Model-Based Clustering II

Prof. Sharon McNicholas

STATS 780/CSE 780

Introduction

- We have seen model-based clustering using the GPCM models.
- Today, we look at mixture model-based clustering for higher dimensional data.
- As in the last lecture, some of the material is taken from McNicholas (2016).
- Note that bibliographic references are given at the end of these slides.

Factor Analysis

- Consider independent p-dimensional random variables $\mathbf{X}_1, \ldots, \mathbf{X}_n$. The factor analysis model can be written

$$\mathbf{X}_i = \boldsymbol{\mu} + \boldsymbol{\Lambda}\mathbf{U}_i + \boldsymbol{\epsilon}_i, \tag{1}$$

  for $i = 1, \ldots, n$, where $\boldsymbol{\Lambda}$ is a $p \times q$ matrix of factor loadings, the latent factor $\mathbf{U}_i \sim \mathsf{N}(\mathbf{0}, \mathbf{I}_q)$, and $\boldsymbol{\epsilon}_i \sim \mathsf{N}(\mathbf{0}, \boldsymbol{\Psi})$, where $\boldsymbol{\Psi} = \operatorname{diag}(\psi_1, \psi_2, \ldots, \psi_p)$.

- Note that the $\mathbf{U}_i$ are independently distributed and independent of the $\boldsymbol{\epsilon}_i$, which are also independently distributed.

- Factor analysis is a data reduction technique, i.e., $q < p$.

STATS 780/CSE 780 Prof. Sharon McNicholas

Factor Analysis

- From (1), it follows that the marginal distribution of $\mathbf{X}_i$ under the factor analysis model is $\mathsf{N}(\boldsymbol{\mu}, \boldsymbol{\Lambda}\boldsymbol{\Lambda}' + \boldsymbol{\Psi})$. There are $pq + p - q(q-1)/2$ free parameters in the covariance matrix $\boldsymbol{\Lambda}\boldsymbol{\Lambda}' + \boldsymbol{\Psi}$ (Lawley and Maxwell, 1962).

- Therefore, the reduction in free covariance parameters under the factor analysis model is

$$\frac{1}{2}p(p+1) - \left[pq + p - \frac{1}{2}q(q-1)\right] = \frac{1}{2}\left[(p-q)^2 - (p+q)\right], \tag{2}$$

  and there is a reduction in the number of free parameters provided that (2) is positive, i.e., provided that

$$(p-q)^2 > (p+q).$$

Mixture of Factor Analyzers

- Analogous to the factor analysis model, the mixture of factor analyzers model assumes that

$$\mathbf{X}_i = \boldsymbol{\mu}_g + \boldsymbol{\Lambda}_g \mathbf{U}_{ig} + \boldsymbol{\epsilon}_{ig} \tag{3}$$

  with probability π_g, for $i = 1, \ldots, n$ and $g = 1, \ldots, G$.

- $\boldsymbol{\Lambda}_g$ is a $p \times q$ matrix of factor loadings, the $\mathbf{U}_{ig}$ are independently $\text{N}(\mathbf{0}, \mathbf{I}_q)$ and are independent of the $\boldsymbol{\epsilon}_{ig}$, which are independently $\text{N}(\mathbf{0}, \boldsymbol{\Psi}_g)$, where $\boldsymbol{\Psi}_g$ is a $p \times p$ diagonal matrix with positive diagonal elements.

- It follows that the density of $\mathbf{X}_i$ from the mixture of factor analyzers model is

$$f(\mathbf{x}_i \mid \boldsymbol{\vartheta}) = \sum_{g=1}^{G} \pi_g \phi(\mathbf{x}_i \mid \boldsymbol{\mu}_g, \boldsymbol{\Lambda}_g \boldsymbol{\Lambda}_g' + \boldsymbol{\Psi}_g). \tag{4}$$

Mixture of Factor Analyzers

- Ghahramani and Hinton (1997) were the first to introduce a mixture of factor analyzers model; they constrain $\boldsymbol{\Psi}_g = \boldsymbol{\Psi}$ to facilitate an interpretation of $\boldsymbol{\Psi}$ as sensor noise.

- Tipping and Bishop (1999) introduce the closely related mixture of probabilistic principal component analyzers (MPPCA) model, where each $\boldsymbol{\Psi}_g$ matrix is isotropic, i.e., $\boldsymbol{\Psi}_g = \psi_g \mathbf{I}_p$.

- McLachlan and Peel (2000) use the unconstrained mixture of factor analyzers model, i.e., with $\boldsymbol{\Sigma}_g = \boldsymbol{\Lambda}_g \boldsymbol{\Lambda}_g' + \boldsymbol{\Psi}_g$.

PGMMs

- One can view the mixture of factor analyzers models and the MPPCA model, collectively, as a family of three models.

- This family can easily be extended to a four-member family by adding the model with component covariance $\mathbf{\Sigma}_g = \mathbf{\Lambda}_g\mathbf{\Lambda}_g' + \psi\mathbf{I}_p$.

- Members of this family of four models have between $G[pq - q(q-1)/2] + 1$ and $G[pq - q(q-1)/2] + Gp$ free parameters in the component covariance matrices.

PGMMs

- A greater level of parsimony can be introduced by constraining the component factor loading matrices to be equal, i.e., $\mathbf{\Lambda}_g = \mathbf{\Lambda}$.
- McNicholas and Murphy (2008) develop a family of eight parsimonious Gaussian mixture models (PGMMs) for clustering by imposing, or not, each of the constraints $\mathbf{\Lambda}_g = \mathbf{\Lambda}$, $\mathbf{\Psi}_g = \mathbf{\Psi}$, and $\mathbf{\Psi}_g = \psi_g\mathbf{I}_p$.
- Members of the PGMM family have between $pq - q(q-1)/2 + 1$ and $G[pq - q(q-1)/2] + Gp$ free parameters in the component covariance matrices.
- McNicholas (2010) used the PGMM family for model-based classification, and Andrews and McNicholas (2011) applied it for model-based discriminant analysis.

PGMMs

$\mathbf{\Lambda}_g = \mathbf{\Lambda}$	$\mathbf{\Psi}_g = \mathbf{\Psi}$	$\mathbf{\Psi}_g = \psi_g \mathbf{I}_p$	$\mathbf{\Sigma}_g$	Free Cov. Paras.
C	C	C	$\mathbf{\Lambda\Lambda}' + \psi \mathbf{I}_p$	$pq - q(q-1)/2 + 1$
C	C	U	$\mathbf{\Lambda\Lambda}' + \mathbf{\Psi}$	$pq - q(q-1)/2 + p$
C	U	C	$\mathbf{\Lambda\Lambda}' + \psi_g \mathbf{I}_p$	$pq - q(q-1)/2 + G$
C	U	U	$\mathbf{\Lambda\Lambda}' + \mathbf{\Psi}_g$	$pq - q(q-1)/2 + Gp$
U	C	C	$\mathbf{\Lambda}_g\mathbf{\Lambda}_g' + \psi \mathbf{I}_p$	$G[pq - q(q-1)/2] + 1$
U	C	U	$\mathbf{\Lambda}_g\mathbf{\Lambda}_g' + \mathbf{\Psi}$	$G[pq - q(q-1)/2] + p$
U	U	C	$\mathbf{\Lambda}_g\mathbf{\Lambda}_g' + \psi_g \mathbf{I}_p$	$G[pq - q(q-1)/2] + G$
U	U	U	$\mathbf{\Lambda}_g\mathbf{\Lambda}_g' + \mathbf{\Psi}_g$	$G[pq - q(q-1)/2] + Gp$

Expanded PGMMs

- McNicholas and Murphy (2010) further parameterize the mixture of factor analyzers component covariance structure by writing

$$\mathbf{\Psi}_g = \omega_g \mathbf{\Delta}_g,$$

where $\omega_g \in \mathbb{R}^+$ and $\mathbf{\Delta}_g$ is a diagonal matrix with $|\mathbf{\Delta}_g| = 1$.

- The resulting mixture of modified factor analyzers model has component covariance structure

$$\mathbf{\Sigma}_g = \mathbf{\Lambda}_g\mathbf{\Lambda}_g' + \omega_g \mathbf{\Delta}_g.$$

- In addition to the constraint $\mathbf{\Lambda}_g = \mathbf{\Lambda}$, all legitimate combinations of the constraints $\omega_g = \omega$, $\mathbf{\Delta}_g = \mathbf{\Delta}$, and $\mathbf{\Delta}_g = \mathbf{I}_p$ are imposed, resulting in a family of 12 PGMMs.

Expanded PGMMs

Expanded PGMM Nomenclature					
$\mathbf{\Lambda}_g = \mathbf{\Lambda}$	$\mathbf{\Delta}_g = \mathbf{\Delta}$	$\omega_g = \omega$	$\mathbf{\Delta}_g = \mathbf{I}_p$	PGMM Equiv.	$\mathbf{\Sigma}_g$
C	C	C	C	CCC	$\mathbf{\Lambda}\mathbf{\Lambda}' + \omega\mathbf{I}_p$
C	C	U	C	CUC	$\mathbf{\Lambda}\mathbf{\Lambda}' + \omega_g\mathbf{I}_p$
U	C	C	C	UCC	$\mathbf{\Lambda}_g\mathbf{\Lambda}_g' + \omega\mathbf{I}_p$
U	C	U	C	UUC	$\mathbf{\Lambda}_g\mathbf{\Lambda}_g' + \omega_g\mathbf{I}_p$
C	C	C	U	CCU	$\mathbf{\Lambda}\mathbf{\Lambda}' + \omega\mathbf{\Delta}$
C	C	U	U	–	$\mathbf{\Lambda}\mathbf{\Lambda}' + \omega_g\mathbf{\Delta}$
U	C	C	U	UCU	$\mathbf{\Lambda}_g\mathbf{\Lambda}_g' + \omega\mathbf{\Delta}$
U	C	U	U	–	$\mathbf{\Lambda}_g\mathbf{\Lambda}_g' + \omega_g\mathbf{\Delta}$
C	U	C	U	–	$\mathbf{\Lambda}\mathbf{\Lambda}' + \omega\mathbf{\Delta}_g$
C	U	U	U	CUU	$\mathbf{\Lambda}\mathbf{\Lambda}' + \omega_g\mathbf{\Delta}_g$
U	U	C	U	–	$\mathbf{\Lambda}_g\mathbf{\Lambda}_g' + \omega\mathbf{\Delta}_g$
U	U	U	U	UUU	$\mathbf{\Lambda}_g\mathbf{\Lambda}_g' + \omega_g\mathbf{\Delta}_g$

Comments

- Note that an AECM algorithm (Meng and van Dyk, 1997) is used for parameter estimation.
- I want to look at quite a few clustering examples now.
- We can use the `pgmm` package in R.
- We can also compare with other clustering approaches.
- Then, on to clustering longitudinal data.

References

- Andrews, J.L. and P.D. McNicholas (2011). 'Mixtures of modified t-factor analyzers for model-based clustering, classification, and discriminant analysis'. *Journal of Statistical Planning and Inference* **141**(4), 1479–1486.
- Ghahramani, Z. and G.E. Hinton (1997). The EM algorithm for factor analyzers. Technical Report CRG-TR-96-1, University of Toronto, Toronto, Canada.
- Lawley, D.N. and A.E. Maxwell (1962). 'Factor analysis as a statistical method'. *Journal of the Royal Statistical Society: Series D* **12**(3), 209–229.
- McLachlan, G. J. and D. Peel (2000). Mixtures of factor analyzers. In *Proceedings of the Seventh International Conference on Machine Learning*, pp. 599–606. San Francisco: Morgan Kaufmann.
- McNicholas, P.D. (2010). 'Model-based classification using latent Gaussian mixture models'. *Journal of Statistical Planning and Inference* **140**(5), 1175–1181.
- McNicholas, P.D. (2016), *Mixture Model-Based Classification*, Boca Raton: Chapman & Hall/CRC Press.
- McNicholas, P.D. and T.B. Murphy (2008). 'Parsimonious Gaussian mixture models'. *Statistics and Computing* **18**(3), 285–296.
- McNicholas, P.D. and T.B. Murphy (2010). 'Model-based clustering of microarray expression data via latent Gaussian mixture models'. *Bioinformatics* **26**(21), 2705–2712.
- Meng, X.-L. and D. van Dyk (1997). 'The EM algorithm — An old folk song sung to a fast new tune (with discussion)'. *Journal of the Royal Statistical Society: Series B* **59**(3), 511–567.