Supervised-Component versus PLS regression. The case of GLMMs with autoregressive random effect

Jocelyn Chauvet, Xavier Bry, Catherine Trottier

Supervised–Component versus PLS regression
The case of GLMMs with autoregressive random effect

Jocelyn Chauvet*[1], Xavier Bry[1] and Catherine Trottier[1,2]

[1]IMAG, CNRS, Univ. Montpellier, France.
[2]Univ. Paul–Valéry Montpellier 3, F34000, Montpellier, France.
*Email: jocelyn.chauvet@umontpellier.fr

Abstract: We address two regularised versions of the EM algorithm for Generalised Linear Mixed Models (GLMM) for panel data. A random response y is modelled by a GLMM, using a set X of explanatory variables and two random effects. The first random effect models the dependence within individuals on which data is repeatedly collected while the second one embodies the serially correlated time–specific effect shared by all the individuals. Variables in X are assumed many and redundant, so that regression demands regularisation. In this context, we first propose a ridge–penalised EM algorithm, and then a supervised component–based regularised EM algorithm as an alternative. An attempt will be made to compare the latter with the PLS regression.

Introduction

In the context of GLMMs having a large number of redundant covariates, penalty–based approaches such as ridge (Eliot *et al.* (2011)) or lasso (Groll and Tutz (2014)) on the one hand and component–based approaches on the other (Chauvet *et al.* (2016)) have already been highlighted. Focussing on situations where variable selection is inappropriate, we propose both ridge and Supervised Component (SC) estimation techniques for fitting a GLMM in the context of panel data with an autoregressive time–specific random effect.

General principles

We first consider a Gaussian balanced panel data with q_1 individuals, each of them observed at the same q_2 time–points. With $n = q_1 q_2$ and $q = q_1 + q_2$, the model writes $y = X\beta + U\xi + \varepsilon$, where $y \in \mathbb{R}^n$ is the response vector, $\beta \in \mathbb{R}^p$ and $\xi \in \mathbb{R}^q$ respectively the fixed and random effects vectors (X and U being their associated design matrices), and $\varepsilon \sim \mathcal{N}_n(0, \sigma_0^2 \mathrm{Id}_n)$ the residuals. We further assume that $\xi = (\xi_1^{\mathrm{T}}, \xi_2^{\mathrm{T}})^{\mathrm{T}}$, where $\xi_1 \sim \mathcal{N}_{q_1}(0, \sigma_1^2 \mathrm{Id}_{q_1})$ is the individual–specific random effect and $\xi_2 \sim \mathcal{N}_{q_2}(0, \sigma_2^2 A_2(\rho))$, with $A_2(\rho) = \left(\frac{\rho^{|i-j|}}{1-\rho^2}\right)_{1 \leqslant i,j \leqslant q_2}$, the order–1 autoregressive time–specific random effect. For rank–1 component, we set $\beta = u\gamma$, with $||u|| = 1$ and $\gamma \in$

$\mathbb{R}$ the regression parameter associated with component Xu. Instead of subtracting a penalty term to the likelihood, we add a bonus term favouring the alignment of the component on the most interpretable directions in the explanatory subspace. For that, we take into account the structural relevance of component Xu, defined as $\phi(u) = \left(\sum_j (u^{\mathrm{T}} N_j u)^{\ell}\right)^{\frac{1}{\ell}}$, where the N_j's are s.d.p matrices encoding the type of structures of interest in X and $\ell \geqslant 1$ is a parameter tuning the locality of bundles to be considered. L denoting the complete log–likelihood and $\theta = (\beta, \sigma_0^2, \sigma_1^2, \sigma_2^2, \rho)$, we present the current iteration of both ridge– and single SC–regularised EMs.

---

ridge E–step: Define $\mathcal{Q}_{\mathbf{rid}}\left(\theta \,|\, \theta^{[t]}\right) = \mathbb{E}_{\xi|y}\left[L\left(\theta; y, \xi\right) - \boldsymbol{\lambda}\,\|\boldsymbol{\beta}\|_{\mathbf{2}}^{\mathbf{2}} \;|\, \theta^{[t]}\right]$

SC E–step: Define $\mathcal{Q}_{\mathbf{SC}}\left(\theta \,|\, \theta^{[t]}\right) = \mathbb{E}_{\xi|y}\left[\boldsymbol{(1-s)}L\left(\theta; y, \xi\right) + \boldsymbol{s}\,\boldsymbol{\log \phi(u)} \;|\, \theta^{[t]}\right]$

M–step: Set $\theta^{[t+1]} = \arg\max_{\theta} \mathcal{Q}_{\mathbf{rid}}\left(\theta \,|\, \theta^{[t]}\right)$ or $\theta^{[t+1]} = \arg\max_{\theta:||u||=1} \mathcal{Q}_{\mathbf{SC}}\left(\theta \,|\, \theta^{[t]}\right)$

Trade–off parameter $s \in [0, 1]$ and parameter ℓ are tuned by cross–validation (as shrinkage parameter $\lambda \geqslant 0$ for ridge) and higher rank components are computed like the rank–1, subject to extra orthogonality constraints. The extension to GLMMs is inspired by the Schall's iterative scheme alternating linearisation of the model and parameters' estimation. The idea is to keep the same linearisation step, but replace the usual estimation step with a "local" regularised EM.

Conclusion

Both methods were tested on simulated Gaussian and Poisson data and perform well in terms of estimation and prediction. But unlike ridge, SC gives access to interesting graphical diagnoses that reveal multidimensional predictive structures and greatly facilitate the interpretation of the model.

References

Chauvet, J., Bry, X., Trottier, C. and Mortier, F. (2016). Extension to mixed models of the Supervised Component–based Generalised Linear Regression. In: *Proceedings COMPSTAT 2016*, Springer.

Eliot, M., Ferguson, J., Reilly, M.P. and Foulkes, A.S. (2011). Ridge Regression for Longitudinal Biomarker Data. *The International Journal of Biostatistics*, **7**, pp. 1 – 11.

Groll, A. and Tutz, F. (2014). Variable selection for generalized linear mixed models by L_1-penalized estimation. *Statistics and Computing*, **24**, pp. 137 – 154.