators of multivariate location and scatter.
The Annals of Statistics, 4(1):51–67.

Ollila, E. and Tyler, D. E. (2012).
Distribution-free detection under complex elliptically symmetric clutter distribution.
In *2012 IEEE 7th Sensor Array and Multichannel Signal Processing Workshop (SAM)*, pages 413–416.

Peel, D. and McLachlan, G. J. (2000).
Robust mixture modelling using the t distribution.
Statistics and Computing, 10(4):339–348.

Roizman, V., Draskovic, G., and Pascal, F. (2019).
A new clustering algorithm for PolSAR images segmentation.
In *IEEE CAMSAP 2019 - IEEE International Workshop on Computational Advances in Multi-Sensor Adaptive Processing*, Guadeloupe, West Indies, France.

Roizman, V., Jonckheere, M., and Pascal, F. (2020).
Robust clustering and outlier rejection using the mahalanobis distance distribution.
In *28th European Signal Processing Conference, EUSIPCO 2018*. IEEE.

Roizman, V., Jonckheere, M., and Pascal, F. (2023).
A flexible em-like clustering algorithm for noisy data.
IEEE Transactions on Pattern Analysis and Machine Intelligence.

Tyler, D. E. (1987).
A distribution-free M-estimator of multivariate scatter.
The Annals of Statistics, 15(1):234–251.

Yu, K., Dang, X., Bart, H., and Chen, Y. (2015).
Robust model-based learning via spatial-em algorithm.
IEEE Transactions on Knowledge and Data Engineering, 27(6):1670–1682.