Variable Selection In Additive Gene Environment Interactions with the Group Lasso

Sahir Bhatnagar and Yi Yang

January 31, 2017

1 Introduction

We consider a regression model for an outcome variable $\mathbf{Y} = (Y_1, \ldots, Y_n)$ where n is the number of subjects. Let $E = (E_1, \ldots, E_n)$ be a binary or continuous environment vector and $\mathbf{X} = (X_1, \ldots, X_n)^T$ be the $n \times p$ matrix of high-dimensional data where $X_i = (X_{i1}, \ldots, X_{ij}, \ldots, X_{ip}) \in [0,1]^p$, and $\boldsymbol{\varepsilon} = (\varepsilon_1, \ldots, \varepsilon_n)$ a vector of errors. Consider the regression model with main effects and their interactions with E:

$$Y_i = \beta_0^* + \sum_{j=1}^{p} \beta_j^* X_{ij} + \beta_E^* E_i + \sum_{j=1}^{p} \alpha_j^* E_i X_j + \varepsilon_i, \qquad i = 1, \ldots, n, \tag{1}$$

where $\beta_0^*, \beta_j^*, \beta_E^*, \alpha_j^*$ are the true unknown model parameters for $j = 1, \ldots, p$. This can be extended to the more general additive model:

$$Y_i = \beta_0^* + \sum_{j=1}^{p} f_j^*(X_{ij}) + f_E^*(E_i) + \sum_{j=1}^{p} f_{jE}^*(X_{ij}, E_i) + \varepsilon_i \qquad i = 1, \ldots, n \tag{2}$$

As in (Radchenko and James, 2010), we can express (2) as

$$\mathbf{Y} = \sum_{j=1}^{p} \mathbf{f}_j^* + \mathbf{f}_E^* + \sum_{j=1}^{p} \mathbf{f}_{jE}^* + \boldsymbol{\varepsilon} \tag{3}$$

where $\mathbf{f}_j^* = \left(f_j^*(X_{1j}), \ldots, f_j^*(X_{nj})\right)^T$, $\mathbf{f}_{jE}^* = \left(f_{jE}^*(X_{1j}, X_{1E}), \ldots, f_j^*(X_{nj}, X_{nE})\right)^T$ and $\mathbf{f}_E^* = f_E^*(E_i)$. We consider the candidate vectors $\{\mathbf{f}_j, \mathbf{f}_E, \mathbf{f}_{jE}\}$. The general approach for fitting (3) is to minimize the following penalized regression criterion:

$$\frac{1}{2}||\mathbf{Y} - \mathbf{f}||^2 + P(\mathbf{f}) \tag{4}$$

where

$$\mathbf{f} = \sum_{j=1}^{p} \mathbf{f}_j + \mathbf{f}_E + \sum_{j=1}^{p} \mathbf{f}_{jE} \tag{5}$$

and $P(\mathbf{f})$ is a penalty function on $\mathbf{f}$

The smoothing method for variable X_j is a projection on to a set of basis functions. Consider

$$f_j(\cdot) = \sum_{\ell=1}^{p_j} \psi_{j\ell}\beta_{j\ell} \tag{6}$$

where the $\{\psi_{j\ell}\}_1^{p_j}$ are a family of basis functions in X_j (Hastie et al., 2015).

Bibliography

Peter Radchenko and Gareth M James. Variable selection using adaptive nonlinear interaction structures in high dimensions. *Journal of the American Statistical Association*, 105 (492):1541–1553, 2010. 1

Trevor Hastie, Robert Tibshirani, and Martin Wainwright. *Statistical Learning with Sparsity: The Lasso and Generalizations*. CRC Press, 2015. 2