Variable Selection in Nonlinear Interactions with the Group Lasso

Sahir Rai Bhatnagar

SSC 2017

sahirbhatnagar.com, @syfi_24

McGill University

Supervisors

Yi Yang

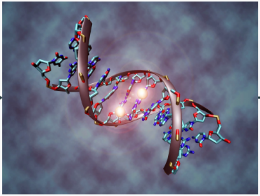
Celia Greenwood

Post-doc Opportunities with Celia in Statistical Genetics

Robert Platt @robertwplatt · 13h
Setting up the McGill Biostat booth at #SSC2017WPG @McGillEBOH
pic.twitter.com/VSUsV7cBhZ

Motivation

Interactions with Environment



Environment
Gestational Diabetes

Large Data
Child's epigenome
($p \approx 450\mathrm{k}$)

Phenotype
Obesity measures

Variable Selection in Interaction Models

Classic Interaction Model

- $Y \rightarrow$ response
- $X_E \rightarrow$ environment
- $X_j \rightarrow$ predictors, $j = 1, \ldots, p$

$$Y = \beta_0 \cdot \mathbf{1} + \sum_{j=1}^{p} \beta_j X_j + \beta_E X_E + \sum_{j=1}^{p} \alpha_j X_E X_j + \varepsilon$$

Heredity Property[1]

$$Y = \beta_0 \cdot \mathbf{1} + \sum_{j=1}^{p} \beta_j X_j + \beta_E X_E + \sum_{j=1}^{p} \alpha_j X_E X_j + \varepsilon$$

Strong Hierarchy

$$\hat{\alpha}_j \neq 0 \qquad \Rightarrow \qquad \hat{\beta}_j \neq 0 \quad \text{and} \quad \hat{\beta}_E \neq 0$$

Weak Hierarchy

$$\hat{\alpha}_j \neq 0 \qquad \Rightarrow \qquad \hat{\beta}_j \neq 0 \quad \text{or} \quad \hat{\beta}_E \neq 0$$

[1] Chipman, 1996, *Canadian Journal of Statistics*

Hierarchical Interactions: Current State of the Art

Type	Model	Software
Linear	`CAP` (Zhao et al. 2009, *Ann. Stat*)	✗
	`SHIM` (Choi et al. 2009, *JASA*)	✗
	`hiernet` (Bien et al. 2013, *Ann. Stat*)	`hierNet(x, y)`
	`GRESH` (She and Jiang 2014, *JASA*)	✗
	`FAMILY` (Haris et al. 2014, *JCGS*)	`FAMILY(x, z, y)`
	`glinternet` (Lim and Hastie 2015, *JCGS*)	`glinternet(x, y)`
	`RAMP` (Hao et al. 2016, *JASA*)	`RAMP(x, y)`
Non-linear	`VANISH` (Radchenko and James 2010, *JASA*)	✗
	`funshim` (Bhatnagar et al. 2017+)	`funshim(x, e, y)`

Lasso interaction model

$$Y = \beta_0 \cdot \mathbf{1} + \sum_{j=1}^{p} \beta_j X_j + \beta_E X_E + \sum_{j=1}^{p} \alpha_j X_E X_j + \varepsilon$$

$$\underset{\beta_0, \boldsymbol{\beta}, \boldsymbol{\alpha}}{\operatorname{argmin}} \quad \mathcal{L}(Y; \Theta) + \lambda(\|\boldsymbol{\beta}\|_1 + \|\boldsymbol{\alpha}\|_1)$$

Reparametrization[2]

Reparametrization

$$\alpha_j = \gamma_j \beta_j \beta_E$$

Model

$$Y = \beta_0 \cdot \mathbf{1} + \sum_{j=1}^{p} \beta_j X_j + \beta_E X_E + \sum_{j=1}^{p} \gamma_j \beta_j \beta_E X_E X_j + \varepsilon$$

Objective Function

$$\underset{\beta_0, \boldsymbol{\beta}, \boldsymbol{\gamma}}{\arg\min} \quad \mathcal{L}(Y; \Theta) + \lambda_\beta \sum_{j=1}^{p} w_j |\beta_j| + \lambda_\gamma \sum_{j=1}^{p} w_{jE} |\gamma_{jE}|$$

---

[2] Choi et al. 2010, **JASA**

`funshim`: An Extension to Nonlinear Effects

Basis Expansion

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

$$f(X_1) = \underbrace{\begin{bmatrix} \psi_{11}(X_{11}) & \psi_{12}(X_{12}) & \cdots & \psi_{11}(X_{15}) \\ \vdots & \vdots & \cdots & \vdots \\ \vdots & \vdots & \cdots & \vdots \\ \psi_{11}(X_{i1}) & \psi_{12}(X_{i2}) & \cdots & \psi_{11}(X_{i5}) \\ \vdots & \vdots & \cdots & \vdots \\ \vdots & \vdots & \cdots & \vdots \\ \psi_{11}(X_{N1}) & \psi_{12}(X_{N2}) & \cdots & \psi_{11}(X_{N5}) \end{bmatrix}_{N\times 5}}_{\boldsymbol{\Psi}_1} \times \underbrace{\begin{bmatrix} \beta_{11} \\ \beta_{12} \\ \beta_{13} \\ \beta_{14} \\ \beta_{15} \end{bmatrix}_{5\times 1}}_{\theta_1}$$

funshim

- $\theta_j = (\beta_{j1}, \ldots, \beta_{jp_j}) \in \mathbb{R}^{p_j}$
- $\alpha_j = (\alpha_{j1}, \ldots, \alpha_{jp_j}) \in \mathbb{R}^{p_j}$
- $\Psi_j \rightarrow n \times p_j$ matrix of evaluations of the $\psi_{j\ell}$

Model

$$Y = \beta_0 \cdot \mathbf{1} + \sum_{j=1}^{p} \Psi_j \theta_j + \beta_E X_E + \sum_{j=1}^{p} X_E \Psi_j \alpha_j + \varepsilon$$

Reparametrization

$$\alpha_j = \gamma_j \beta_E \theta_j$$

Model

$$Y = \beta_0 \cdot \mathbf{1} + \sum_{j=1}^{p} \Psi_j \theta_j + \beta_E X_E + \sum_{j=1}^{p} \gamma_j \beta_E X_E \Psi_j \theta_j + \varepsilon$$

Objective Function

$$\underset{\beta_E, \boldsymbol{\theta}, \boldsymbol{\gamma}}{\operatorname{argmin}} \quad \mathcal{L}(Y; \Theta) + \lambda_\beta \left(w_E |\beta_E| + \sum_{j=1}^{p} w_j \|\theta_j\|_2 \right) + \lambda_\gamma \sum_{j=1}^{p} w_{jE} |\gamma_j|$$

Algorithm

Algorithm 1: Block Relaxation Algorithm

Set the iteration counter $k \leftarrow 0$, initial values for the parameter vector $\Theta^{(0)}$;

for *each pair* $(\lambda_\beta, \lambda_\gamma)$ **do**

repeat

$$\boldsymbol{\gamma}^{(k+1)} \leftarrow \underset{\boldsymbol{\gamma}}{\operatorname{argmin}} \quad Q_{\lambda_\beta,\lambda_\gamma}\left(\boldsymbol{\gamma}, \beta_E^{(k)}, \boldsymbol{\theta}^{(k)}\right)$$

$$\boldsymbol{\theta}^{(k+1)} \leftarrow \underset{\boldsymbol{\theta}}{\operatorname{argmin}} \quad Q_{\lambda_\beta,\lambda_\gamma}\left(\boldsymbol{\theta}, \beta_E^{(k)}, \boldsymbol{\gamma}^{(k+1)}\right)$$

$$\beta_E^{(k+1)} \leftarrow \underset{\boldsymbol{\beta}_E}{\operatorname{argmin}} \quad Q_{\lambda_\beta,\lambda_\gamma}\left(\boldsymbol{\theta}^{(k+1)}, \beta_E, \boldsymbol{\gamma}^{(k+1)}\right)$$

$k \leftarrow k + 1$

until *convergence criterion is satisfied*;

end

Objective Function

$$\underset{\beta_E, \boldsymbol{\theta}, \boldsymbol{\gamma}}{\operatorname{argmin}} \quad \mathcal{L}(Y; \Theta) + \lambda_\beta \left(w_E |\beta_E| + \sum_{j=1}^{p} w_j \|\theta_j\|_2 \right) + \lambda_\gamma \sum_{j=1}^{p} w_{jE} |\gamma_j|$$

Lasso problem (`glmnet`, Friedman, Hastie & Tibshirani 2010)

$$\underset{\boldsymbol{\gamma}}{\operatorname{argmin}} \quad \mathcal{L}(Y; \Theta) + \lambda_\beta \left(w_E |\beta_E| + \sum_{j=1}^{p} w_j \|\theta_j\|_2 \right) + \lambda_\gamma \sum_{j=1}^{p} w_{jE} |\gamma_j|$$

[3] https://github.com/sahirbhatnagar/funshim

Implementation[4]

Objective Function

$$\underset{\beta_E, \boldsymbol{\theta}, \boldsymbol{\gamma}}{\operatorname{argmin}} \quad \mathcal{L}(Y; \Theta) + \lambda_\beta \left(w_E |\beta_E| + \sum_{j=1}^{p} w_j \|\theta_j\|_2 \right) + \lambda_\gamma \sum_{j=1}^{p} w_{jE} |\gamma_j|$$

Group Lasso problem (`gglasso`, Yang and Zou 2015)

$$\underset{\beta_E, \boldsymbol{\theta}}{\operatorname{argmin}} \quad \mathcal{L}(Y; \Theta) + \lambda_\beta \left(w_E |\beta_E| + \sum_{j=1}^{p} w_j \|\theta_j\|_2 \right) + \lambda_\gamma \sum_{j=1}^{p} w_{jE} |\gamma_j|$$

[4] https://github.com/sahirbhatnagar/funshim

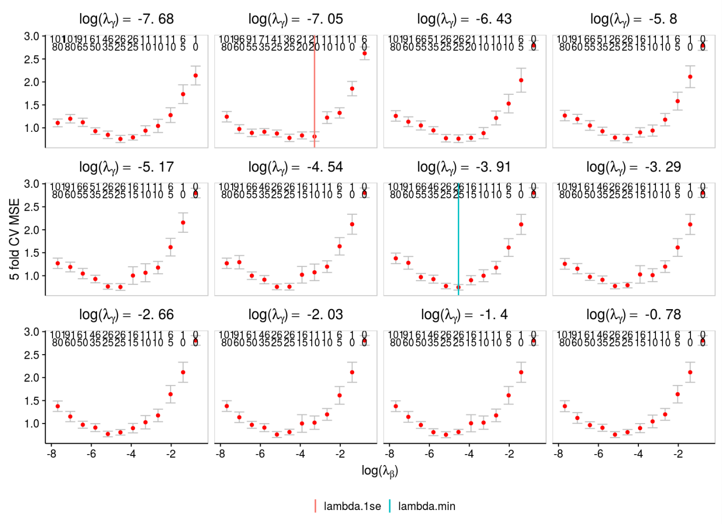
Simulations

Scenario 1

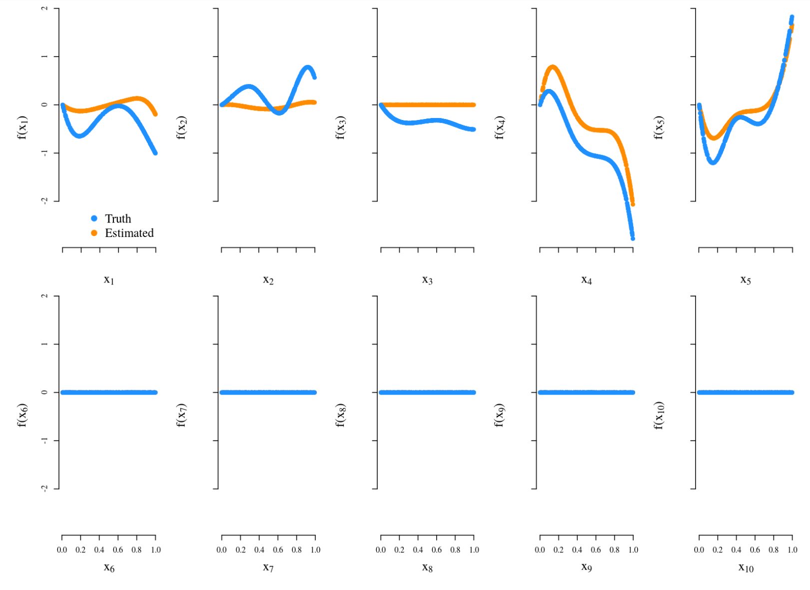
- $Y = \sum_{j=1}^{5} f(X_j) + X_E + E(f(X_1) + f(X_2))$
- $f(\cdot) \rightarrow$ B-splines with 5 df
- $\theta_j \sim \mathcal{N}(0, 1)$
- $N = 400, p = 50$
- $50 \times 5 \times 2 + 1 = 501$ parameters to estimate

Scenario 1: Cross-validation results

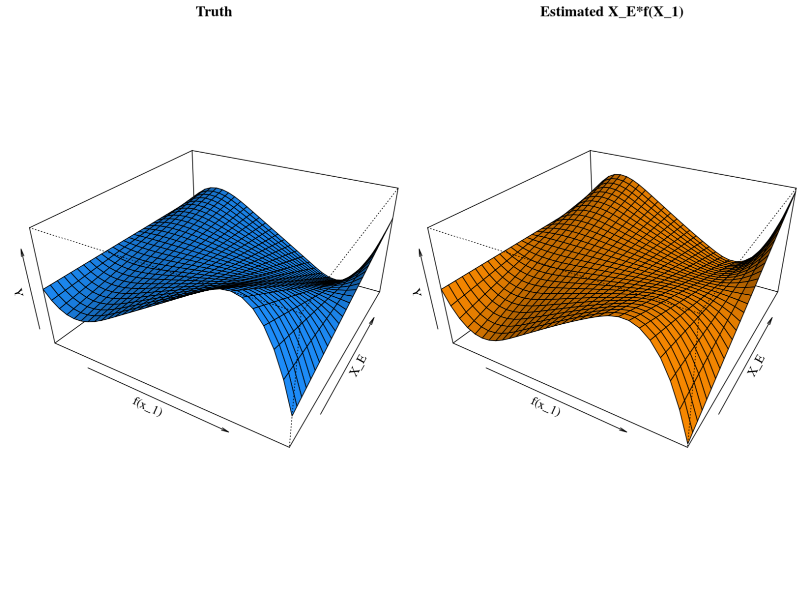
```r
funshim::plot(cvfit)
```



Scenario 1: Main Effects



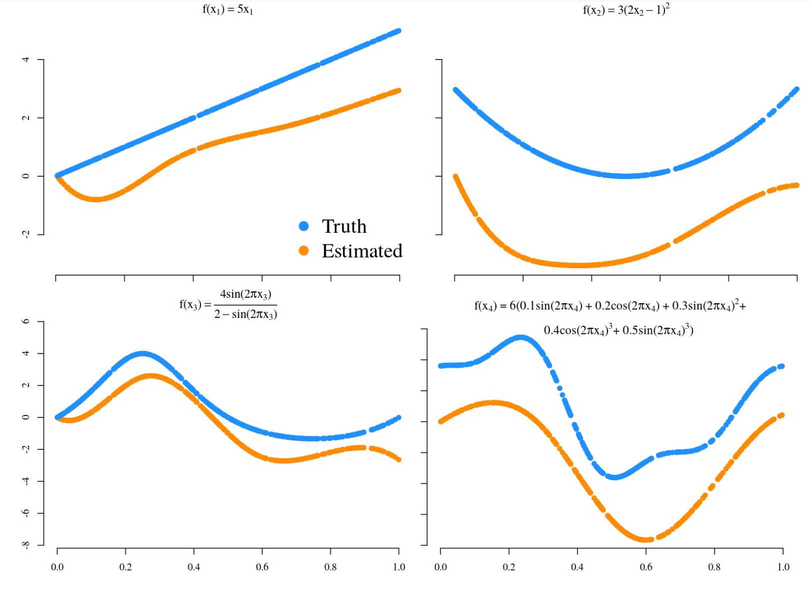
Scenario 1: Interaction Effects



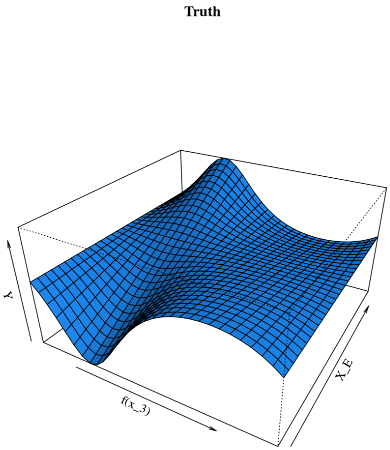
Scenario 2

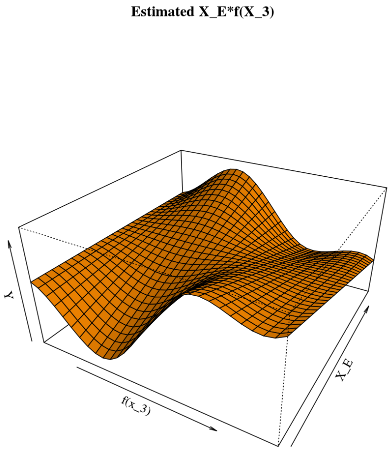
- $Y = \sum_{j=1}^{4} f(X_j) + X_E + E(f(X_3) + f(X_4))$
- $f(X_1) \rightarrow$ linear
- $f(X_2) \rightarrow$ quadratic
- $f(X_3) \rightarrow$ sinusoidal
- $f(X_4) \rightarrow$ complicated sinusoidal

Scenario 2: Main Effects

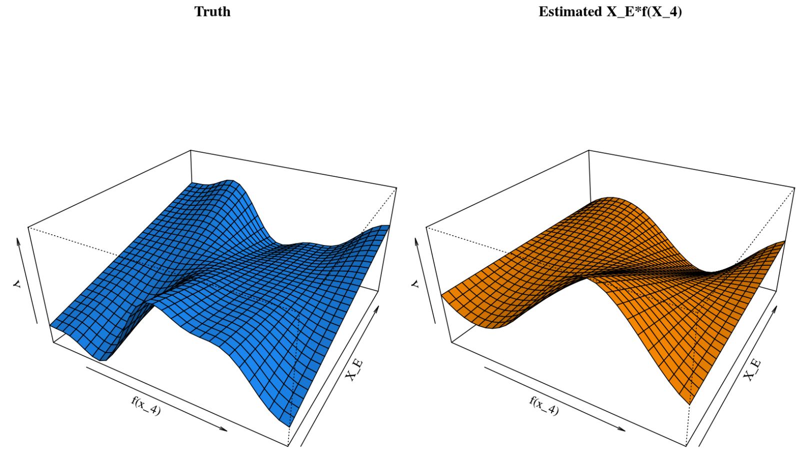


Scenario 2: Interaction Effects





Scenario 2: Interaction Effects



Discussion

Strengths and Limitations

Strengths

- Environment interactions with strong heredity property in $p >> N$
- `funshim` allows for flexible modeling of input variables (imaging, gene expression, DNA methylation)
- R package provided with tuning parameter selection routines

Limitations

- Can only handle $E \cdot f(X)$ or $f(E) \cdot X$
- Does not allow for $f(X_1, E)$ or $f(X_1, X_2)$
- Current implementation is slow due to cross validation for 2 tuning parameters

Future Directions

- Are two tuning parameters really necessary ?

$$\lambda \left\{ (1-\alpha) \left[w_E|\beta_E| + \sum_{j=1}^{p} w_j \|\theta_j\|_2 \right] + \alpha \sum_{j=1}^{p} w_{jE}|\gamma_j| \right\}$$

- Weak heredity property $\rightarrow \alpha_j = \gamma_j(|\beta_j| + |\beta_E|)$
- Numerical convergence and KKT checks
- Information Criterion instead of CV for tuning parameters
- Extension to GLM
- User defined interactions
- Non-parametric screening prior to model fitting
- Real data analysis

Acknowledgements

CENTRE LUDMER CENTER
NEUROINFORMATIQUE & SANTÉ MENTALE | NEUROINFORMATICS & MENTAL HEALTH

Jewish General Hospital
Lady Davis Institute for Medical Research

McGill

References

- Radchenko, P., & James, G. M. (2010). Variable selection using adaptive nonlinear interaction structures in high dimensions. Journal of the American Statistical Association, 105(492), 1541-1553.
- Choi, N. H., Li, W., & Zhu, J. (2010). Variable selection with the strong heredity constraint and its oracle property. Journal of the American Statistical Association, 105(489), 354-364.
- Chipman, H. (1996). Bayesian variable selection with related predictors. Canadian Journal of Statistics, 24(1), 17-36.
- Friedman, J., Hastie, T., & Tibshirani, R. (2010). Regularization paths for generalized linear models via coordinate descent. Journal of statistical software, 33(1)
- Yang, Y., & Zou, H. (2015). A fast unified algorithm for solving group-lasso penalize learning problems. Statistics and Computing, 25(6), 1129-1141
- De Leeuw, J. (1994). Block-relaxation algorithms in statistics. In Information systems and data analysis (pp. 308-324). Springer Berlin Heidelberg.

sahirbhatnagar.com