# A tractable Multi-Partitions Clustering 

## Vincent VANDEWALLE ${ }^{1,2}$ Joint work with Matthieu MARBAC ${ }^{3}$

${ }^{1}$ Univ. Lille, EA2694 Santé publique: épidémiologie et qualité des soins ${ }^{2}$ Inria
${ }^{3}$ CREST, ENSAI

COMPSTAT 2018
Friday $31^{\text {th }}$ August 2018
lasi

## Outline

(1) Extension of the variable selection to variable clustering

- Variable selection in clustering
- Multiple Gaussian Mixture
- Proposed Multiple Partitions Mixture
- Properties of the model
(2) Parameters estimation and model selection
- Maximum likelihood inference
- Penalized observed-data likelihood
- Integrated complete-data likelihood
(3) Numerical experiments
- NBA team data
- Wine data
(4) Conclusion and perspectives


## Data

$\mathbf{x}=\left(\boldsymbol{x}_{1}, \ldots, \boldsymbol{x}_{n}\right)$ composed of $n$ independent observations $\boldsymbol{x}_{i}=\left(x_{i 1}, \ldots, x_{i d}\right)$ defined on $\mathbb{R}^{d}$.

## Goal

Cluster the data in $G$ clusters.

## What variables use in clustering?

- Well-posed problem in the supervised classification setting with objective criteria: error rate, AUC, ...
- III-posed problem in clustering since the class variable is not known by advance. Thus, what are the most relevant variables with respect to this unknown variable?
- Pragmatic solution 1: Prior choice of the practitioner among available variables (according to some focus)
- Pragmatic solution 2: Posterior analysis of the correlation between the predicted cluster (based on all the variables) and each variable


## The model based clustering solution

- Mixture models allow to perform clustering by modelling the distribution of the data as a mixture of $G$ components each one corresponding to a cluster.
- Thus possibility to suppose that some variables do not depend (directly) on the cluster in the probabilistic model.


## Some references

- Raftery \& Dean (2006): some classifying, and some redundant variables. Redundant variables independent of the cluster given the classifying variables.
- Maugis \& al. (2009): refinement of Raftery \& Dean (2006) by specifying the role of each variable (classifying, redundant or independant).


## Advantages of these approaches

- Improve the accuracy of the clustering by decreasing the variance of the estimators.
- Allow some specific interpretation of the classifying variables.


## Limitations of these approaches

- Combinatorial problem to select the best model with these refined approaches
- Search too hard to perform when the number of variables is large


## Solution: use simpler models for a better search (Marbac \& Serdki 2016,2017)

- Assumption of conditional independence of the classifying variables given the cluster
- Non-classifying variables are independent
- Optimisation of the integrated classification likelihood (ICL)
- Better results than previous approaches on large number of variables with moderated sample size
- The independence assumption allows to easily consider the heterogeneous data setting


## Several clustering variables

- The variables in the data can convey several clustering view points with respect to different groups of variables
- Allow to find some clustering which could be hidden by other variables


## Some references

- Galimberti \& al. (2007): First proposition of a multiple Gaussian mixture model
- Galimberti \& al. (2017): Refinement of the previous model with ideas similar to to Raftery \& Dean (2006) et Maugis \& al. (2009)


## Remarks

- Smart modelling of the role of each variable
- Search hard to perform when the number of variables is large
- Specific to the Gaussian setting

Extension of the variable selection to variable clustering

## Illustration of a multiple partition

| id | $\mathrm{x}_{1}$ | $\mathrm{x}_{2}$ | $\mathrm{x}_{3}$ | $\mathrm{x}_{4}$ | $\mathbf{z}_{1}$ | $\mathbf{z}_{2}$ |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| 1 | 3.23 | 3.28 | 3.14 | 4.08 | 1 | 1 |
| 2 | 4.26 | 4.7 | 4.41 | 5.36 | 1 | 1 |
| 3 | 6.43 | 3.94 | -6.09 | -3.03 | 1 | 2 |
| 4 | 2.93 | 3.22 | -5.05 | -3.2 | 1 | 2 |
| 5 | 3.77 | 4.88 | -3.21 | -4 | 1 | 2 |
| 6 | -2.03 | -4.9 | 2.81 | 4 | 2 | 1 |
| 7 | -2.78 | -5.87 | 2.11 | 3.57 | 2 | 1 |
| 8 | -2.38 | -4.06 | 3.35 | 4.91 | 2 | 1 |
| 9 | -4.88 | -5.26 | -2.86 | -4.42 | 2 | 2 |
| 10 | -5.01 | -4.83 | -3.3 | -6.42 | 2 | 2 |
| 11 | -3.25 | -5.84 | -5.23 | -5.35 | 2 | 2 |
| 12 | -4.28 | -4.36 | -3.38 | -5.14 | 2 | 2 |



## Main assumptions

- The variables can be decomposed in $B$ independent blocks.
- The block $b$ follows a mixture with $G_{b}$ components (for $b=1, \ldots, B$ ), with the assumption of class conditional independence of the variables.


## Notations

- $\boldsymbol{\omega}=\left(\omega_{j} ; j=1, \ldots, d\right)$ the repartition of the variables in blocks; $\omega_{j}=b$ if variable $j$ belongs to block $b$.
- $\boldsymbol{m}=\left(G_{1}, \ldots, G_{B}, \boldsymbol{\omega}\right)$ defines the model
- $\Omega_{b}=\left\{j: \omega_{j}=b\right\}$ the subset of variables belonging to block $b$
- $\boldsymbol{\theta}=(\boldsymbol{\pi}, \boldsymbol{\alpha})$ model parameters
- $p\left(. \mid \boldsymbol{\alpha}_{j g}\right)$ the pdf of the distribution of parameters $\boldsymbol{\alpha}_{j g}$.


## Probability distribution function of $\boldsymbol{x}_{i}$

$$
p\left(\boldsymbol{x}_{i} \mid \boldsymbol{m}, \boldsymbol{\theta}\right)=\prod_{b=1}^{B} p\left(\boldsymbol{x}_{i\{b\}} \mid \boldsymbol{m}, \boldsymbol{\theta}\right) \text { with } p\left(\boldsymbol{x}_{i\{b\}} \mid \boldsymbol{m}, \boldsymbol{\theta}\right)=\sum_{g=1}^{G_{b}} \pi_{b g} \prod_{j \in \Omega_{b}} p\left(x_{i j} \mid \boldsymbol{\alpha}_{j g}\right),
$$

Illustration :

- $n=100$ from a MPM with $B=2$ blocks of two variables.
- Variable 1 and 2 belong to block 1 and variables 3 and 4 in block 2
- Each block follows a bi-component Gaussian mixture (i.e., $G_{b}=2$ ) with equal proportions (i.e., $\pi_{b g}=1 / 2$ ) and $\mu_{j 1}=4, \mu_{j 2}=-4$ and $\sigma_{j g}=1$.



## Remarks

- Different partitions explained by subsets of variables.
- Generalizes approaches used for variable selection in model-based clustering (if $B=2$ and $G_{1}=1$ then variables belonging to block 1 are not relevant for the clustering, while variables belonging to block 2 are relevant)
- MGMM permits variable selection and multiple partitions explained by subsets of variables (variables classification).
- Sparse model: number of parameters $\nu_{m}=\sum_{b=1}^{B}\left(G_{b}-1\right)+G_{b} \sum_{j \in \Omega_{b}} \nu_{j}$
- Better model search expected than in the model of Galimberti \& al. (2017)
- Natural extension to the heterogeneous data setting


## Identifiability

Model identifiability is directly obtained from the identifiability of Gaussian mixture with local independence (Teicher, 1963, 1967).

## Observed-data likelihood for sample x and model $\boldsymbol{m}$

$$
\ell(\boldsymbol{\theta} \mid \boldsymbol{m}, \mathbf{x})=\sum_{b=1}^{B} \sum_{i=1}^{n} \ln \left(\sum_{g=1}^{G_{b}} \pi_{b g} \prod_{j \in \Omega_{b}} p\left(x_{i j} \mid \boldsymbol{\alpha}_{j g}\right)\right)
$$

## Latent partitions

- $B$ independent mixtures
- $\mathbf{z}=\left(\boldsymbol{z}_{i b} ; i=1, \ldots, n ; b=1, \ldots, B\right)$ vectors of the component memberships
- $\boldsymbol{z}_{i b}=\left(z_{i b 1}, \ldots, z_{i b G_{b}}\right)$ where $z_{i b g}=1$ if observation $i$ arose from component $g$ for block $b$, and $z_{i b g}=0$ otherwise


## Completed-data likelihood for sample x and model $\boldsymbol{m}$

$$
\ell(\boldsymbol{\theta} \mid \boldsymbol{m}, \mathbf{x}, \mathbf{z})=\sum_{b=1}^{B} \ln p\left(\mathbf{z}_{b} \mid \boldsymbol{\pi}_{b}\right)+\sum_{j=1}^{d} \ln p\left(\mathbf{x}_{j} \mid \mathbf{z}_{\omega_{j}}, \boldsymbol{\alpha}_{j}\right)
$$

where $\ln p\left(\mathbf{z}_{b} \mid \boldsymbol{\pi}_{b}\right)=\sum_{i=1}^{n} \sum_{g=1}^{G_{b}} z_{i b g} \ln \pi_{b g}$ and $\ln p\left(\mathbf{x}_{j} \mid \mathbf{z}_{b}, \boldsymbol{\alpha}_{j}\right)=\sum_{i=1}^{n} \sum_{g=1}^{G_{b}} z_{i b g} \ln p\left(x_{i j} \mid \boldsymbol{\alpha}_{j g}\right)$.

## EM algorithm

Starting from the initial value $\boldsymbol{\theta}^{[0]}$, iteration $[r]$ is composed of two steps: E-step Computation of the fuzzy partitions $t_{i b g}^{[r]}:=\mathbb{E}\left[Z_{i b g} \mid \boldsymbol{x}_{i\{b\}}, \boldsymbol{m}, \boldsymbol{\theta}^{[r-1]}\right]$, hence for $b=1, \ldots, B$, for $g=1, \ldots, G_{b}$, for $i=1, \ldots, n$

$$
t_{i b g}^{[r]}=\frac{\pi_{b g}^{[r-1]} \prod_{j \in \Omega_{b}} p\left(x_{i j} \mid \boldsymbol{\alpha}_{j g}^{[r-1]}\right)}{\sum_{k=1}^{G_{b}} \pi_{b k}^{[r-1]} \prod_{j \in \Omega_{b}} p\left(x_{i j} \mid \boldsymbol{\alpha}_{j g}^{[r-1]}\right)}
$$

M-step Maximization of the expected value of the complete-data log-likelihood over the parameters,

$$
\pi_{b g}^{[r]}=\frac{n_{b g}^{[r]}}{n} \text { and } \boldsymbol{\alpha}_{j g}^{[r]}=\underset{\boldsymbol{\alpha}_{j g} \in \Theta_{j}}{\arg \max } Q\left(\boldsymbol{\alpha}_{j g} \mid \mathbf{x}_{j}, \mathbf{t}_{\omega_{j} g}^{[r]}\right)
$$

where $Q\left(\boldsymbol{\alpha}_{j g} \mid \mathbf{x}_{j}, \mathbf{t}_{b}\right)=\sum_{i=1}^{n} t_{i b g} \ln p\left(x_{i j} \mid \boldsymbol{\alpha}_{j g}\right)$.

## Remarks

- Independence between the $B$ blocks of variables permits to maximize the observed-data log-likelihood on each block separately.
- Possible modification to perform the block estimation and the parameter inference simultaneously.
- In practice the number of blocks, the repartition of variables into blocks, and the number of classes per block are unknown.


## Model collection $\mathcal{M}$

$$
\mathcal{M}=\left\{\boldsymbol{m}: \omega_{j} \leq B_{\max } \text { and } G_{b} \leq G_{\max } ; j=1, \ldots, d ; b=1, \ldots, B_{\max }\right\}
$$

where $B_{\text {max }}$ is the maximum number of blocks and $G_{\max }$ is the maximum number of components within block.

## Model selection

Model selection often achieved by searching the model $\boldsymbol{m}^{\star}$ maximizing the BIC criterion which is defined by

$$
\operatorname{BIC}(\boldsymbol{m})=\max _{\boldsymbol{\theta}_{\boldsymbol{m}}} \ell_{p e n}\left(\boldsymbol{\theta}_{\boldsymbol{m}} \mid \boldsymbol{m}, \mathbf{x}\right)
$$

where

$$
\ell_{p e n}\left(\boldsymbol{\theta}_{\boldsymbol{m}} \mid \boldsymbol{m}, \mathbf{x}\right)=\ell\left(\boldsymbol{\theta}_{\boldsymbol{m}} \mid \boldsymbol{m}, \mathbf{x}\right)-\frac{\nu_{\boldsymbol{m}}}{2} \ln n
$$

## Remark

$$
\left(\boldsymbol{m}, \boldsymbol{\theta}_{m}\right)
$$

## Penalised completed likelihood

$\ell_{p e n}\left(\boldsymbol{\theta}_{\boldsymbol{m}} \mid \boldsymbol{m}, \mathbf{x}, \mathbf{z}\right)=\ell\left(\boldsymbol{\theta}_{\boldsymbol{m}} \mid \boldsymbol{m}, \mathbf{x}, \mathbf{z}\right)-\frac{\boldsymbol{\nu}_{\boldsymbol{m}}}{2} \log n$

$$
=\sum_{b=1}^{B} \ln p\left(\mathbf{z}_{b} \mid \boldsymbol{\pi}_{b}\right)-\frac{G_{b}-1}{2} \ln n+\sum_{j=1}^{d} \ln p\left(\mathbf{x}_{j} \mid \mathbf{z}_{\omega_{j}}, \boldsymbol{\alpha}_{j}\right)-\frac{\nu_{j} G_{\omega_{j}}}{2} \ln n,
$$

## Consequence

- For $\mathbf{z}_{b}$ fixed: possibility to re-affect each variable individually to the most accurate block.
- Thus computationnaly attractive


## Combinatorial model selection through a modified the EM algorithm for $B$ and $\left(G_{1}, \ldots, G_{B}\right)$ fixed : choice of $\boldsymbol{m} \Leftrightarrow$ choice of $\boldsymbol{\omega}$

The EM algorithm to achieve $\arg \max _{(\boldsymbol{\omega}, \boldsymbol{\theta})} \ell_{p e n}\left(\boldsymbol{\theta}_{\boldsymbol{m}} \mid \boldsymbol{m}, \mathbf{x}\right)$, starting from $\left(\boldsymbol{\omega}^{[0]}, \boldsymbol{\theta}^{[0]}\right)$ is at iteration $[r]$ :
E-step Computation of the fuzzy partitions $t_{i b g}^{[r]}:=\mathbb{E}\left[Z_{i b g} \mid \boldsymbol{x}_{i}, \boldsymbol{m}, \boldsymbol{\theta}^{[r-1]}\right]$, hence for $b=1, \ldots, B$, for $g=1, \ldots, G_{b}$, for $i=1, \ldots, n$

$$
t_{i b g}^{[r]}=\frac{\pi_{b g}^{[r-1]} \prod_{j \in \Omega_{b}^{[r-1]}} p\left(x_{i j} \mid \boldsymbol{\alpha}_{j g}^{[r-1]}\right)}{\sum_{k=1}^{G_{b}} \pi_{b k}^{[r-1]} \prod_{j \in \Omega_{b}^{[r-1]}} p\left(x_{i j} \mid \boldsymbol{\alpha}_{j g}^{[r-1]}\right)},
$$

M-step1 Updating the affectation of the variables to blocks

$$
\omega_{j}^{[r]}=\underset{\omega_{j} \in\{1, \ldots, B\}}{\arg \max }\left(\sum_{g=1}^{G_{\omega_{j}}} \max _{\boldsymbol{\alpha}_{j g} \in \Theta_{j}} Q\left(\boldsymbol{\alpha}_{j g} \mid \mathbf{x}_{j}, \mathbf{t}_{\omega_{j} g}^{[r]}\right)-\frac{\nu_{j} G_{\omega_{j}}}{2} \ln n\right)
$$

M-step2 Updating the model parameters

$$
\pi_{b g}^{[r]}=\frac{n_{b g}^{[r]}}{n} \text { and } \boldsymbol{\alpha}_{j g}^{[r]}=\underset{\boldsymbol{\alpha}_{j g} \in \Theta_{j}}{\arg \max } Q\left(\boldsymbol{\alpha}_{j g} \mid \mathbf{x}_{j}, \mathbf{t}_{\omega_{j}^{[r]} g}^{[r]}\right) .
$$

## Integrated complete-data likelihood

$$
p(\mathbf{x}, \mathbf{z} \mid \boldsymbol{m})=\int p(\mathbf{x}, \mathbf{z} \mid \boldsymbol{m}, \boldsymbol{\theta}) p(\boldsymbol{\theta} \mid \boldsymbol{m}) d \boldsymbol{\theta}
$$

## Assumptions

- Independence between the prior distributions
- Standard conjugate priors
- Closed form of the complete-data integrated likelihood


## MICL (maximum integrated complete-data likelihood) criterion

$$
\operatorname{MICL}(\boldsymbol{m})=\ln p\left(\mathbf{x}, \mathbf{z}_{\boldsymbol{m}}^{\star} \mid \boldsymbol{m}\right) \text { with } \mathbf{z}_{\boldsymbol{m}}^{\star}=\underset{\mathbf{z}_{m}}{\arg \max } \ln p(\mathbf{x}, \mathbf{z} \mid \boldsymbol{m}) .
$$

Thus

$$
\left(\boldsymbol{m}^{\star}, \mathbf{z}_{\boldsymbol{m}^{\star}}^{\star}\right)=\underset{\left(\boldsymbol{m}, \mathbf{z}_{m}\right)}{\arg \max } \ln p(\mathbf{x}, \mathbf{z} \mid \boldsymbol{m}) .
$$

## Motivations

- Criteria based on the integrated complete-data likelihood are popular for model-based clustering
- Take into account the clustering purpose : model the data distribution and provide well-separated components


## Optimisation of MICL over $(\mathbf{z}, \boldsymbol{\omega})$ for $B$ and $G_{1}, \ldots, G_{B}$ fixed

Starting at the initial value $\boldsymbol{\omega}^{[0]}$, each $\omega_{j}$ is uniformly sampled among $\{1, \ldots, B\}$, the algorithm at iteration $[r]$ is
Partition step: find $\mathbf{z}_{b}^{[r]}$ such that for all $b=1, \ldots, B$

$$
p\left(\mathbf{x}_{\{b\}}^{[r-1]}, \mathbf{z}_{b}^{[r]}\right) \geq p\left(\mathbf{x}_{\{b\}}^{[r-1]}, \mathbf{z}_{b}^{[r-1]}\right),
$$

where $\mathbf{x}_{\{b\}}^{[r-1]}=\left(\mathbf{x}_{j} ; \boldsymbol{\omega}^{[r-1]}=b\right)$.
Model step: find $\boldsymbol{\omega}^{[r]}$ such that for $j=1, \ldots, d$

$$
\omega_{j}^{[r]}=\underset{b \in\{1, \ldots, B\}}{\arg \max } p\left(\mathbf{x}_{j} \mid \mathbf{z}_{b}^{[r]}\right)
$$

## Data description

## NBA teams for the season 2016/2017 described by 16 numerical variables ${ }^{1}$

- total minutes played
$(\mathrm{min})$
- field goals made rate (fgmr)
- field goals attempted (fga)
- three-pointers made rate (3pmr)
- three-pointer attempted (3pa)
- free throws made (ftm)
- free throw attempted (fta)
- offensive rebounds (or)
- total rebounds (tr)
- assists (as)
- steals (st)
- turnovers (to)
- blocks (bk)
- personal fouls (pf)
- technical fouls (tc)
- points (pts)


## Model selection with BIC

| $B$ | BIC | Time (s) | Block | G | variables |
| :---: | :---: | :---: | :---: | :---: | :---: |
| 1 | -1932 | 7 | 1 | 2 | all the variables |
| 2 | -1915 | 47 | 1 | 3 | fgmr, fga, 3pmr, 3pa, tr, as, to, pts |
|  |  |  | 2 | 1 | min, ftmr, fta, orr, st, bk, pf, tc |
| 3 | -1909 | 170 | 1 | 2 | fgmr, 3pmr, pf |
|  |  |  | 2 | 2 | fga, 3pa, tr, as, st, to, pts |
|  |  |  | 3 | 1 | min, ftmr, fta, orr, bk, tc |

[^0]
## Data description

## NBA teams for the season 2016/2017 described by 16 numerical variables ${ }^{1}$

- total minutes played (min)
- field goals made rate (fgmr)
- field goals attempted (fga)
- three-pointers made rate (3pmr)
- three-pointer attempted (3pa)
- free throws made (ftm)
- free throw attempted (fta)
- offensive rebounds (or)
- total rebounds (tr)
- assists (as)
- steals (st)
- turnovers (to)
- blocks (bk)
- personal fouls (pf)
- technical fouls (tc)
- points (pts)

Model selection with BIC

| $B$ | BIC | Time (s) | Block | G | variables |
| :---: | :---: | :---: | :---: | :---: | :---: |
| 3 | -1909 | 170 | 1 | 2 | fgmr, 3pmr, pf |
|  |  |  | 2 | 2 | fga, 3pa, tr, as, st, to, pts |
|  |  | 3 | 1 | min, ftmr, fta, orr, bk, tc |  |

[^1]
## Model parameters

## First block: offensive vs defensive teams

Three features: field goals made rate, three points made rate and personal fouls

|  |  | fmgr | 3pmr | pf |
| :---: | :---: | :---: | :---: | :---: |
| offensive teams: ( $\pi_{11}=0.57$ ) | mean | 0.468 | 0.371 | 1628.042 |
| high shooting ability low personal fouls | sd | 0.009 | 0.010 | 76.138 |
| defensive teams: ( $\pi_{12}=0.43$ ) | mean | 0.446 | 0.342 | 1635.503 |
| low shooting ability | sd | 0.005 | 0.010 | 173.701 |

high personal fouls

Second block: two better statistics teams for general performances vs others
Seven features: field goal attempted, 3 points attempted, total rebounds, assists, steals, turnovers and points

|  |  | fga | 3pa | tr | as | st | to | pts |
| :--- | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| GS Warriors: $\left(\pi_{11}=0.07\right)$ | mean | 7144 | 2934 | 3642 | 2281 | 728 | 1186 | 9481 |
| Houston Rockets | sd | 3 | 372 | 3 | 211 | 59 | 3 | 22 |
| Other: $\left(\pi_{11}=0.93\right)$ | mean | 6992 | 2162 | 3564 | 1825 | 626 | 1091 | 8600 |
| teams | sd | 183 | 262 | 141 | 131 | 45 | 105 | 259 |

Third block: Six features detected as irrelevant for clustering.

## Data description

## Data ${ }^{1}$

- 27 chemical and physical properties of three types of Italian wines: Barolo, Grignolino, Barbera
- Data collected during the time period of 1970-1979


## Models selected by BIC

| $B$ | BIC | Time | Block | G | ARI |
| :---: | :---: | :---: | :---: | :---: | :---: |
| 1 | -6025.00 | 30 | 1 | 4 | 0.78 |
| 2 | -5947.88 | 280 | 1 | 3 | 0.87 |
|  |  |  | 2 | 4 | 0.16 |
| 3 | -5921.42 | 1590 | 1 | 4 | 0.74 |
|  |  |  | 2 | 4 | 0.20 |
|  |  |  | 3 | 2 | 0.02 |
| 4 | -5918.06 | 6065 | 1 | 4 | 0.75 |
|  |  |  | 2 | 2 | 0.21 |
|  |  |  | 3 | 3 | 0.02 |
|  |  |  | 4 | 2 | 0.00 |

${ }^{1}$ available in the package pgmm

## Model interpretation

## Block 1: the type of wines ( $\mathrm{ARI}=0.75$ )

19 variables: Alcohol, Sugar-free Extract, Tartaric Acid, Uronic Acids, Alcalinity of Ash, Calcium, Magnesium, Phosphate, Total Phenols, Flavanoids, Non-flavanoid Phenols, Proanthocyanins, Color Intensity, Hue, OD280/OD315 of Diluted Wines, OD280/OD315 of Flavanoids, Glycerol, 2-3-Butanediol, Proline

|  | Barolo | Grignolino | Barbera |
| :--- | ---: | ---: | ---: |
| Class 1 | 0 | 45 | 0 |
| Class 2 | 0 | 5 | 48 |
| Class 3 | 58 | 1 | 0 |
| Class 4 | 1 | 20 | 0 |

Block 2: year of production
4 variables: Fixed Acidity, Malic Acid, pH, Total Nitrogen

|  |  | Year |  |  |  |  |  |  |  |
| :--- | ---: | ---: | ---: | ---: | ---: | ---: | ---: | ---: | ---: |
|  | 1970 | 1971 | 1972 | 1973 | 1974 | 1975 | 1976 | 1978 | 1979 |
| Class 1 | 8 | 25 | 4 | 27 | 31 | 3 | 1 | 0 | 0 |
| Class 2 | 1 | 3 | 3 | 2 | 14 | 6 | 16 | 29 | 5 |

## Conclusion

- Proposition of model-based clustering with several class variables, each one explaining the heterogeneity of a block of variables:
- Find groups of variables producing the same clustering of the individuals - Interpret the clustering produced for each group of variables
- Model search performed simultaneously with parameters estimation
- Proposed model can be used in the heterogeneous data settings


## Perspectives

- Consider the semi-supervised setting in the multi-partition framework for new partitions discovery: i. e. $\mathbf{z}_{1}$ known and $\mathbf{z}_{2}$ unknown.
- Extension to heterogeneous co-clustering by adding one level of latent variable to summarize the multi-partition by a single partition, while keeping the partition of the variables.
- Derive some k-means type multi-partition similary to Witten \& Tibshirani (2010) in the variable selection framework to deal with the very high dimensional setting.


[^0]:    ${ }^{1}$ http://www.dougstats.com/16-17RD.Team. Opp.txt

[^1]:    ${ }^{1}$ http://www.dougstats.com/16-17RD.Team. Opp.txt

