A Comparison Between Two Algorithms in Subgroup Analysis

Bin Li
Advisor: Dr. Xin Wang
Committee Members: Dr.Jing Zhang; Dr.O'Connell

July 26, 2021

Miami University

Table of contents

Introduction

- **Subgroup analysis: widely used**
    1. Medical: develop individualized treatment strategies to achieve precision medicine (Ma and Huang, 2017).
    2. Environmental science: the spatial effect in a large region may change with location but tend to stay homogeneous within clusters (Li and Sang, 2019).

Introduction

- **Clustering analysis: identifying subgroups**
  1. Hierarchical clustering: looks for nested clusters either in agglomerative mode or in a divisive model.
  2. Partitional clustering: searches for clusters simultaneously as a partition of the data without considering any hierarchical structure, like K-means.
  3. Other clustering algorithms: use different objective functions, probabilistic generative models, and heuristics.(Jain, 2010).

Introduction

- **Clustering analysis: investigating the relationship between Y and X through regression analysis**
  1. Pairwise coefficient difference: penalize on coefficients to achieve homogeneity among coefficients, like concave pairwise fusion penalized least squares approach (Ma and Huang, 2017).
  2. Sparially clustered coefficient: construct regularization to incorporate spatial neighborhood information and capture clustered coefficients (Li and Sang, 2019).

Penaly Function

- **Penalty function selection**
  1. Smoothly Clipped Absolute Deviation Penalty (SCAD)
  2. Minimax Concave Penalty (MCP)
  3. Lasso Penalty (L1 Penalty)

The Model and the Algorithm

The model

- **Details about subgroups and regression analysis**
  1. Multiple measurements
  2. Two categories of covariates: "common" and "specific"
  3. Linear regression model considered in Wang et al. (2020)

$$y_{ih} = \boldsymbol{z}_{ih}^T \boldsymbol{\eta} + \boldsymbol{x}_{ih}^T \boldsymbol{\beta}_i + \epsilon_{ih}$$

where $\boldsymbol{y}_{ih}$: the hth observation for the ith subject for $i = 1, \ldots, n$ and $h = 1, \ldots, n_i$.

$\boldsymbol{z}_{ih}$: common covariates

$\boldsymbol{x}_{ih}$: specific covariates

$\boldsymbol{\eta}$: vector of common regression coefficients

$\boldsymbol{\beta}_i$: unit-specific regression coefficients

ϵ_{ih} : i.i.d random errors with $E(\epsilon_{ih}) = 0$ and $Var(\epsilon_{ih}) = \sigma^2$

Penalization

- **Penalization on pairwise coefficient difference**
  1. Vector penalty $(\boldsymbol{\beta}_i - \boldsymbol{\beta}_j)$ (Wang et al., 2020; Ma et al., 2019)
  2. Coordinate penalty $(\beta_{li} - \beta_{lj})$ (Li and Sang, 2019; Yang et al., 2019)

  An example:

$$\begin{gathered}
\boldsymbol{\beta}_1 = (\beta_{11}, \beta_{21}, \beta_{31})^T \\
\boldsymbol{\beta}_2 = (\beta_{12}, \beta_{22}, \beta_{32})^T \\
\boldsymbol{\beta}_3 = (\beta_{13}, \beta_{23}, \beta_{33})^T \\
\vdots \\
\boldsymbol{\beta}_n = (\beta_{1n}, \beta_{2n}, \beta_{3n})^T
\end{gathered}$$

- **Objective function**

$$Q_n(\boldsymbol{\eta}, \boldsymbol{\beta}; \lambda) = \frac{1}{2} \sum_{i=1}^{n} \frac{1}{n_i} \sum_{h=1}^{n_i} \left(y_{ih} - \boldsymbol{z}_{ih}^T \boldsymbol{\eta} - \boldsymbol{x}_{ih}^T \boldsymbol{\beta}_i \right)^2 + \boxed{\sum_{1 \le i < j \le n} p_\gamma \left(\|\boldsymbol{\beta}_i - \boldsymbol{\beta}_j\|, c_{ij}\lambda \right)}$$

where

$p_\gamma(\cdot, \lambda)$: penalty function

γ : built-in constant, $\gamma = 3$ (Fan and Li, 2001)

$\|\cdot\|$: Euclidean norm

c_{ij} : pairwise weights. In spatial data (Wang et al., 2020), c_{ij} can be defined based on locations. Here $c_{ij} = 1$ is considered in the simulation study.

λ : tuning parameter, $\lambda \geq 0$

Algorithm II Coordinate Penalty

- **Objective function**

$$Q(\boldsymbol{\eta}, \boldsymbol{\beta}; \lambda) = \frac{1}{2} \sum_{i=1}^{n} \frac{1}{n_i} \sum_{h=1}^{n_i} \left(y_{ih} - \boldsymbol{z}_{ih}^T \boldsymbol{\eta} - \boldsymbol{x}_{ih}^T \boldsymbol{\beta}_i \right)^2 + \boxed{\sum_{l=1}^{p} \sum_{1 \le i < j \le n} p_\gamma \left(\left| \beta_{li} - \beta_{lj} \right|, c_{ij} \lambda \right)},$$

where

$p_\gamma(\cdot, \lambda)$: penalty function

γ : built-in constant, $\gamma = 3$ (Fan and Li, 2001)

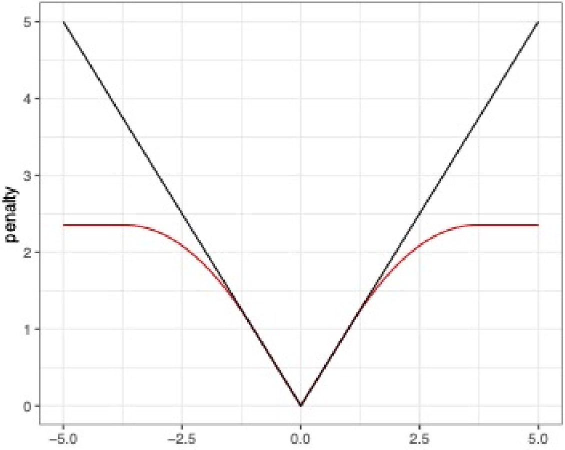
c_{ij} : pairwise weights. In spatial data (Wang et al., 2020), c_{ij} can be defined based on locations. Here $c_{ij} = 1$ is considered in the simulation study.

λ : tuning parameter, $\lambda \ge 0$

Penalty Function

- **The choice of the penalty function $p_\gamma(\cdot, \lambda)$**

  SCAD Penalty VS. Lasso Penalty



ADMM

- **Alternating direction method of multiplier algorithm**
  1. ADMM is used to compute the objective functions efficiently.
  2. `Spgr` package in R (`https://github.com/wangx23/Spgr`)

Simulation

Methods

- **Combinations of the two layer penalization**
  1. SCAD with the coordinate penalty (M1)
  2. SCAD with the vector penalty (M2)
  3. Lasso with the coordinate penalty (M3)

Data Simulation

- **Model used in simulation**

$$y_{ih} = \boldsymbol{z}_{ih}^T \boldsymbol{\eta} + \boldsymbol{x}_{ih}^T \boldsymbol{\beta}_i + \epsilon_{ih}$$

- **Data generation**
  1. $\boldsymbol{z}_{ih}$=$(z_{i,1}, z_{i,2}, z_{i,3}, z_{i,4}, z_{i,5})^T$
     1.1 $z_{ih,1} = 1$
     1.2 $(z_{ih,2}, ..., z_{ih,5})^T$: $MVN, \mu = 0, \sigma = 1, \rho = 0.3$
  2. $\boldsymbol{\eta} = (\eta_1, ..., \eta_5)^T$: $Uniform[1, 2]$
  3. x_{ih}: $N(0, 1)$; $Bin(n, 0.7)$
  4. ϵ_{ih}: $N(0, 0.5^2)$

Hyperparameters

- **Hyperparameters in penalty function** $p_\gamma(\cdot, \lambda)$
  1. $\gamma = 3$ (SCAD penalty)
  2. λ: chosen by BIC (SCAD and lasso penalty)

     2.1 SCAD and lasso with coordinate penalty (M1 and M3)

$$BIC\,(\lambda) = \log\left[\frac{1}{n}\sum_{i=1}^{n}\left(y_i - \boldsymbol{z}_i^T\hat{\boldsymbol{\eta}}(\lambda) - \boldsymbol{x}_i^T\hat{\boldsymbol{\beta}}_i(\lambda)\right)^2\right] + \boxed{C_n\frac{log\;\; n}{n}\left(\sum_{l=1}^{p}\hat{K}_l(\lambda) + q\right)}$$

     2.2 SCAD with vector penalty (M2)

$$BIC\,(\lambda) = \log\left[\frac{1}{n}\sum_{i=1}^{n}\frac{1}{n_i}\left(y_{ih} - \boldsymbol{z}_{ih}^T\hat{\boldsymbol{\eta}}(\lambda) - \boldsymbol{x}_{ih}^T\hat{\boldsymbol{\beta}}_i(\lambda)\right)^2\right] + \boxed{C_n\frac{log\;\; n}{n}\left(\hat{K}(\lambda)p + q\right)}$$

where $C_n = \log(\log(np + q))$

Criterion for Model Performance

- **Adjusted Rand Index (ARI)**
  1. The quantity ARI measures the degree of agreement between two partitions.(Rand, 1971; Vinh et al., 2010)
  2. ARI: [0, 1]
- **Root Mean Square Error (RMSE)**
  1. Average RMSE for coefficient coefficient

$$\sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{\beta}_{li}-\beta_{li})^2}.$$

  2. Average RMSE for vector coefficient

$$\sqrt{\frac{1}{n}\sum_{i=1}^{n}\|\hat{\boldsymbol{\beta}}_i-\boldsymbol{\beta}_i\|^2}.$$

Balanced Groups I

- **Simulation scenarios**

Parameter	Test 1	Test 2	Test 3	Test 4
n	100	200	100	200
n_i	2	2	5	5
β_{1i}	(-1,1,2)	(-1,1,2)	(-1,1,2)	(-1,1,2)
β_{2i}	(-1,1,2)	(-1,1,2)	(-1,1,2)	(-1,1,2)
β_{3i}	(-1,1,2)	(-1,1,2)	(-1,1,2)	(-1,1,2)

Balanced Groups II

- **Simulation results**

1. Average ARI for coordinate coefficients

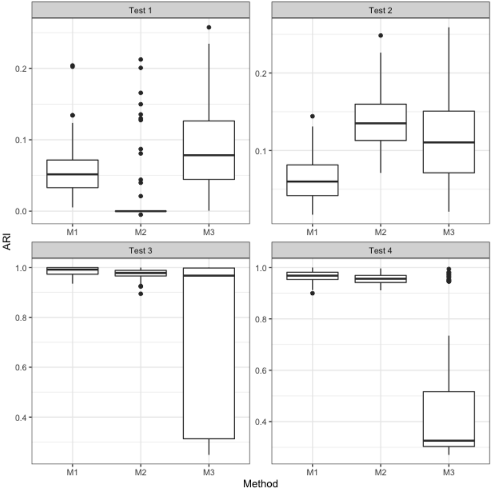
Test	M1: β_1	M1: β_2	M1: β_3	M3: β_1	M3: β_2	M3: β_3
Test 1	0.196	0.196	0.199	0.217	0.204	0.225
Test 2	0.195	0.194	0.197	0.247	0.248	0.262
Test 3	0.99	0.988	0.989	0.824	0.816	0.821
Test 4	0.977	0.977	0.978	0.665	0.676	0.667

2. Average RMSE for coordinate coefficients

Test	M1: β_1	M1: β_2	M1: β_3	M3: β_1	M3: β_2	M3: β_3
Test 1	0.761	0.756	0.750	0.905	0.914	0.894
Test 2	0.756	0.759	0.751	0.824	0.827	0.807
Test 3	0.037	0.040	0.041	0.189	0.188	0.185
Test 4	0.076	0.073	0.071	0.335	0.326	0.33

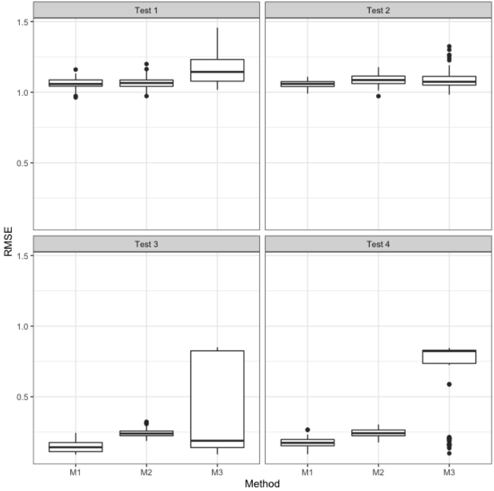
Balanced Groups III

- **ARI for vector coefficients**



Balanced Groups IV

- **RMSE for vector coefficients**



Unbalanced Groups I

- **Simulation scenarios**

Parameter	Test 1	Test 2
n	100	200
n_i	5	5
β_{1i}	(1,1,1)	(1,1,1)
β_{2i}	(-1,1,-1)	(-1,1,-1)
β_{3i}	(-1,1,2)	(-1,1,2)

Unbalanced Groups II

- **Simulation results**

1. Average ARI for coordinate coefficients

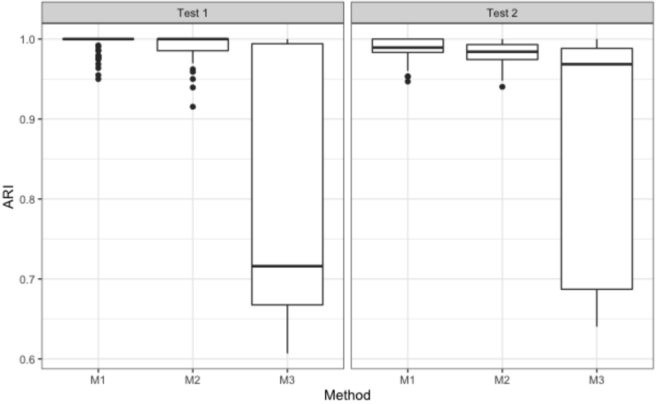
Test	M1: β_1	M1: β_2	M1: β_3	M3: β_1	M3: β_2	M3: β_3
Test 1	0.999	0.998	0.996	0.998	0.998	0.734
Test 2	0.998	0.994	0.988	0.999	0.993	0.812

2. Average RMSE for coordinate coefficients

Test	M1: β_1	M1: β_2	M1: β_3	M3: β_1	M3: β_2	M3: β_3
Test 1	0.007	0.01	0.02	0.012	0.02	0.256
Test 2	0.007	0.02	0.049	0.009	0.024	0.204

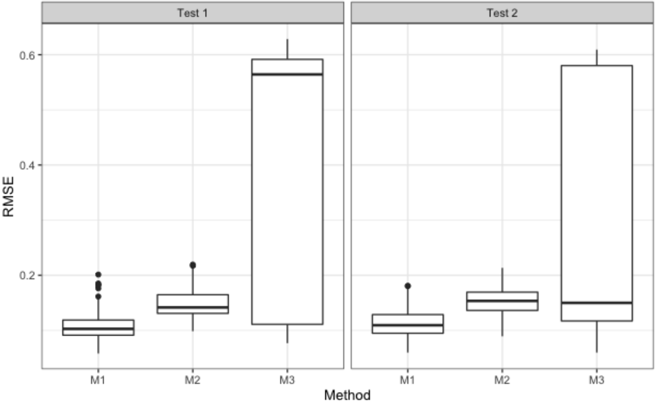
Unbalanced Groups III

- **ARI for vector coefficients**



Unbalanced Groups IV

- **RMSE for vector coefficients**



Empirical Example

Covid-19 Dataset

- **Dataset**
  1. Response variable: monthly case confirmed rate in 48 continental states
  2. Common effect covariates: state party; household average income
  3. Unit-specfic effect covariates: vaccine completion rate; unemployment rate
  4. Replicates: monthly rate from Jan 2021 to April 2021

- **Data source**
  1. the U.S. Bureau of Labor Statistics
  2. U.S. Census Bureau
  3. `covid19.analysis` package in R

Algorithm Applied

- **Algorithm**

  SCAD with coordinate penalty (M1)

- **Weights**

  1. Equal weights: $c_{ij} = 1$
  2. Unequal weights:

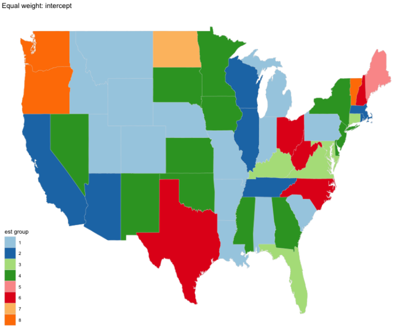
$$c_{ij} = \exp(\psi(1 - a_{ij}))$$

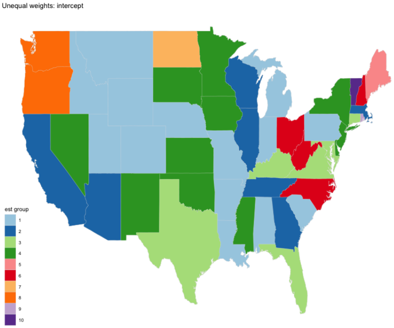
where ψ: tuning parameter to be chosen using BIC

a_{ij}: neighborhood order defined based on the neighborhood structure

Application Results I

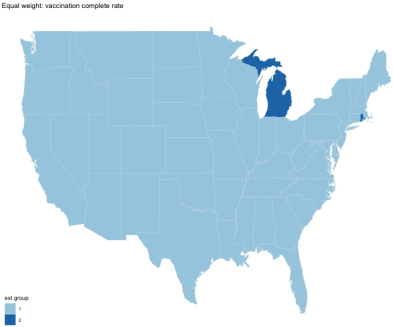
- **Estimated group for intercept**

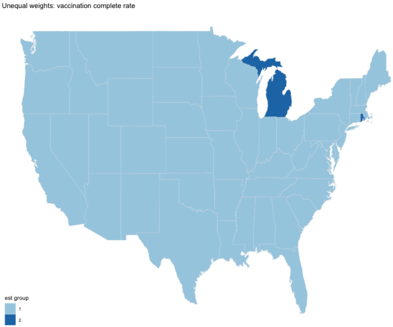




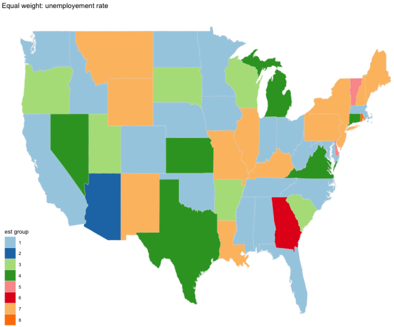
Application Results II

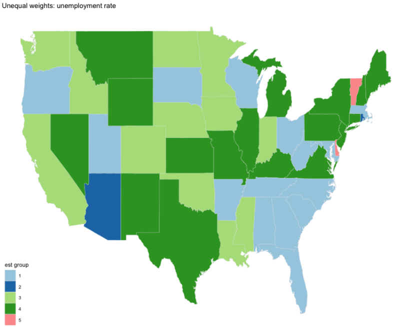
- **Estimated group for vaccine completion rate**





- **Estimated group for unemployment rate**





Discussion

Conclusion

- **Conclusion**
  1. Tests with more replicates tend to bring more desirable estimates.
  2. Lasso penalty performs the worst when dealing with situations with more groups.
  3. SCAD with the coordinate penalty (M1) performs the best when enough replicates are given.

- **Discussion**
  1. That SCAD with vector penalty (M2) performs worse than SCAD with coordinate penalty (M1) may be caused by that fewer observations would be assigned to each group under M2.
  2. Coordinate penalty could be easy to extend to variable selection problems.

References

Fan, J. and Li, R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American statistical Association*, 96(456):1348–1360.

Jain, A. K. (2010). Data clustering: 50 years beyond k-means. *Pattern recognition letters*, 31(8):651–666.

Li, F. and Sang, H. (2019). Spatial homogeneity pursuit of regression coefficients for large datasets. *Journal of the American Statistical Association*, 114(527):1050–1062.

Ma, S. and Huang, J. (2017). A concave pairwise fusion approach to subgroup analysis. *Journal of the American Statistical Association*, 112(517):410–423.

Ma, S., Huang, J., Zhang, Z., and Liu, M. (2019). Exploration of heterogeneous treatment effects via concave fusion. *The international journal of biostatistics*, 16(1).

Rand, W. M. (1971). Objective criteria for the evaluation of clustering methods. *Journal of the American Statistical association*, 66(336):846–850.

Vinh, N. X., Epps, J., and Bailey, J. (2010). Information theoretic measures for clusterings comparison: Variants, properties, normalization and correction for chance. *The Journal of Machine Learning Research*, 11:2837–2854.

Wang, X., Zhu, Z., and Zhang, H. H. (2020). Spatial heterogeneity automatic detection and estimation.

Yang, X., Yan, X., and Huang, J. (2019). High-dimensional integrative analysis with homogeneity and sparsity recovery. *Journal of Multivariate Analysis*, 174:104529.

Thanks