ROBUST
13. září 2016

Shluková analýza ve směsích regresních modelů

Kristýna Vaňkátová, Eva Fišerová

Department of Mathematical Analysis and Applications of Mathematics
Faculty of Science, Palacký University Olomouc
Czech Republic

kristyna.vankatova@upol.cz

Standard mixture regression models

- Standard linear regression model

$$y_j = \mathbf{x}_j^{\mathrm{T}} \boldsymbol{\beta} + \varepsilon_j$$

- The mixture of linear regression models can be written as

$$y_j = \begin{cases} \mathbf{x}_j^{\mathrm{T}} \boldsymbol{\beta}_1 + \varepsilon_{1j} & \text{with probability } \pi_1 \\ \mathbf{x}_j^{\mathrm{T}} \boldsymbol{\beta}_2 + \varepsilon_{2j} & \text{with probability } \pi_2 \\ & \vdots \\ \mathbf{x}_j^{\mathrm{T}} \boldsymbol{\beta}_c + \varepsilon_{cj} & \text{with probability } \pi_c \end{cases} \qquad \text{where } \varepsilon_{ij} \sim \mathrm{N}(0, \sigma_i^2)$$

- Accounting for the mixture structure, the conditional density of $y_j | \mathbf{x}_j$ is

$$f(y_j | \mathbf{x}_j, \boldsymbol{\Psi}) = \sum_{i=1}^{c} \pi_i \phi(y_j | \mathbf{x}_j^{\mathrm{T}} \boldsymbol{\beta}_i, \sigma_i^2).$$

- $\boldsymbol{\Psi}$ denotes the vector of all parameters

$$\boldsymbol{\Psi} = (\pi_1, \ldots, \pi_c, (\boldsymbol{\beta}_1^{\mathrm{T}}, \sigma_1^2), \ldots, (\boldsymbol{\beta}_c^{\mathrm{T}}, \sigma_c^2))^{\mathrm{T}}.$$

- Maximizing the log likelihood function

$$\log L(\boldsymbol{\Psi}|\mathbf{x}_1,\ldots,\mathbf{x}_n,y_1,\ldots,y_n)=\sum_{j=1}^{n}\log\left(\sum_{i=1}^{c}\pi_i\phi(y_j|\mathbf{x}_j^{\mathrm{T}}\boldsymbol{\beta}_i,\sigma_i^2)\right).$$

- The **EM algorithm** is an iterative procedure which alternates between an Expectation step and a Maximization step.

$(k+1)$th iteration of EM algorithm:

1. **E-step**: Given the observed data $\mathbf{y}$ and current parameter estimates $\widehat{\boldsymbol{\Psi}}^{(k)}$ in the kth iteration, calculates estimated a-posteriori probabilities, assigning each observation j to each cluster i.

$$\widehat{\tau}_{ij}^{(k)}=\frac{\widehat{\pi}_i^{(k)}\phi(y_j|\mathbf{x}_j^{\mathrm{T}}\widehat{\boldsymbol{\beta}}_i^{(k)},\widehat{\sigma}_i^{2(k)})}{\sum_{h=1}^{c}\widehat{\pi}_h^{(k)}\phi(y_j|\mathbf{x}_j^{\mathrm{T}}\widehat{\boldsymbol{\beta}}_h^{(k)},\widehat{\sigma}_h^{2(k)})}.$$

2. **M-step**: Given the estimates $\widehat{\tau}_{ij}^{(k)}$ for the a-posteriori probabilities τ_{ij} (which are functions of $\widehat{\boldsymbol{\Psi}}^{(k)}$), obtain new estimates $\widehat{\boldsymbol{\Psi}}^{(k+1)}$ of the parameters by maximizing expected complete log–likelihood, which is in general easier to maximize than the original likelihood.

$$Q(\boldsymbol{\Psi},\widehat{\boldsymbol{\Psi}}^{(k)})=\sum_{j=1}^{n}\sum_{i=1}^{c}\widehat{\tau}_{ij}^{(k)}\log\phi(y_j|\mathbf{x}_j^{\mathrm{T}}\boldsymbol{\beta}_i,\sigma_i^2)$$

- The maximization is equivalent to solving the weighted least squares problem:

$$\widehat{\boldsymbol{\beta}}_i^{(k+1)} = (\mathbf{X}^{\mathrm{T}}\mathbf{W}_i^{(k)}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{W}_i^{(k)}\mathbf{y} \qquad \text{for} \quad i = 1, \ldots, c,$$

  where $\mathbf{y} = (y_1, \ldots, y_n)^T$ is the vector of observations and $\mathbf{X} = (\mathbf{x}_1, \ldots, \mathbf{x}_n)^T$ is the design matrix, both are weighted by $\widehat{\tau}_{ij}^{\,1/2\,(k)}$.

- In order to estimate $\widehat{\sigma}_i^2$ and $\widehat{\pi}_i$ we use

$$\widehat{\sigma}^{2\,(k+1)} = \frac{(\mathbf{y} - \mathbf{X}\widehat{\boldsymbol{\beta}}_i^{(k+1)})^{\mathrm{T}}\mathbf{W}_i(\mathbf{y} - \mathbf{X}\widehat{\boldsymbol{\beta}}_i^{(k+1)})}{n}, \qquad \text{for} \quad i = 1, \ldots, c,$$

- and

$$\widehat{\pi}_i^{(k+1)} = \frac{\sum_{j=1}^n \tau_{ij}^{(k)}}{n} \qquad \text{for} \quad i = 1, \ldots, c.$$

Mixture regression models with concomitant variables

- We consider mixture of regressions with concomitant variables of form

$$f(y_j|\mathbf{x}_j, \boldsymbol{\omega}_j) = \sum_{i=1}^{c} \pi_i(\boldsymbol{\omega}_j, \boldsymbol{\alpha}_i)\phi(y_j|\mathbf{x}_j^{\mathrm{T}}\boldsymbol{\beta}_i, \sigma_i^2),$$

- Multinomial logit model for the π_i is given by

$$\pi_i(\boldsymbol{\omega}_j, \boldsymbol{\alpha}_i) = \frac{\exp^{\boldsymbol{\omega}_j^{\mathrm{T}}\boldsymbol{\alpha}_i}}{\sum_{h=1}^{c}\exp^{\boldsymbol{\omega}_j^{\mathrm{T}}\boldsymbol{\alpha}_h}} \qquad \text{for} \qquad i = 1, \ldots, c,$$

with $\boldsymbol{\alpha} = (\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_c)$ and $\boldsymbol{\alpha}_1 \equiv \mathbf{0}$.
- The vector of all parameters is given by

$$\boldsymbol{\Psi} = ((\boldsymbol{\alpha}_1^{\mathrm{T}}, \boldsymbol{\beta}_1^{\mathrm{T}}, \sigma_1^2), \ldots, (\boldsymbol{\alpha}_c^{\mathrm{T}}, \boldsymbol{\beta}_c^{\mathrm{T}}, \sigma_c^2))^{\mathrm{T}}.$$

- Parameters estimated via EM algorithm.
- M-step: maximization of expected complete log likelihood

$$Q(\boldsymbol{\Psi}, \widehat{\boldsymbol{\Psi}}^{(k)}) = Q_1(\boldsymbol{\beta}_i, \sigma_i^2, i = 1, \ldots, c\,; \widehat{\boldsymbol{\Psi}}^{(k)}) + Q_2(\boldsymbol{\alpha}, \widehat{\boldsymbol{\Psi}}^{(k)}),$$

where

$$Q_1(\boldsymbol{\beta}_i, \sigma_i^2, i = 1, \ldots, c\,; \widehat{\boldsymbol{\Psi}}^{(k)}) = \sum_{j=1}^{n} \sum_{i=1}^{c} \widehat{\tau}_{ij}^{(k)} \log\left(\phi(y_j | \mathbf{x}_j^{\mathrm{T}} \boldsymbol{\beta}_i, \sigma_i^2)\right)$$

and

$$Q_2(\boldsymbol{\alpha}, \widehat{\boldsymbol{\Psi}}^{(k)}) = \sum_{j=1}^{n} \sum_{i=1}^{c} \widehat{\tau}_{ij}^{(k)} \log\left(\pi_i(\boldsymbol{\omega}_j, \boldsymbol{\alpha}_i)\right).$$

- Formulas Q_1 and Q_2 can be maximized separately. The maximization of Q_1 gives new estimates $\widehat{\boldsymbol{\beta}}_i^{(k+1)}, \widehat{\sigma}_i^{2(k+1)}$, $i = 1, \ldots, c$, and the maximization of Q_2 gives $\widehat{\boldsymbol{\alpha}}^{(k+1)}$. Q_1 is maximized using the weighted ML estimation of linear models and Q_2 by means of the weighted ML estimation of multinomial logit models.

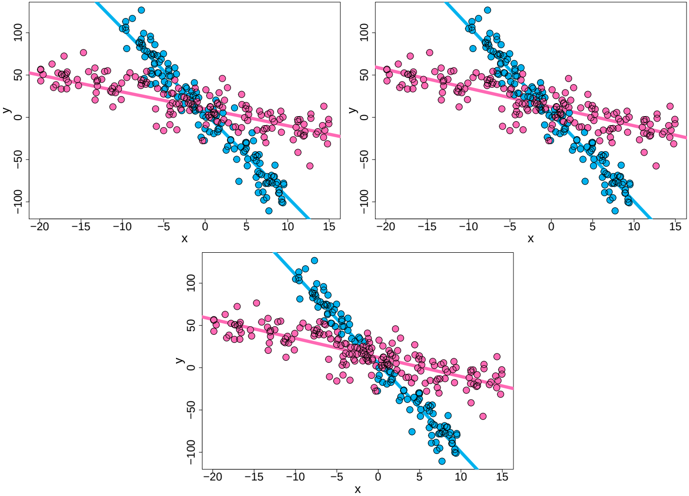
Categorical concomitant variable

- The data comes from two component mixture of the following structure:

$$
\begin{aligned}
150\,\mathrm{obs.}: \quad & y = 5 - 10\,x + \varepsilon; \qquad & \varepsilon \sim \mathrm{N}(0, 15) \\
150\,\mathrm{obs.}: \quad & y = 10 - 2\,x + \varepsilon; \qquad & \varepsilon \sim \mathrm{N}(0, 15)
\end{aligned}
$$

- Categorical concomitant variable assigns value 1 and 2 to observations from the first component (with accuracy of 90%) and value 3 and 4 to observations from the second component (with accuracy of 90%)

- Comparison of true data classification (top left), classification via ordinary mixture of regressions (bottom) and classification via mixture of regressions with concomitant variable (top right).



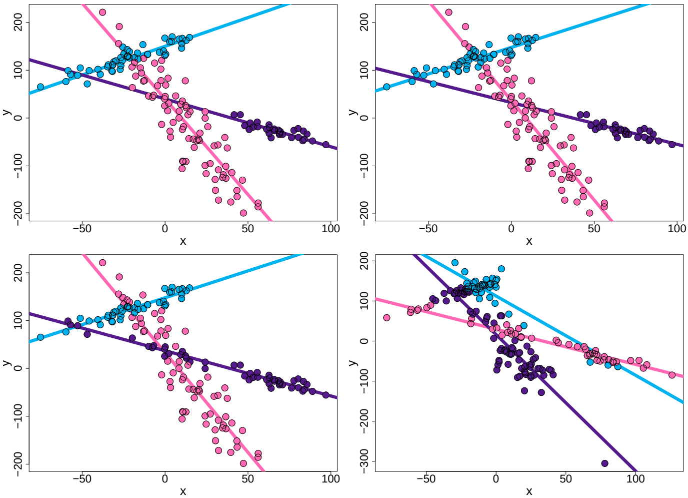
Continuous concomitant variable

- The data comes from three component mixture of the following structure:

$$\begin{array}{lll} 50\,\text{obs.}: & y = 150 - 1.2\,x + \varepsilon; & \varepsilon \sim \mathrm{N}(0,10) \\ 80\,\text{obs.}: & y = 40 - 4\,x + \varepsilon; & \varepsilon \sim \mathrm{N}(0,40) \\ 30\,\text{obs.}: & y = 40 - 1\,x + \varepsilon; & \varepsilon \sim \mathrm{N}(0,10) \end{array}$$

- Explanatory variable can be used as concomitant variable for better estimates of observations membership.

- Comparison of true data classification (top left), classification via ordinary mixture of regressions (bottom left & bottom right) and classification via mixture of regressions with concomitant variable (top right).



- The *MSE* and variance of the regression parameters and standard error estimates for both mixture models:

$MSEPAR(\widehat{\psi}_m)$	β_{10}	β_{11}	β_{20}	β_{21}	β_{10}	β_{11}	β_{20}	β_{21}	β_{30}	β_{31}
klasická směs	2.098	0.095	2.189	0.023	5.508	0.006	55.368	0.069	149.421	0.028
směs s dopr. prom.	**1.763**	**0.057**	**1.809**	**0.017**	**5.017**	0.006	**30.709**	**0.059**	**61.040**	**0.011**
$VAR(\widehat{\psi}_m)$	β_{10}	β_{11}	β_{20}	β_{21}	β_{10}	β_{11}	β_{20}	β_{21}	β_{30}	β_{31}
klasická směs	2.033	0.051	2.063	0.018	**3.683**	0.005	34.297	0.059	**13.816**	**0.003**
směs s dopr. prom.	**1.713**	**0.048**	**1.799**	**0.016**	4.772	0.005	**29.577**	0.059	56.016	0.010

- For three component mixture - most of the standard mixture estimates are completely inaccurate and not included in the estimates quality comparison.

Conclusions

- Mixtures of regression models can be applied to data where observations originate from various groups and the group affiliations are not known.
- Mixture of regressions with concomitant variables provides quality estimates in the sense of lower MSE and variance of parameters.
- Concomitant variable enhances the classification of observations.

Thank you for your attention!

References

- BENAGLIA, T., CHAUVEAU, D., HUNTER, D.R. a YOUNG, D.S.. Mixtools: An R Package for Analyzing Finite Mixture Models. Journal of Statistical Software. 2009, Vol. 32, No. 6. www.jstatsoft.org [online 27.2.2014]
- DE VEAUX, R.D. Mixtures of linear regressions. Comput. Statist. Data Anal. 1989, No. 8, pp. 227–245.
- DESARBO, W.S. a CRON, W.L.. A Maximum Likelihood Methodology for Clusterwise Linear Regression. Journal of Classification. 1988, No. 5. URL: deepblue.lib.umich.edu [online 3.3.2014]
- FARIA, Susana, SOROMENHO, Gilda. Fitting mixtures of linear regressions. Journal of Statistical Computation and Simulation (Impact Factor: 0.63). 03/2010; 80:201-225. DOI:10.1080/00949650802590261. URL: http://www.researchgate.net [online 10. 2. 2014]
- GRÜN, B. a LEISCH. F.. Fitting Finite Mixtures of Generalized Linear Regressions in R. Computational Statistics and Data Analysis. 2006. URL: http://www.ci.tuwien.ac.at [online 5. 5. 2015]
- GRÜN, B. a LEISCH. F.. FlexMix Version 2: Finite Mixtures with Concomitant Variables and Varying and Constant Parameters. Journal of Statistical Software. 2008, Vol. 28, No. 4. URL: cran.at.r-project.org [online 3.3.2014]
- HOSHIKAWA, Toshiya. Mixture regression for observational data, with application to functional regression models. 2013. URL: arxiv.org [online 10.3.2014]
- JONES, P.N., MCLACHLAN, G.J. Fitting finite mixture models in a regression context. Aust. J. Statist. 1992, No. 34, pp. 233–240.
- LEISCH, Friedrich. FlexMix: A General Framework for Finite Mixture Models and Latent Class Regression in R. Journal of Statistical Software. 2004, Vol. 11, No. 8.