Penalties available in grpreg

Patrick Breheny

May 18, 2017

1 Introduction

This vignette summarizes the penalties available in grpreg and describes how specifying various options in the code corresponds to different mathematical forms of the penalty.

The following notation will be used. The objective function $Q(\boldsymbol{\beta})$ being minimized can be decomposed into

$$Q(\boldsymbol{\beta}|\mathbf{X}, \mathbf{y}) = L(\boldsymbol{\beta}|\mathbf{X}, \mathbf{y}) + P(\boldsymbol{\beta}),$$

where $L(\boldsymbol{\beta}|\mathbf{X}, \mathbf{y})$ denotes the loss function (scaled negative log-likelihood) and $P(\boldsymbol{\beta})$ denotes the penalty. In what follows,

- $\boldsymbol{\beta}$ denotes the entire vector of regression coefficients

- $\boldsymbol{\beta}_j$ denotes the vector of regression coefficients corresponding to the jth group

- β_{jk} denotes kth regression coefficient in the jth group

- $\|\boldsymbol{\beta}_j\|$ denotes the Euclidean (L_2) norm of $\boldsymbol{\beta}_j$

- $\|\boldsymbol{\beta}_j\|_1$ denotes the L_1 norm of β_j (Manhattan distance)

To clarify, by "scaled" loss function, we mean that the loss function has been multiplied by a factor of $1/n$. So, for least-squares loss,

$$L(\boldsymbol{\beta}|\mathbf{X}, \mathbf{y}) = \frac{1}{2n}\|\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\|^2$$

while for binomial loss

$$L(\boldsymbol{\beta}|\mathbf{X}, \mathbf{y}) = -\frac{1}{n}\sum_{i=1}^{n}\left\{y_i \log \pi_i + (1 - y_i) \log (1 - \pi_i)\right\},$$

and so on.

2 Group selection

2.1 Group lasso

```
> grpreg(X, y, group, penalty="grLasso")
```

$$P(\beta) = \lambda \sum_{j}\|\boldsymbol{\beta}_j\|$$

2.2 Group MCP

```
> grpreg(X, y, group, penalty="grMCP")
```

$$P(\boldsymbol{\beta}) = \sum_j \text{MCP}_{\lambda,\gamma}(\|\boldsymbol{\beta}_j\|)$$

where $\text{MCP}_{\lambda,\gamma}(\cdot)$ denotes the MCP penalty with regularization parameter λ and tuning parameter γ.

2.3 Group SCAD

```
> grpreg(X, y, group, penalty="grSCAD")
```

$$P(\boldsymbol{\beta}) = \sum_j \text{SCAD}_{\lambda,\gamma}(\|\boldsymbol{\beta}_j\|)$$

where $\text{SCAD}_{\lambda,\gamma}(\cdot)$ denotes the SCAD penalty with regularization parameter λ and tuning parameter γ. If you use any of the above penalties, please cite

Breheny P and Huang J (2015). Group descent algorithms for nonconvex penalized linear and logistic regression models with grouped predictors. *Statistics and Computing*, **25**: 173-187. [pdf].

The article goes into more mathematical details, discusses issues of standardization in the group sense, and provides references.

3 Bi-level selection

3.1 Group exponential lasso (GEL)

```
> grpreg(X, y, group, penalty="gel")
```

$$P(\beta) = \sum_j f_{\lambda,\tau}(\|\boldsymbol{\beta}_j\|)$$

where $f(\cdot)$ denotes the exponential penalty with regularization parameter λ and tuning parameter τ:

$$f_{\lambda,\tau}(\theta) = \frac{\lambda^2}{\tau}\left\{1 - \exp\left(-\frac{\tau\theta}{\lambda}\right)\right\}$$

If you use the GEL penalty, please cite

Breheny P (2015). The group exponential lasso for bi-level variable selection. *Biometrics*, **71**: 731-740. [pdf]

3.2 Composite MCP

```
> grpreg(X, y, group, penalty="cMCP")
```

$$P(\boldsymbol{\beta}) = \sum_j \text{MCP}_{\lambda,\gamma_1}\left(\sum_k \text{MCP}_{\lambda,\gamma_2}(|\beta_{jk}|)\right)$$

where $\text{MCP}_{\lambda,\gamma}(\cdot)$ denotes the MCP penalty with regularization parameter λ and tuning parameter γ. If you use the composite MCP penalty, please cite either of the following papers:

Breheny P and Huang J (2009). Penalized methods for bi-level variable selection. *Statistics and Its Interface*, **2**: 369-380. [pdf]

Huang J, Breheny P and Ma S (2012). A Selective Review of Group Selection in High-Dimensional Models. *Statistical Science*, **27**: 481-499. [pdf]

Please note that there is some confusion around the name "group MCP". In the first paper above (2009), the composite MCP penalty was referred to as the "group MCP" penalty; the second paper (2012), in reviewing the various kinds of group penalties that had been proposed, recommended changing the name to "composite MCP" to avoid confusion with the "group MCP" of Section 2.2.

3.3 Group bridge

```
> gBridge(X, y, group)
```

$$P(\boldsymbol{\beta}) = \lambda \sum_j K_j^\gamma \|\boldsymbol{\beta}_j\|_1^\gamma$$

where K_j denotes the number of elements in group j.

Please note that the group bridge penalty uses a very different algorithm from the other penalties. Due to the nature of the penalty, model fitting is slower and less stable for group bridge models. This is, in fact, the main motivation of the GEL penalty of Section 3.1: to offer a more tractable alternative to group bridge that has similar estimation properties but is much better behaved from a numerical optimization perspective.

If you use the group bridge penalty, please cite either of the following papers:

Huang J, Ma S, Xie H and Zhang C (2009). A group bridge approach for variable selection. *Biometrika*, **96**: 339-355.

Breheny P and Huang J (2009). Penalized methods for bi-level variable selection. *Statistics and Its Interface*, **2**: 369-380. [pdf]

The first paper proposed the method; the second paper proposed the algorithm that is used in the grpreg package.

4 Specifying an additional ridge component

For all of the penalties in the previous section, grpreg allows the specification of an additional ridge (L_2) component to the penalty. This will set $\lambda_1 = \alpha\lambda$ and $\lambda_2 = (1-\alpha)\lambda$, with the penalty given by

$$P(\boldsymbol{\beta}) = P_1(\boldsymbol{\beta}|\lambda_1) + \frac{\lambda_2}{2}\|\boldsymbol{\beta}\|^2,$$

where P_1 is any of the penalties from the earlier sections. So, for example

```
> grpreg(X, y, group, penalty="grLasso", alpha=0.75)
```

will fit a model with penalty

$$P(\beta) = 0.75\lambda \sum_j \|\boldsymbol{\beta}_j\| + \frac{0.25\lambda}{2}\|\boldsymbol{\beta}\|^2.$$