Model-Based Clustering III

Prof. Sharon McNicholas

STATS 780/CSE 780

Introduction

- Now, we are going to look at clustering longitudinal data in a mixture framework.
- As in the last couple of lectures, some of the material is taken from McNicholas (2016).
- Bibliographic references are given at the end of these slides.

Cholesky Decomposition

- The Cholesky decomposition is a method for decomposing a matrix into the product of a lower triangular matrix and its transpose.

- Let $\mathbf{A}$ be a real, positive definite matrix; then the Cholesky decomposition of $\mathbf{A}$ is
$$\mathbf{A} = \mathbf{L}\mathbf{L}', \tag{1}$$
where $\mathbf{L}$ is a unique lower triangular matrix.

- The decomposition in (1) is often used in numerical analysis applications to simplify the solution to a system of linear equations.

Modified Cholesky Decomposition

- Pourahmadi (1999) applies a modified Cholesky decomposition to the covariance matrix $\mathbf{\Sigma}$ of a random variable, i.e.,
$$\mathbf{T}\mathbf{\Sigma}\mathbf{T}' = \mathbf{D}, \tag{2}$$
where $\mathbf{T}$ is a unique unit lower triangular matrix and $\mathbf{D}$ is a unique diagonal matrix with strictly positive diagonal entries.

- Note that a unit lower triangular matrix is a lower triangular matrix with diagonal elements 1.

- Alternatively, the relationship in (2) can be written
$$\mathbf{\Sigma}^{-1} = \mathbf{T}'\mathbf{D}^{-1}\mathbf{T}. \tag{3}$$

Modified Cholesky Decomp. contd.

- The values of $\mathbf{T}$ and $\mathbf{D}$ can be interpreted as generalized autoregressive parameters and innovation variances, respectively (Pourahmadi, 1999), so that the linear least-squares predictor of X_t, based on $X_{t-1}, \ldots, X_1$, is given by

$$\hat{X}_t = \mu_t + \sum_{s=1}^{t-1}(-\varphi_{ts})(X_s - \mu_s) + \sqrt{d_t}\varepsilon_t, \tag{4}$$

  where φ_{ts} is the sub-diagonal element of $\mathbf{T}$ in position (t, s), d_t is the tth diagonal element of $\mathbf{D}$, and $\varepsilon_t \sim \mathsf{N}(0, 1)$.

- To cluster longitudinal data, McNicholas and Murphy (2010) use (3) as the component covariance in a Gaussian mixture, i.e.,

$$\boldsymbol{\Sigma}_g^{-1} = \mathbf{T}_g'\mathbf{D}_g^{-1}\mathbf{T}_g.$$

STATS 780/CSE 780 Prof. Sharon McNicholas

A Family

- A family of eight Gaussian mixture models arises from the option to constrain $\mathbf{T}_g$ and/or $\mathbf{D}_g$ to be equal across components together with the option to impose the isotropic constraint $\mathbf{D}_g = \delta_g\mathbf{I}_p$ (McNicholas and Murphy, 2010).

- Each member of this family has a natural interpretation for longitudinal data.
  - Constraining $\mathbf{T}_g = \mathbf{T}$ suggests that the autoregressive relationship between time points, cf. (4), is the same across components.
  - The constraint $\mathbf{D}_g = \mathbf{D}$ suggests that the variability at each time point is the same for each component.
  - The isotropic constraint $\mathbf{D}_g = \delta_g\mathbf{I}_p$ suggests that the variability is the same at each time point in component g.

A Family contd.

Model	$\mathbf{T}_g$	$\mathbf{D}_g$	$\mathbf{D}_g$	Free Cov. Parameters
EEA	Equal	Equal	Anisotropic	$p(p-1)/2+p$
VVA	Variable	Variable	Anisotropic	$G[p(p-1)/2]+Gp$
VEA	Variable	Equal	Anisotropic	$G[p(p-1)/2]+p$
EVA	Equal	Variable	Anisotropic	$p(p-1)/2+Gp$
VVI	Variable	Variable	Isotropic	$G[p(p-1)/2]+G$
VEI	Variable	Equal	Isotropic	$G[p(p-1)/2]+1$
EVI	Equal	Variable	Isotropic	$p(p-1)/2+G$
EEI	Equal	Equal	Isotropic	$p(p-1)/2+1$

Comments

- Note that an EM algorithm can be used for parameter estimation.
- I want to look at some clustering examples now.
- We can use the `longclust` package in R.
- Here, it is not so easy to compare with other approaches.
- Among other things, I hope you are getting a sense of the flexibility of mixture model-based approaches.
- In the next lecture, we will finish up with model-based approaches and move on.

References

- Benoît (1924). Note sur une méthode de résolution des equations normales provenant de l'application de la méthode des moindres carrés à un système d'équations linéaires en nombre inférieur celui des inconnues (Procédé du Commandant Cholesky). *Bulletin Géodésique* **2**, 67–77.

- McNicholas, P.D. (2016), *Mixture Model-Based Classification*, Boca Raton: Chapman & Hall/CRC Press.

- McNicholas, P.D. and T.B. Murphy (2010). 'Model-based clustering of longitudinal data'. *The Canadian Journal of Statistics* **38**(1), 153–168.

- Pourahmadi, M. (1999). 'Joint mean-covariance models with applications to longitudinal data: Unconstrained parameterisation'. *Biometrika* **86**(3), 677–690.