Stat Comput manuscript No.
(will be inserted by the editor)

Nonlinear Mixed-effects Scalar-on-function Models and Variable Selection - Supplementary material

Yafeng Cheng · Jian Qing Shi · Janet Eyre

Received: date / Accepted: date

1 Normalization methods to illuminate effetcs from different dimensions of variables

The group lars algorithm introduces one extra parameter to correct the influence of the difference of the dimensions among different groups of variables. In our case, such difference is more dramatic.

Recall that the LARS algorithm uses the following equation to find the distance to move on the direction unit vector with respect to the scalar candidate variable z:

$$\mathrm{Cor}(r - \alpha u, z)^2 = \mathrm{Cor}(r - \alpha u, u)^2$$

where u is the unit vector representing the direction of the current iteration. The iteration number is omitted.

The group LARS from Yuan and Lin (2006) has a similar equation:

$$||(r - \alpha u)^T \mathbf{z}_j||_2^2 / p_j = ||(r - \alpha u)^T \mathbf{z}_i||_2^2 / p_i$$

where $||\cdot||_2$ means the Euclidean norm of the vector; the group of variables $\mathbf{z}_i$ is one of the variables selected; $\mathbf{b}_i$ is the corresponding coefficient; $\mathbf{z}_j$ is one of the candidate variables; p_i and p_j are the dimensions of $\mathbf{x}_i$ and $\mathbf{x}_j$, respectively. The dimensions p_i and p_j are used to remove the effect of the difference in dimensions. Note that the direction vector u here is not unit vector.

Yafeng Cheng
MRC Biostatistics Unit, University of Cambridge
E-mail: yafeng.cheng@mcr-bsu.cam.ac.uk

Jian Qing Shi
School of Mathematics, Statistics and Physics, Newcastle University
E-mail: j.q.shi@ncl.ac.uk

Janet Eyre
Institute of Neurosciences, Newcastle University
E-mail: janet.eyre@ncl.ac.uk

This equation is equivalent to:

$$\frac{(r-\alpha u)^T(\mathbf{z}_j\mathbf{z}_j^T)(r-\alpha u)}{p_j} = \frac{(r-\alpha u)^T(\mathbf{z}_i\mathbf{z}_i^T)(r-\alpha u)}{p_i}$$
$$(r-\alpha u)^T\frac{\mathbf{z}_j\mathbf{z}_j^T}{p_j}(r-\alpha u) = (r-\alpha u)^T\frac{\mathbf{z}_i\mathbf{z}_i^T}{p_i}(r-\alpha u).$$

The normalization is on the matrices $\mathbf{z}_j\mathbf{z}_j^T$ and $\mathbf{z}_i\mathbf{z}_i^T$, as these two matrices are the only different elements between left and right hand sides of the equation.

For functional candidates, we have

$$\mathrm{Cor}\left(r-\alpha u, \int x(t)\beta(t)dt\right)^2/N_f = \mathrm{Cor}(r-\alpha u, u)^2/N_u; \tag{1}$$

while for scalar candidates, we have

$$\mathrm{Cor}\,(r-\alpha u, z)^2/N_z = \mathrm{Cor}(r-\alpha p, p)^2/N_u. \tag{2}$$

The parameters N_f, N_z and N_u are constants for normalization.

1.1 For functional candidate variables

We omit the subscript of the variables in the following derivation. In the main manuscript, we obtained the solution of the modified functional canonical correlation:

$$\text{correlation:}\quad \rho^2 = \frac{V_{x,y}^T P_{x,x}^{-1} V_{x,y}}{V_y} \tag{3}$$

$$\text{coefficients for } x(t):\quad \tilde{C}_b = \frac{P_{x,x}^{-1}V_{x,y}}{\rho||y||_2}, \tag{4}$$

$$\text{coefficients for } y:\quad a = 1/\mathrm{sd}(y). \tag{5}$$

After the k-th iteration in the flars algorithm, we have scalar variable $r\alpha u$ instead of y from Eqn (3). Thus we have:

$$\rho^2 = \frac{V_{xy}^T P_{x,x}^{-1} V_{x,y}}{V_y} \tag{6}$$

$$= \frac{(r\alpha u)^T\mathbf{x}WK^{-1}W^T\mathbf{x}^T(r\alpha u)}{(r\alpha u)^T(r\alpha u)}, \tag{7}$$

where $K = P_{x,x}$. If we substitute Eqn (7) into left hand side of Eqn (1) and expand the right hand side of Eqn (1), we can get:

$$\frac{(r-\alpha u)^T\mathbf{x}WK^{-1}W^T\mathbf{x}^T(r-\alpha u)}{(r-\alpha u)^T(r-\alpha u)N_f} = \frac{[(r-\alpha u)^T u/(n-1)]^2}{(r-\alpha u)^T(r-\alpha u)u^T u/(n-1)^2 N_u}$$
$$\frac{(r-\alpha u)^T\mathbf{x}WK^{-1}W^T\mathbf{x}^T(r-\alpha u)}{(r-\alpha u)^T(r-\alpha u)N_f} = \frac{(r-\alpha u)^T uu^T(r-\alpha u)}{(r-\alpha u)^T(r-\alpha u)u^T pN_u}$$
$$(r-\alpha u)^T\bar{S}(r-\alpha u) = (r-\alpha u)^T\bar{U}(r-\alpha u).$$

where $\bar{S} = \frac{\mathbf{x}WK^{-1}W^T\mathbf{x}^T}{N_f}$,$\bar{U} = \frac{u(u^Tu)^{-1}u^T}{N_u}$; $\mathbf{x}$ is the discrete data points of the functional variable $x(t)$. The estimated α is the solution of the quadratic function:

$$\alpha^2[u^T(\bar{S}-\bar{U})u] - 2\alpha[r^T(\bar{S}-\bar{U})u] + [r^T(\bar{S}-\bar{U})r] = 0 \tag{8}$$

1.2 For scalar candidate variables

If we expand both side of Eqn (2):

$$\frac{[(r-\alpha u)^T z/(n-1)]^2}{(r-\alpha u)^T(r-\alpha u)z^T z/(n-1)^2 N_z} = \frac{[(r-\alpha u)^T u/(n-1)]^2}{(r-\alpha u)^T(r-\alpha u)u^T u/(n-1)^2 N_u}$$
$$\frac{(r-\alpha u)^T z z^T(r-\alpha u)}{z^T z N_z} = \frac{(r-\alpha u)^T u u^T(r-\alpha u)}{u^T u N_u}$$
$$(r-\alpha u)^T \bar{Z}(r-\alpha u) = (r-\alpha u)^T \bar{U}(r-\alpha u).$$

where $\bar{Z} = \frac{z(z^T z)^{-1} z^T}{N_z}$. The solution of α is the solution of the quadratic function:

$$\alpha^2[u^T(\bar{Z}-\bar{U})u] - 2\alpha[r^T(\bar{X}-\bar{U})u] + [r^T(\bar{X}-\bar{U})r] = 0 \tag{9}$$

1.3 Normalization method

The intuitions from Eqn(8) and Eqn(9) are identical. Thus we can unify the formulas for scalar and functional cases. Follow the spirit of the group LARS, we tested rank, trace and Frobenius norm in the simulation study, and found out that Frobenius norm is the most stable and best performed one. Thus we use Frobenius norm as the normalization method in the algorithm.

2 Calculation of the degrees of freedom for fLARS

Here we propose our definition of the degrees of freedom based on the hat matrix. In the functional LARS algorithm, the residual after iteration k can be written as

$$r^{(k+1)} = r^{(k)} - \alpha^{(k)} u^{(k)}$$

where $u^{(k)}$ is the direction vector, calculated by:

$$u^{(k)} = \frac{H_k r^{(k)}}{\mathrm{sd}(H_k r^{(k)})}.$$

Therefore, the true "hat" matrix at iteration k is :

$$H_k^* = \frac{H_k \alpha^{(k)}}{\mathrm{sd}(H_k r^{(k)})},$$

where $H_k = \mathbf{X}W(W^T\mathbf{X}^T\mathbf{X}W + \lambda_1 W_2 + \lambda_2 W^T W)^{-1}W^T\mathbf{X}^T$ from Eqn (4).

The residual after iteration k becomes:

$$r^{(k+1)} = (I - H_k^*)r^{(k)}.$$

Recursively, the fitted value after iteration K with respect to the response variable is:

$$\begin{aligned}\hat{y} =& y - r^{(K+1)} \\ =& y - [\prod_{k=1}^{K}(I - H_k^*)]y \\ =& [I - \prod_{k=1}^{K}(I - H_k^*)]y,\end{aligned}$$

hence the "hat" matrix $\bar{H}_K$ after iteration K is

$$\bar{H}_K = I - \prod_{k=1}^{K}(I - H_k^*). \tag{10}$$

Similar to that in Efron et al. (2003); Barreca et al. (2005) We then define the degrees of freedom for fLARS as follows:

$$\begin{aligned}\mathrm{df}^* &= \mathrm{tr}\left(\frac{\mathrm{Cov}(\hat{\boldsymbol{\mu}^*}^T, \mathbf{y}^{*T})}{\sigma^2}\right) \\ &= \mathrm{tr}\left(\frac{\mathrm{Cov}(\mathbf{y}^{*T}\bar{H}_k^T, \mathbf{y}^{*T})}{\sigma^2}\right) \\ &= \mathrm{tr}\left(\frac{\bar{H}_k\mathbf{y}^*\mathbf{y}^{*T}/(n-1)}{\sigma^2}\right)\end{aligned}$$

where $\mathbf{y}^*\mathbf{y}^{*T}/(n-1)$ is an $n \times n$ matrix. The i,j-th element is $\mathrm{Cov}(y_i, y_j)$. Its value is σ^2 if $i = j$ and 0 otherwise. Hence:

$$\begin{aligned}\mathrm{df}^* &= \mathrm{tr}\left(\frac{\bar{H}_k\mathbf{y}^*\mathbf{y}^{*T}/(n-1)}{\sigma^2}\right) \\ &= \mathrm{tr}\left(\frac{\bar{H}_k\sigma^2 I}{\sigma^2}\right) \\ &= \mathrm{tr}(\bar{H}_k).\end{aligned} \tag{11}$$

3 Variables selected in linear regression with mixed functional and scalar variables using group lasso

We extend the valriable selection algorithm for functinal linear regression proposed by Gertheiss et al. (2013) to the case where both scalar and functional are in the regression. This extended version of the algorithm is used in the comparison in the simulation study. Recall that in the manuscript, the target linear model is a regression with mixed functional and scalar candidate variables:

$$y_{i,d} = \sum_{j=1}^{J}\int x_{i,d,j}(t)\beta_j(t)dt + \sum_{m=J+1}^{J+M} z_{i,d,m}\gamma_m + \epsilon_{i,d}, \tag{12}$$

The notation used here is the same as those in the manuscript. To simplify the notation, I will drop the subscripts i and d in the equation. The objective function of the algorithm is:

$$G = (y-\sum_{j=1}^{J} \mathbf{x}_j \boldsymbol{\Phi}^T \tilde{b}_j^T/k- \sum_{m=J+1}^{J+M} z_m\gamma_m)^2+\lambda(\sum_{j=1}^{J} \sqrt{p_j}(\tilde{b}_j^T(\boldsymbol{\Phi}^T\boldsymbol{\Phi}+\varphi\boldsymbol{\Phi}_2\boldsymbol{\Phi}_2^T)\tilde{b}_j)^{1/2}+ \sum_{m=J+1}^{J+M} ||\gamma_m||_2), \tag{13}$$

where $\mathbf{x}_j$ is the functional variable represented by RDP with dimension k, $\boldsymbol{\Phi}$ is the basis function for functional coefficients, $\tilde{b}$ is the coefficients for the functional coefficients. The tuning parameter λ is for lasso penalty and φ is for roughness penalty. The difference between our extension and the original version algorithm is that we included the scalar variables and added corresponding penalty for lasso variable selection.

The objective function can then be written as:

$$\begin{aligned}
G =&(y - \sum_{j=1}^{J} \mathbf{x}_j \boldsymbol{\Phi}^T \tilde{b}^T/k - \sum_{m=1}^{M} z_m\gamma_m)^2 + \lambda(\sum_{j=1}^{J} \sqrt{p_j}(\tilde{b}_j^T L^T L\tilde{b}_j)^{1/2} + \sum_{m=1}^{M} ||\gamma_m||_2)\\
=&(y - \sum_{j=1}^{J} \mathbf{x}_j \boldsymbol{\Phi}^T L^{-1} L\tilde{b}^T/k - \sum_{m=1}^{M} z_m\gamma_m)^2 + \lambda(\sum_{j=1}^{J} \sqrt{p_j}(\tilde{b}_j^T L^T L\tilde{b}_j)^{1/2} + \sum_{m=1}^{M} ||\gamma_m||_2)\\
=&(y - \sum_{j=1}^{J} \mathbf{x}_j^* \tilde{b}^{*T}/k - \sum_{m=1}^{M} z_m\gamma_m)^2 + \lambda(\sum_{j=1}^{J} \sqrt{p_j}(\tilde{b}_j^{*T} \tilde{b}_j^*)^{1/2} + \sum_{m=1}^{M} ||\gamma_m||_2),
\end{aligned}$$

where L is from the Cholesky decomposition of the penalty term $K = \boldsymbol{\Phi}^T\boldsymbol{\Phi} + \varphi\boldsymbol{\Phi}_2\boldsymbol{\Phi}_2^T$, and

$$\begin{aligned}
\mathbf{x}_j^* &= \mathbf{x}_j \boldsymbol{\Phi}^T L^{-1}\\
\mathbf{b}_j^* &= L\tilde{b}_j.
\end{aligned}$$

After such transformation, the we can directly use the group lasso method and the corresponding code for selection. The Cholesky decomposition requires the targeting matrix to be positive definite. The penalty term K is not guaranteed to be so in practice.

References

Efron, B., Hastie, T., Johnstone, I., Tibshirani, R., et al. (2003). Least angle regression. *The Annals of statistics*, 32, 407-499.

Yuan, M. and Lin, Y. (2006). Model selection and estimation in regression with grouped variables. *Journal of the Royal Statistical Society: Series B*, 68, 49-67.

Zou, H., Hastie, T., and Tibshirani, R. (2007) On the "degrees of freedom" of the lasso. *The Annals of Statistics*, 35, 2173–2192.

Gertheiss, J., Maity, A., and Staicu, A. M. (2013) Variable selection in generalized functional linear models. *Stat*, 2, 86-101.