An algorithm for high-dimensional generalized linear mixed models using ℓ_1-penalization

Jürg Schelldorfer

joint work with Peter Bühlmann

encouraged by Stephan Dlugosz

FOR916 Workshop on ”Mathematical Statistics meets Econometrics”

June 9, 2011

Motivation

data set of Stephan Dlugosz:

administrative data set about employment:

Binary response variable $Y \in \{\text{employed}, \text{unemployed}\}$

covariates X: income, sex, age group, employment duration,....

quarterly results of (Y, X) of many workers over several years

General framework

- response variable from the exponential family
- continuous covariates
- grouped observations (think of longitudinal data, repeated measures data)

Goal:
Performing variable selection in the setup where AIC, BIC, cAIC, mAIC, ... are computationally infeasible (i.e. $n \approx p, n \ll p$)

Table of Contents

Overview

	$n > p$	$n \ll p$
Generalized Linear Models (GLMs)	MLE [IRLS]	Lasso [`R:glmnet`]
Generalized Linear Mixed Models (GLMMs)	MLE [`R:glmer`]	?

n: number of observations

p: number of variables

Generalized Linear Model (GLM)

For n observations (y_i, x_i^T)

- (y_i, x_i^T) are independent for $i = 1, \ldots, n$
- y_i has a density of the form

$$\exp\left\{\phi^{-1}\Big(y_i\xi_i - b(\xi_i)\Big) + c(y_i, \phi)\right\}\text{with } \boldsymbol{\mu}_i = \mathbb{E}[y_i]$$

- $g(\boldsymbol{\mu}) = \boldsymbol{\eta}$ with $\boldsymbol{\eta} = \boldsymbol{X}\boldsymbol{\beta}$

Then estimate $\boldsymbol{\beta}$ by

$$\hat{\boldsymbol{\beta}}_{MLE} = argmin_{\boldsymbol{\beta}} - \ell(\boldsymbol{\beta})$$

ℓ_1-regularized Generalized Linear Model

For $n \ll p$ we should not use the MLE. Use the Lasso (Tibshirani, 1996)

$$\hat{\beta}(\lambda) = argmin_{\beta} - \ell(\beta) + \lambda\|\beta\|_1 \quad , \quad \lambda > 0$$

with the following properties:

- The Lasso does variable selection (i.e. some coefficients are set exactly to zero)
- Convex optimiziation problem, which can be solved efficiently

Generalized Linear Mixed Model (GLMM)

$g = 1, ..., N$ independent groups/clusters/subjects
$j = 1, .., n_g$ observations for group/cluster/subject g
$n = \sum_{g=1}^{N} n_g$ total number of observations

$\boldsymbol{y}$: n-dim response variable
$\boldsymbol{b}$: q-dim (correlated) random effects

$\boldsymbol{\beta} \in \mathbb{R}^p$ fixed-effects parameters
$\boldsymbol{\theta} \in \mathbb{R}^L$ covariance parameters
ϕ dispersion parameter

$\boldsymbol{X}$: $n \times p$ model matrix for $\boldsymbol{\beta}$
$\boldsymbol{Z}$: $n \times q$ model matrix for $\boldsymbol{b}$
$\boldsymbol{\Sigma}_{\boldsymbol{\theta}}$: $q \times q$ covariance matrix, determined by $\boldsymbol{\theta}$

Generalized Linear Mixed Model (GLMM)

Model Assumptions:

- $y_i|\boldsymbol{b}$ are independent for $i = 1, \ldots, n$
- $y_i|\boldsymbol{b}$ has a density of the form

$$\exp\left\{\phi^{-1}\Big(y_i\xi_i - b(\xi_i)\Big) + c(y_i, \phi)\right\} \text{ with } \boldsymbol{\mu}_i = \mathbb{E}[y_i|\boldsymbol{b}]$$

- $g(\boldsymbol{\mu}) = \boldsymbol{\eta}$ with $\boldsymbol{\eta} = \boldsymbol{X}\beta + \boldsymbol{Z}\boldsymbol{b}$
- $\boldsymbol{b} \sim \mathcal{N}_q(\mathbf{0}, \boldsymbol{\Sigma}_{\boldsymbol{\theta}})$ with $\boldsymbol{\Sigma}_{\boldsymbol{\theta}} \geq 0$ for $\boldsymbol{\theta} \in \mathbb{R}^L$

$$(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\theta}}, \hat{\phi})_{MLE} = argmin_{\boldsymbol{\beta},\boldsymbol{\theta},\phi} - \log L(\boldsymbol{\beta}, \boldsymbol{\theta}, \phi)$$

Recap

	$n > p$	$n \ll p$
Generalized Linear Models (GLMs)	MLE [IRLS] ✔	Lasso [`R:glmnet`] ✔
Generalized Linear Mixed Models (GLMMs)	MLE [`R:glmer`] ✔	!

High-dimensional GLMM Set-up

Additionally to a GLMM, assume

- $n = \sum_{i=1}^{N} n_g \ll p$
- the true β_0 is sparse
- L small

Aim: Estimate $\boldsymbol{\beta}, \boldsymbol{\theta}$, ϕ and predict $\boldsymbol{b}$

The GLMMLasso estimator

KeyIdea 1: Lasso-type penalty

Estimate the parameters $(\boldsymbol{\beta}, \theta, \phi)$ by minimizing

$$Q_\lambda(\boldsymbol{\beta}, \boldsymbol{\theta}, \phi) := -2 \log L(\boldsymbol{\beta}, \boldsymbol{\theta}, \phi) + \lambda \|\boldsymbol{\beta}\|_1,$$

$$(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\theta}}, \hat{\phi}) := \mathit{argmin}_{\boldsymbol{\beta}, \boldsymbol{\theta}, \phi} Q_\lambda(\boldsymbol{\beta}, \boldsymbol{\theta}, \phi).$$

Remark: In general, $L(\boldsymbol{\beta}, \boldsymbol{\theta}, \phi)$ cannot be computed explicitly.

Laplace approximation

$\boxed{\textbf{KeyIdea 2}\text{: Laplace approximation}}$ to approximate the integrand of $L(\boldsymbol{\beta}, \boldsymbol{\theta}, \phi)$ by a quadratic function.

$$I = \int_{\mathbb{R}^q} e^{-S(\boldsymbol{b})} d\boldsymbol{b} \approx (2\pi)^{q/2} |S''(\tilde{\boldsymbol{b}})|^{-1/2} e^{-S(\tilde{\boldsymbol{b}})}$$

where $\tilde{\boldsymbol{b}} = argmin_{\boldsymbol{b}} S(\boldsymbol{b})$ is the mode of $-S(\boldsymbol{b})$.

Hence

$$Q_\lambda(\boldsymbol{\beta}, \boldsymbol{\theta}, \phi) \rightsquigarrow \tilde{Q}_\lambda^{LA}(\boldsymbol{\beta}, \boldsymbol{\theta}, \phi)$$

The GLMMLasso estimator

The GLMMLasso estimator is defined by

$$(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\theta}}, \hat{\phi}) := \textit{argmin}_{\boldsymbol{\beta},\boldsymbol{\theta},\phi} \tilde{Q}_{\lambda}^{LA}(\boldsymbol{\beta}, \boldsymbol{\theta}, \phi)$$

Remark: It is a non-convex optimization problem!

The GLMMLasso algorithm I

How to calculate

$$(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\theta}}, \hat{\phi}) := argmin_{\boldsymbol{\beta},\boldsymbol{\theta},\phi} \tilde{Q}_{\lambda}^{LA}(\boldsymbol{\beta}, \boldsymbol{\theta}, \phi)?$$

KeyIdea 3: coordinate-wise optimization with inexact line search,
i.e. optimize $\tilde{Q}_{\lambda}^{LA}$ w.r.t. one coordinate keeping all other coordinates fixed (Tseng and Yun, 2009):

- **Quadratic approximation** of the objective function
- calculate the **gradient**
- **Inexact line search** using the Armijo rule

The GLMMLasso algorithm II

GLMMLasso algorithm

(0) *Choose a starting value* $(\boldsymbol{\beta}^{(0)}, \boldsymbol{\theta}^{(0)}, \phi^{(0)})$.

Repeat *for* $s = 1, 2, \ldots$

(1) *(Fixed-effects parameter optimization) For* $k = 1, \ldots, p$

a) *(Laplace approximation)*
Calculate the Laplace approximation $\tilde{Q}_{\lambda}^{LA}(., ., .)$

b) *(Quadratic approximation and inexact line search)*

i) *Approximate the second derivative by* $h_k^{(s)} > 0$.

ii) *Calculate the descent direction* $d_k^{(s)} \in \mathbb{R}$

iii) *Choose a step size* $\alpha_k^{(s)} > 0$ *such that there is a decrease in the objective function.*

(2) *(Covariance parameter optimization) For* $l = 1, \ldots, L$

$$\theta_l^{(s)} = argmin_{\theta_l} \tilde{Q}_{\lambda}^{LA}(., ., .)$$

(3) *(Dispersion parameter optimization)*

$$\phi^{(s)} = argmin_{\phi} \tilde{Q}_{\lambda}^{LA}(., ., .)$$

until *convergence.*

Tools to speed up

Two ingredients which speed up the algorithm remarkably:

- $\boxed{\textbf{KeyIdea 4a}\text{: regard } \tilde{\boldsymbol{b}} \text{ as fixed}}$ for the quadratic approximation w.r.t. β_k
- $\boxed{\textbf{KeyIdea 4b}\text{: active-set algorithm}}$ cycle through the non-zero coefficients β_k, and only through all p coefficients every Dth iteration

This two ingredients make it feasable to calculate large data sets (i.e. $n = 400$ and $p = 4000$)!

The price to pay

Small additional bias in the parameter estimates, and similar variable selection properties.

Application in Econometrics

This is ongoing work with Stephan Dlugosz on administrative data.

Take-home message

	$n > p$	$n \ll p$
Generalized Linear Models (GLMs)	MLE [IRLS]	Lasso [`R:glmnet`]
Generalized Linear Mixed Models (GLMMs)	MLE [`R:glmer`]	GLMMLasso **KeyIdea 1-4**

Thank you!

Questions?

References

- P. Tseng and S. Yun ; A Coordinte Gradient Descent Method for Nonsmooth Separable Minimization ; Mathematical Programming (2009)
- R. Tibshirani ; Regression Shrinkage and Selection via the Lasso ; J. R. Stat. Soc. (1996)
- J. Schelldorfer, P. Bühlmann and S. van de Geer ; Estimation for High-Dimensional Linear Mixed-Effects Models Using ℓ_1-penalization ; The Scandinavian Journal of Statistics (2011)
- D. Bates ; lme4: Mixed-effects modeling with R (to appear)