On Mixtures of Factor Mixture Analyzers

Cinzia Viroli
cinzia.viroli@unibo.it

Department of Statistics,
University of Bologna, Italy

- In *model based clustering* the data are assumed to come from a *finite mixture model* (McLachlan and Peel, 2000) with each component corresponding to a cluster.

State of the art (1)

- In *model based clustering* the data are assumed to come from a *finite mixture model* (McLachlan and Peel, 2000) with each component corresponding to a cluster.

- For quantitative data each mixture component is usually modeled as a multivariate Gaussian distribution (Fraley and Raftery, 2002):

$$f(\mathbf{y};\boldsymbol{\theta}) = \sum_{i=1}^{k} w_i \phi^{(p)}(\mathbf{y};\boldsymbol{\mu}_i,\boldsymbol{\Sigma}_i)$$

State of the art (1)

- In *model based clustering* the data are assumed to come from a *finite mixture model* (McLachlan and Peel, 2000) with each component corresponding to a cluster.

- For quantitative data each mixture component is usually modeled as a multivariate Gaussian distribution (Fraley and Raftery, 2002):

$$f(\mathbf{y};\boldsymbol{\theta}) = \sum_{i=1}^{k} w_i \phi^{(p)}(\mathbf{y}; \boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)$$

- **However** when the number of observed variables is large, it is well known that Gaussian mixture models represent an over-parameterized solution.

State of the art (2)

Some solutions (among the others):

Model based clustering

Dimensionally reduced model based clustering

State of the art (2)

Some solutions (among the others):

Model based clustering

- Banfield and Raftery (1993): proposed a parameterization of the generic component-covariance matrix based on its spectral decomposition: $\mathbf{\Sigma}_i = \lambda_i \mathbf{A}_i^\top \mathbf{D}_i \mathbf{A}_i$
- Bouveyron et al. (2007): proposed a different parameterization of the generic component-covariance matrix

Dimensionally reduced model based clustering

State of the art (2)

Some solutions (among the others):

Model based clustering

- Banfield and Raftery (1993): proposed a parameterization of the generic component-covariance matrix based on its spectral decomposition: $\boldsymbol{\Sigma}_i = \lambda_i \mathbf{A}_i^\top \mathbf{D}_i \mathbf{A}_i$
- Bouveyron et al. (2007): proposed a different parameterization of the generic component-covariance matrix

Dimensionally reduced model based clustering

- Ghahrami and Hilton (1997) and McLachlan et al. (2003): *Mixtures of Factor Analyzers (MFA)*

State of the art (2)

Some solutions (among the others):

Model based clustering

- Banfield and Raftery (1993): proposed a parameterization of the generic component-covariance matrix based on its spectral decomposition: $\mathbf{\Sigma}_i = \lambda_i \mathbf{A}_i^\top \mathbf{D}_i \mathbf{A}_i$
- Bouveyron et al. (2007): proposed a different parameterization of the generic component-covariance matrix

Dimensionally reduced model based clustering

- Ghahrami and Hilton (1997) and McLachlan et al. (2003): *Mixtures of Factor Analyzers (MFA)*
- Yoshida et al. (2004), Baek and McLachlan (2008), Montanari and Viroli (2010) : *Factor Mixture Analysis (FMA)*

Mixture of factor analyzers (MFA)

- **Dimensionality reduction is performed through k factor models with Gaussian factors**

Mixture of factor analyzers (MFA)

- **Dimensionality reduction is performed through k factor models with Gaussian factors**
- The distribution of each observation is modelled, with probability π_j ($j = 1, \ldots, k$), according to an ordinary factor analysis model $\mathbf{y} = \boldsymbol{\eta}_j + \boldsymbol{\Lambda}_j \mathbf{z} + \mathbf{e}_j$, with $\mathbf{e}_j \sim \phi^{(p)}(\mathbf{0}, \boldsymbol{\Psi}_j)$, where $\boldsymbol{\Psi}_j$ is a diagonal matrix and $\mathbf{z}_j \sim \phi^{(q)}(\mathbf{0}, \mathbf{I}_q)$

Mixture of factor analyzers (MFA)

- **Dimensionality reduction is performed through k factor models with Gaussian factors**
- The distribution of each observation is modelled, with probability π_j ($j = 1, \ldots, k$), according to an ordinary factor analysis model $\mathbf{y} = \boldsymbol{\eta}_j + \boldsymbol{\Lambda}_j \mathbf{z} + \mathbf{e}_j$, with $\mathbf{e}_j \sim \phi^{(p)}(\mathbf{0}, \boldsymbol{\Psi}_j)$, where $\boldsymbol{\Psi}_j$ is a diagonal matrix and $\mathbf{z}_j \sim \phi^{(q)}(\mathbf{0}, \mathbf{I}_q)$
- In the observed space we obtain a finite mixture of multivariate Gaussians with heteroscedastic components:

$$f(\mathbf{y}) = \sum_{j=1}^{k} \pi_j \phi^{(p)}(\boldsymbol{\eta}_j, \boldsymbol{\Lambda}_j \boldsymbol{\Lambda}_j^\top + \boldsymbol{\Psi}_j)$$

Factor Mixture Analysis (FMA)

- **Dimensionality reduction is performed through a single factor model with factors modelled by a multivariate Gaussian mixture**

Factor Mixture Analysis (FMA)

- **Dimensionality reduction is performed through a single factor model with factors modelled by a multivariate Gaussian mixture**
- The observed centred data are described as $\mathbf{y} = \mathbf{\Lambda z} + \mathbf{e}$ with $\mathbf{e} \sim \phi^{(p)}(\mathbf{0}, \mathbf{\Psi})$ where $\mathbf{\Psi}$ is diagonal.

Factor Mixture Analysis (FMA)

- **Dimensionality reduction is performed through a single factor model with factors modelled by a multivariate Gaussian mixture**
- The observed centred data are described as $\mathbf{y} = \mathbf{\Lambda}\mathbf{z} + \mathbf{e}$ with $\mathbf{e} \sim \phi^{(p)}(\mathbf{0}, \mathbf{\Psi})$ where $\mathbf{\Psi}$ is diagonal.
- The q factors are assumed to be standardized and are modelled as a finite mixture of multivariate Gaussians

$$f(\mathbf{z}) = \sum_{i=1}^{k} \gamma_i \phi_i^{(q)}(\boldsymbol{\mu}_i, \mathbf{\Sigma}_i).$$

Factor Mixture Analysis (FMA)

- **Dimensionality reduction is performed through a single factor model with factors modelled by a multivariate Gaussian mixture**
- The observed centred data are described as $\mathbf{y} = \boldsymbol{\Lambda}\mathbf{z} + \mathbf{e}$ with $\mathbf{e} \sim \phi^{(p)}(\mathbf{0}, \boldsymbol{\Psi})$ where $\boldsymbol{\Psi}$ is diagonal.
- The q factors are assumed to be standardized and are modelled as a finite mixture of multivariate Gaussians

$$f(\mathbf{z}) = \sum_{i=1}^{k} \gamma_i \phi_i^{(q)}(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i).$$

- In the observed space we obtain a finite mixture of multivariate Gaussians with heteroscedastic components:

$$f(\mathbf{y}) = \sum_{i=1}^{k} \gamma_i \phi_i^{(p)}(\boldsymbol{\Lambda}\boldsymbol{\mu}_i, \boldsymbol{\Lambda}\boldsymbol{\Sigma}_i\boldsymbol{\Lambda}^\top + \boldsymbol{\Psi}).$$

MFA vs FMA

MFA

- k factor models with q Gaussian factors;

FMA

- one factor model with q non Gaussian factors (distributed as a multivariate mixture of Gaussians);

MFA vs FMA

MFA

- k factor models with q Gaussian factors;

- The number of clusters corresponds to the number of factor models; $\Rightarrow$ 'local' dimension reduction within each group

FMA

- one factor model with q non Gaussian factors (distributed as a multivariate mixture of Gaussians);

- The number of clusters is defined by the number of components of the Gaussian mixture; $\Rightarrow$ 'global' dimension reduction and clustering is performed in the latent space.

MFA vs FMA

MFA

- k factor models with q Gaussian factors;
- The number of clusters corresponds to the number of factor models; $\Rightarrow$ 'local' dimension reduction within each group
- A flexible solution with less parameters than model based clustering;

FMA

- one factor model with q non Gaussian factors (distributed as a multivariate mixture of Gaussians);
- The number of clusters is defined by the number of components of the Gaussian mixture; $\Rightarrow$ 'global' dimension reduction and clustering is performed in the latent space.
- A flexible solution with less parameters than model based clustering;

Mixtures of Factor Mixture Analyzers

The model

We assume the data can be described by k_1 factor models with probability π_j $(j = 1, \ldots, k_1)$:

$$\mathbf{y} = \boldsymbol{\eta}_j + \boldsymbol{\Lambda}_j \mathbf{z} + \mathbf{e}_j. \tag{1}$$

The model

We assume the data can be described by k_1 factor models with probability π_j $(j = 1, \ldots, k_1)$:

$$\mathbf{y} = \boldsymbol{\eta}_j + \boldsymbol{\Lambda}_j \mathbf{z} + \mathbf{e}_j. \tag{1}$$

Within all the factor models, the factors are assumed to be distributed according to a finite mixture of k_2 Gaussians:

$$f(\mathbf{z}) = \sum_{i=1}^{k_2} \gamma_i \phi^{(q)}(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i), \tag{2}$$

with mixture parameters supposed to be equal across the factor models $j = 1, \ldots, k_1$.

The model

From the previous assumptions it follows that the distribution of the observed variables becomes a '*double*' mixture of Gaussians:

$$f(\mathbf{y};\boldsymbol{\theta}) = \sum_{j=1}^{k_1} \pi_j \sum_{i=1}^{k_2} \gamma_i \phi^{(p)}(\boldsymbol{\eta}_j + \boldsymbol{\Lambda}_j \boldsymbol{\mu}_i, \boldsymbol{\Lambda}_j \boldsymbol{\Sigma}_i \boldsymbol{\Lambda}_j^\top + \boldsymbol{\Psi}_j). \quad (3)$$

which leads to a '*double*' interpretation:

(1) a mixture of k_1 factor analyzers with non-Gaussian factors, jointly modelled by a mixture of k_2 Gaussians, or
(2) a non-linear factor mixture analysis model.

From the previous assumptions it follows that the distribution of the observed variables becomes a '*double*' mixture of Gaussians:

$$f(\mathbf{y};\boldsymbol{\theta}) = \sum_{j=1}^{k_1} \pi_j \sum_{i=1}^{k_2} \gamma_i \phi^{(p)}(\boldsymbol{\eta}_j + \boldsymbol{\Lambda}_j \boldsymbol{\mu}_i, \boldsymbol{\Lambda}_j \boldsymbol{\Sigma}_i \boldsymbol{\Lambda}_j^\top + \boldsymbol{\Psi}_j). \quad (3)$$

which leads to a '*double*' interpretation:

(1) a mixture of k_1 factor analyzers with non-Gaussian factors, jointly modelled by a mixture of k_2 Gaussians, or
(2) a non-linear factor mixture analysis model.

Moreover it coincides with MFA when $k_2 = 1$ and with FMA when $k_1 = 1$. Thus the method includes MFA and FMA as special cases.

Classification of units

- The double mixture model implies that observations can be classified according to a two-level process:

Classification of units

- The double mixture model implies that observations can be classified according to a two-level process:

  (1) units may be described by one out of the k_1 different factor models;

Classification of units

- The double mixture model implies that observations can be classified according to a two-level process:

  (1) units may be described by one out of the k_1 different factor models;
  (2) then units (within each factor model) may belong to different k_2 sub-populations (defined by the k_2 components of the multivariate factor distribution.)

Classification of units

- The double mixture model implies that observations can be classified according to a two-level process:

  (1) units may be described by one out of the k_1 different factor models;
  (2) then units (within each factor model) may belong to different k_2 sub-populations (defined by the k_2 components of the multivariate factor distribution.)

- The question is: k_1, k_2 or $k_1 \times k_2$ groups?
  i.e. k_1 or k_2 non-Gaussian sub-populations or $k_1 \times k_2$ Gaussian ones?

UCI Wisconsin Diagnostic Breast Cancer Data

The data set contains 569 clinical cases of benignant (62.7%) and malignant (37.3%) diagnoses of breast cancer. Cluster analysis is based on $p = 3$ attributes: extreme area, extreme smoothness, and mean texture.
(ARI by Mclust, k=4 groups: 0.55)

	MFA	FMA	MFMA
k_1	2	1	2
k_2	1	3	3
q	1	1	1
h	16	12	22
logL	-2174	-2167	-2139
BIC	4449	4410	4418
AIC	4379	4385	4323
$\text{ARI}_{(k_1)}$	0.73	0.00	0.80
$\text{ARI}_{(k_2)}$	0.00	0.64	0.05
$\text{ARI}_{(k_1 k_2)}$	0.73	0.64	0.52

UCI Wisconsin Diagnostic Breast Cancer Data

The data set contains 569 clinical cases of benignant (62.7%) and malignant (37.3%) diagnoses of breast cancer. Cluster analysis is based on $p = 3$ attributes: extreme area, extreme smoothness, and mean texture.
(ARI by Mclust, k=4 groups: 0.55)

	MFA	FMA	MFMA
k_1	2	1	2
k_2	1	3	3
q	1	1	1
h	16	12	22
logL	-2174	-2167	-2139
BIC	4449	4410	4418
AIC	4379	4385	4323
$\text{ARI}_{(k_1)}$	0.73	0.00	0.80
$\text{ARI}_{(k_2)}$	0.00	0.64	0.05
$\text{ARI}_{(k_1 k_2)}$	0.73	0.64	0.52

MFMA: 2, 3 or 6 groups?

UCI Wisconsin Diagnostic Breast Cancer Data

Some indicators to measure the separation of the estimated clusters have been computed:

	k	avg. dist. between	avg. dist. within	avg. silhouette width
$MFMA_{(k_1)}$	2	2.71	1.77	0.32
$MFMA_{(k_2)}$	3	2.67	1.88	0.15
$MFMA_{(k_1 k_2)}$	6	2.57	1.47	0.19
MFA	2	2.68	1.73	0.32
FMA	3	2.72	1.76	0.26
$MCLUST$	4	2.60	1.41	0.27

UCI Wisconsin Diagnostic Breast Cancer Data

Some indicators to measure the separation of the estimated clusters have been computed:

	k	avg. dist. between	avg. dist. within	avg. silhouette width
$MFMA_{(k_1)}$	2	2.71	1.77	0.32
$MFMA_{(k_2)}$	3	2.67	1.88	0.15
$MFMA_{(k_1 k_2)}$	6	2.57	1.47	0.19
MFA	2	2.68	1.73	0.32
FMA	3	2.72	1.76	0.26
$MCLUST$	4	2.60	1.41	0.27

$k_1 = 2$ factor models with $k_2 = 3$ components for modeling the factor

... a mixture of factor analyzers with non-Gaussian components

Conclusion

- MFMA is a double mixture model which extends and combines MFA and FMA

Conclusion

- MFMA is a double mixture model which extends and combines MFA and FMA
- A MFMA model with k_1 and k_2 components may be interpreted in three different ways:
  - as a double mixture which performs clustering into $k = k_1 \times k_2$ groups,
  - as a mixture of factor mixture analysis models which performs clustering into $k = k_2$ groups
  - or as a mixture of factor analyzers with non-Gaussian components which classifies units into $k = k_1$ groups.

Conclusion

- MFMA is a double mixture model which extends and combines MFA and FMA
- A MFMA model with k_1 and k_2 components may be interpreted in three different ways:
  - as a double mixture which performs clustering into $k = k_1 \times k_2$ groups,
  - as a mixture of factor mixture analysis models which performs clustering into $k = k_2$ groups
  - or as a mixture of factor analyzers with non-Gaussian components which classifies units into $k = k_1$ groups.
- In the last two perspectives the proposed model represents a powerful tool for modelling non-Gaussian latent variables.

Conclusion

- MFMA is a double mixture model which extends and combines MFA and FMA
- A MFMA model with k_1 and k_2 components may be interpreted in three different ways:
  - as a double mixture which performs clustering into $k = k_1 \times k_2$ groups,
  - as a mixture of factor mixture analysis models which performs clustering into $k = k_2$ groups
  - or as a mixture of factor analyzers with non-Gaussian components which classifies units into $k = k_1$ groups.
- In the last two perspectives the proposed model represents a powerful tool for modelling non-Gaussian latent variables.
- Some references:

  - A. Montanari and C. Viroli (2010), *Heteroscedastic Factor Mixture Analysis*, Statistical Modelling, forthcoming
  - C. Viroli (2011), *Dimensionally reduced model-based clustering through Mixtures of Factor Mixture Analyzers*, Journal of Classification, forthcoming