Outline
oo

Background
o
ooooo
ooooo
ooooo

Dissertation
ooo
ooooo
ooooo
ooooo
ooooo
ooooo

Future Research
o
oooo

Conclusion
oo

References
oo

Spike-and-Slab Additive Models And Fast Algorithms For High-Dimensional Data Analysis

Boyi Guo

Department of Biostatistics
University of Alabama at Birmingham

July 12th, 2022

Outline
○○

Background
○
○○○○○
○○○○○
○○○○○

Dissertation
○○○
○○○○○
○○○○○
○○○○○
○○○○○

Future Research
○
○○○○

Conclusion
○○

References
○○

Dissertation Committee

- ▶ Chair: Nengjun Yi, Ph.D.
- ▶ Member (in alphabetical order):
  - ▶ AKM Fazlur Rahman, Ph.D.
  - ▶ Byron C. Jaeger, Ph.D.
  - ▶ D. Leann Long, Ph.D.
  - ▶ Michael E. Seifert, M.D.

Outline

Background

Dissertation

Future Research

Conclusion

References

Outline

Outline

- ▶ Background
  - ▶ Spline Model Development
  - ▶ Bayesian Regularization
  - ▶ Bayesian Variable Selection
- ▶ Dissertation
  - ▶ Two-part Spike-and-slab LASSO Prior for Spline Functions
  - ▶ EM-Coordinate Descent Algorithms
  - ▶ Empirical Performance of Prediction & Selection
- ▶ Future Research
  - ▶ Structured Additive Regression with Spike-and-Slab LASSO prior
  - ▶ Spatially Variable Genes Screening
  - ▶ Other Questions of Interest

Outline
oo

Background
●ooooo
ooooo
ooooo

Dissertation
ooo
ooooo
ooooo
ooooo
ooooo
ooooooo

Future Research
o
oooo

Conclusion
oo

References
oo

Background

Outline
○○

Background
●○○○○
○○○○○
○○○○○

Dissertation
○○○
○○○○○
○○○○○
○○○○○
○○○○○
○○○○○

Future Research
○
○○○○

Conclusion
○○

References
○○

Spline Model Development

Spline Model Development

Spline Model Development

Spline Model Development

"It is extremely unlikely that the true (effect) function f(X) (on the outcome) is actually linear in X."

— Hastie, Tibshirani, and Friedman (2009) PP. 139

Question

How to model nonlinear effects?

Spline Model Development

Spline Functions

A *spline* function is a piece-wise polynomial function

$$B(x) = \sum_{k=1}^{K} \beta_k b_k(x) \equiv \boldsymbol{X}^T \boldsymbol{\beta}$$

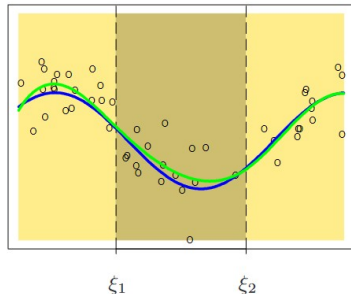$b_k(x)$ are the *basis functions*, possibly truncated power basis and b-spline basis.

Figure 1: A cubic spline function with 2 knots (courtesy of Hastie, Tibshirani, and Friedman (2009))

Outline	Background	Dissertation	Future Research	Conclusion	References
○○	○○○●○	○○○	○	○○	○○
	○○○○○	○○○○○	○○○○		
	○○○○○	○○○○○			
		○○○○○			
		○○○○○			
		○○○○○			

Spline Model Development

Generalized Additive Models with Splines

Generalized additive model (Hastie and Tibshirani 1987) is expressed

$$y_i \overset{\text{iid}}{\sim} EF(\mu_i, \phi), \quad i = 1, \ldots, n$$
$$g(\mu_i) = \beta_0 + B(x_i) = \beta_0 + \boldsymbol{X}_i^T \boldsymbol{\beta}, \quad \mathbb{E}[B(X)] = 0$$

where $B(x_i)$ is the spline function, $g(\cdot)$ is a link function, ϕ is the dispersion parameter

▶ Model fitting follows the generalized linear models, e.g. ordinary least square for Gaussian outcome

$$\hat{\boldsymbol{\beta}} = \arg\min \sum_{i=1}^{n} \left[y_i - \beta_0 - \boldsymbol{X}_i^T \boldsymbol{\beta} \right]^2$$

Outline
○○

Background
○○○○●
○○○○○
○○○○○

Dissertation
○○○
○○○○○
○○○○○
○○○○○
○○○○○
○○○○○

Future Research
○
○○○○

Conclusion
○○

References
○○

Spline Model Development

Problem: Function Smoothness

The estimation of $B(X)$ can be wiggly when the underlying function is smooth, particularly as the number of bases ,K, increases.

[TODO: add two plots, overfitting and not overfitting]

Question

How to balance function smoothness and complexity?

Outline
OO

Background

Dissertation

Future Research

Conclusion
OO

References
OO

Bayesian Regularization

Outline   Background   Dissertation   Future Research   Conclusion   References
OO   OOOOO   OOO   OOOO   OO   OO
                 OO●OO   OOOOO
                 OOOOO   OOOOO
                                     OOOOO
                                     OOOOO
                                     OOOOO

Bayesian Regularization

Smoothing Spline Model

▶ Smoothing penalty $\lambda \int B''(X)^2 dx = \lambda \boldsymbol{\beta}^T \boldsymbol{S} \boldsymbol{\beta}$

    ▶ The smoothing penalty matrix $\boldsymbol{S}$ is known given $\boldsymbol{X}$

    ▶ $\boldsymbol{S}$ is symmetric and positive semi-definite

▶ Penalized Least Square for Gaussian Outcome

$$\hat{\boldsymbol{\beta}} = \arg \min \sum_{i=1}^{n} \sum_{i=1}^{n} \left[y_i - \beta_0 - \boldsymbol{X}_i^T \boldsymbol{\beta} \right]^2 + \lambda \boldsymbol{\beta}^T \boldsymbol{S} \boldsymbol{\beta}$$

▶ The smoothing parameter λ is a tuning parameter, selected via cross-validation

Outline	Background	Dissertation	Future Research	Conclusion	References
oo	ooooo ooo●o ooooo	ooo ooooo ooooo ooooo ooooo ooooo	o oooo	oo	oo

Bayesian Regularization

Problem: Multiple Predictor Model

When a model contains multiple spline functions for variables $X_1, \ldots, X_p$, the penalized least square estimator is

$$\hat{\boldsymbol{\beta}} = \arg\min \sum_{i=1}^{n} \sum_{i=1}^{n} \left[y_i - \beta_0 - \sum \boldsymbol{X}_{ij}^T \beta_j \right]^2 + \lambda_j \beta_j^T \boldsymbol{S}_j \beta_j$$

How to decide λ_i?

▶ Global smoothing, i.e. $\lambda_1 = \cdots = \lambda_p$ assumes all functions shares the same shape
▶ Adaptive smoothing, i.e. examining λ_i combination, are computationally intensive

Outline
OO

Background

Dissertation

Future Research

Conclusion
OO

References
OO

Bayesian Regularization

Bayesian Regularization

- ▶ Bayesian Regularization is the Bayesian analogy of penalized models by using regularizing priors
  - ▶ Bayesian ridge via normal prior

$$\beta \sim N(0, \tau^2) \rightarrow \lambda = \sigma^2/\tau^2$$

- ▶ Adaptive shrinkage with hierarchical priors

$$\tau_j^2 \overset{\text{iid}}{\sim} IG(a, b)$$

  - ▶ Adaptive Smoothing
    - ▶ Random walk prior on b-spline bases with IG hyperprior
    - ▶ Normal prior on truncated power bases with a log-normal spline model for variance

Outline
OO

Background
O
OOOOO
OOOOO
●OOOO

Dissertation
OOO
OOOOO
OOOOO
OOOOO
OOOOO
OOOOO

Future Research
O
OOOO

Conclusion
OO

References
OO

Bayesian Variable Selection

Bayesian Variable Selection

Outline
oo

Background
o
ooooo
ooooo
oo●ooo

Dissertation
o
ooooo
ooooo
ooooo
ooooo
ooooo

Future Research
o
oooo

Conclusion
oo

References
oo

Bayesian Variable Selection

Problem: Functional Selection

In the context of variable selection and high-dimensional statistics, we always assume some variables are not effective or predictive to the outcome.

How to statistically detect

- if a variable is predictive to the outcome, $B_j(X_j) = 0$
- if a variable has a nonlinear relationship with the outcome, $B_j(X_j) = \beta_j X_j$

Bi-level selection is the procedure that simultaneously addresses the two questions above

Outline
○○

Background
○
○○○○○
○○○○○
○○●○○

Dissertation
○○○
○○○○○
○○○○○
○○○○○
○○○○○
○○○○○

Future Research
○
○○○○

Conclusion
○○

References
○○

Bayesian Variable Selection

Spike-and-Slab Priors

Spike-and-slab priors are a family of mixture distributions that deploys a characterizing structure

$$\beta|\gamma \sim (1-\gamma)f_{spike}(\beta) + \gamma f_{slab}(\beta)$$

▶ Latent indicator γ follows a Bernoulli distribution with probability θ
▶ Spike density $f_{spike}(x)$ concentrates around 0 for small effects
▶ Slab density $f_{slab}(x)$ is a flat density for large effects
▶ Natural procedure to select variables via posterior distribution of γ
▶ Markov chain Monte Carlo is not compelling for high-dimensional data analysis

Outline
○○

Background
○
○○○○○
○○○○○
○○○●○

Dissertation
○○○
○○○○○
○○○○○
○○○○○
○○○○○

Future Research
○
○○○○

Conclusion
○○

References
○○

Bayesian Variable Selection

Spike-and-Slab LASSO Priors

▶ Double exponential distributions as the spike and slab distributions

$$\beta|\gamma \sim (1-\gamma)DE(0, s_0) + \gamma DE(0, s_1), 0 < s_0 < s_1$$

   ▶ Seamless variable selection as coefficients shrinkage to 0
   ▶ Computation advantages via Expectation-Maximization (EM) algorithms

▶ Group spike-and-slab LASSO

   ▶ Structure underlying predictors, e.g. gene pathways, bases of a spline function
   ▶ Structured prior on γ

$$\gamma_k|\theta_j \ Binomial(1, \theta_j), k \in j$$

Bayesian Variable Selection

Problem: High-dimensional Spline Model

How to jointly model signal sparsity and function smoothness, while capable of bi-level selection?

▶ Excess shrinkage due to ignoring smooth penalty completely
  ▶ Group lasso penalty (Ravikumar et al. 2009; Huang, Horowitz, and Wei 2010), group SCAD penalty (Wang, Chen, and Li 2007; Xue 2009)
  ▶ Global penalty VS adaptive penalty
▶ All-in-all-out selection
  ▶ Can not detect if a function is linear, e.g. spike-and-slab grouped LASSO prior (Bai et al. 2020; Bai 2021)
  ▶ Failed to select function as whole, e.g. group spike-and-slab LASSO prior
▶ Computational prohibitive algorithms
  ▶ MCMC algorithms doesn't scale well for high-dimensional models (Scheipl, Fahrmeir, and Kneib 2012)

Outline
○○

Background
○
○○○○○
○○○○○
○○○○○

Dissertation
●○○
○○○○○
○○○○○
○○○○○
○○○○○
○○○○○

Future Research
○
○○○○

Conclusion
○○

References
○○

Dissertation

Outline
OO

Background
O
OOOOO
OOOOO
OOOOO

Dissertation
O
O●O
OOOOO
OOOOO
OOOOO
OOOOO

Future Research
O
OOOO

Conclusion
OO

References
OO

Objectives

▶ To develop statistical models that improve curve interpolation and outcome prediction
  ▶ Local adaption of sparse penalty and smooth penalty
  ▶ Bi-level selection for linear and nonlinear effect
▶ To develop a fast and scalable algorithm
▶ To implement a user-friendly statistical software

Projects

- ▶ **Guo, B.**, Jaeger, B. C., Rahman, A. F., Long, D. L., Yi, N. (2022). Spike-and-Slab least absolute shrinkage and selection operator generalized additive models and scalable algorithms for high-dimensional data analysis. *Statistics in Medicine*. doi: https://doi.org/10.1002/sim.9483

- ▶ **Guo, B.**, Jaeger, B. C., Rahman, A. F., Long, D. L., Yi, N. (2022). A scalable and flexible Cox proportional hazard model for high-dimensional survival prediction and functional selection *arXiv*. doi: https://doi.org/10.48550/arXiv.2205.11600

- ▶ **Guo, B.**, Yi, N. (2022). BHAM: An R Package to Fit Bayesian Hierarchical Additive Models for High-dimensional Data Analysis *Work in Progress*

Outline
OO

Background
O
OOOOO
OOOOO
OOOOO

Dissertation
OOOO
●OOOO
OOOOO
OOOOO
OOOOO
OOOOO

Future Research
O
OOOO

Conclusion
OO

References
OO

Two-part Spike-and-slab LASSO (SSL) Prior for Smooth Functions

Two-part Spike-and-slab LASSO (SSL) Prior for Smooth Functions

Two-part Spike-and-slab LASSO (SSL) Prior for Smooth Functions

Generalized Additive Model

Given the data $\{y_i, x_{i1}, \ldots, x_{ip}\}_{i=1}^{n}$ where $p >> n$

$$y_i \overset{\text{i.i.d.}}{\sim} EF(\mu_i, \phi),$$
$$g(\mu_i) = \beta_0 + \sum_{j=1}^{p} B_j(x_{ij}), \quad i = 1, \ldots, n.$$

Outline
○○

Background
○
○○○○○
○○○○○
○○○○○

Dissertation
○○○
○○○●○○
○○○○○
○○○○○
○○○○○
○○○○○○

Future Research
○
○○○○

Conclusion
○○

References
○○

Two-part Spike-and-slab LASSO (SSL) Prior for Smooth Functions

Smoothing Function Reparameterization

▶ Smoothing penalty from Smoothing spline regression (Simon N. Wood 2017)

$$\lambda_j \int B_j''(x)dx = \lambda_j \beta_j^T \boldsymbol{S}_j \beta_j,$$

where S_j is a known smoothing penalty matrix.

▶ Isolate the linear and nonlinear components via eigendecomposing S_j

$$\boldsymbol{X}\boldsymbol{\beta} = X^0\beta + \boldsymbol{X}^*\beta^*$$

▶ Benefits
  ▶ Motivate bi-level selection
  ▶ Implicit modeling of function smoothness
  ▶ Reduce computation load with conditionally independent prior of basis coefficients

Two-part Spike-and-slab LASSO (SSL) Prior for Smooth Functions

Two-part Spike-and-slab LASSO (SSL) Prior

▶ SSL prior for the linear coefficient and group SSL priors for nonlinear coefficients

$$\beta_j | \gamma_j, s_0, s_1 \sim DE(0, (1 - \gamma_j)s_0 + \gamma_j s_1)$$

$$\beta_{jk}^* | \gamma_j^*, s_0, s_1 \overset{\text{iid}}{\sim} DE(0, (1 - \gamma_j^*)s_0 + \gamma_j^* s_1), k = 1, \ldots, K_j$$

▶ Effect hierarchy enforced latent inclusion indicators γ_j and γ_j^* for bi-level selection
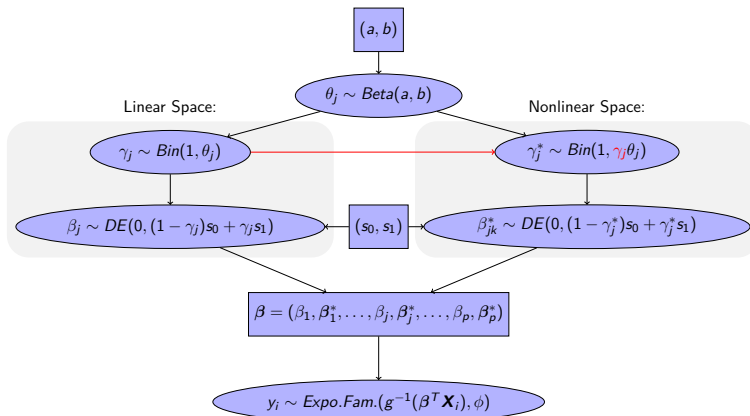
$$\gamma_j | \theta_j \sim Bin(\gamma_j | 1, \theta_j), \quad \gamma_j^* | \gamma_j, \theta_j \sim Bin(1, \gamma_j \theta_j),$$

▶ Local adaptivity of signal sparsity and function smoothness

$$\theta_j \sim \text{Beta}(a, b)$$

Outline
Background
Dissertation
Future Research
Conclusion
References

Two-part Spike-and-slab LASSO (SSL) Prior for Smooth Functions

Visual Representation

Outline
OO

Background
O
OOOOO
OOOOO
OOOOO

Dissertation
OOO
OOOOO
●OOOO
OOOOO
OOOOO
OOOOO

Future Research
O
OOOO

Conclusion
OO

References
OO

EM-Cooridante Descent Algrithm for Scalable Model Fitting

EM-Cooridante Descent Algrithm for Scalable Model Fitting

Outline
○○

Background
○
○○○○○
○○○○○
○○○○○

Dissertation
○○○
○○○○
○●○○○
○○○○○
○○○○○
○○○○○

Future Research
○
○○○○

Conclusion
○○

References
○○

EM-Cooridante Descent Algrithm for Scalable Model Fitting

EM-Cooridante Descent Algrithm for Scalable Model Fitting

We are interested in estimating $\Theta = \{\boldsymbol{\beta}, \boldsymbol{\theta}, \phi\}$ using optimization based algorithm for scalability purpose

▶ Basic Ideas
  ▶ Treat γs as the "missing data" in the EM procedure
  ▶ Quantify the expectation of log posterior density function of Θ with respect to γ conditioning on $\Theta^{(t-1)}$
  ▶ Maximize two parts of the objective function independently
▶ Previous applications in high-dimensional data analysis
  ▶ EMVS (Ročková and George 2014), Spike-and-slab lasso (Ročková and George 2018)
  ▶ BhGLM (Yi et al. 2019)

Outline   Background   Dissertation   Future Research   Conclusion   References
00   O   0000   00   00
           00000   000
           00000   00000
           00000   00●00
                         00000
                         00000
                         00000

EM-Cooridante Descent Algrithm for Scalable Model Fitting

Decomposition of Objective Function

We aim to maximize the log posterior density of Θ by averaging over all possible values of γ

$$\log f(\Theta, \gamma | \mathbf{y}, \mathbf{X}) = Q_1(\beta, \phi) + Q_2(\gamma, \boldsymbol{\theta}),$$

▶ L_1-penalized likelihood function of β, ϕ

$$Q_1 \equiv Q_1(\beta, \phi) = \log f(\mathbf{y}|\beta, \phi) + \sum_{j=1}^{p} \left[\log f(\beta_j|\gamma_j) + \sum_{k=1}^{K_j} \log f(\beta_{jk}^*|\gamma_{jk}^*) \right]$$

▶ Posterior density of θ given data points γs

$$Q_2 \equiv Q_2(\gamma, \boldsymbol{\theta}) = \sum_{j=1}^{p} \left[(\gamma_j + \gamma_j^*) \log \theta_j + (2 - \gamma_j - \gamma_j^*) \log(1 - \theta_j) \right] + \sum_{j=1}^{p} \log f(\theta_j).$$

EM-Cooridante Descent Algrithm for Scalable Model Fitting

Summary of EM-Coordinate Descent Algorithm

▶ E-step

  ▶ Formulate $E_{\gamma|\Theta^{(t)}}[Q(\Theta, \gamma)] = E(Q_1) + E(Q_2)$

    ▶ $E(Q_1)$ is a penalized likelihood function of β, ϕ
    ▶ $E(Q_2)$ is a posterior density of θ given $E(\gamma)$
    ▶ $E(Q_1)$ and $E(Q_2)$ are conditionally independent

  ▶ Calculate $E(\gamma_j)$, $E(\gamma_j^*)$ and the penalties parameters by Bayes' theorem

▶ M-step:

  ▶ Use Coordinate Descent to fit the penalized model in $E(Q_1)$ to update β, ϕ
  ▶ Closed form calculation via $E(Q_2)$ to update θ

Outline
OO

Background
O
OOOOO
OOOOO
OOOOO

Dissertation
OOO
OOOOO
OOOOO●
OOOOO
OOOOO
OOOOO

Future Research
O
OOOO

Conclusion
OO

References
OO

EM-Cooridante Descent Algrithm for Scalable Model Fitting

Tuning Parameter Selection

- ▶ s_0 and s_1 are tuning parameters
- ▶ Empirically, s_1 has extremely small effect on changing the estimates
- ▶ Focus on tuning s_0
- ▶ Consider a sequence of L ordered values $\{s_0^l\} : 0 < s_0^1 < s_0^2 < \cdots < s_0^L < s_1$
- ▶ Cross-validation to choose optimal value for s_0

Simulation Study

Simulation Study

Simulation Study

▶ Follow the data generating process introduced in Bai et al. (2020).

▶ $n_{train} = 500,\ n_{test} = 1000$

▶ $p = 4, 10, 50, 200$

$$\mu = 5\sin(2\pi x_1) - 4\cos(2\pi x_2 - 0.5) + 6(x_3 - 0.5) - 5(x_4^2 - 0.3),$$

▶ $f_j(x_j) = 0$ for $j = 5, \ldots, p$.

▶ 2 types of outcome: Gaussian $(\phi = 1)$, Binomial

▶ Splines are constructed using 10 knots

▶ 50 Iterations

Outline
OO

Background
O
OOOOO
OOOOO
OOOOO

Dissertation
OOOO
OOOOO
OO●OO
OOOOO
OOOOO

Future Research
O
OOOO

Conclusion
OO

References
OO

Simulation Study

Comparison & Metircs

- ▶ Methods of comparison
    - ▶ Proposed model BHAM
    - ▶ Linear LASSO model as the benchmark
    - ▶ mgcv (S. N. Wood 2004)
    - ▶ COSSO (Zhang and Lin 2006) and adaptive COSSO(Storlie et al. 2011)
    - ▶ Sparse Bayesian GAM (Bai 2021)
    - ▶ spikeSlabGAM (Scheipl, Fahrmeir, and Kneib 2012)
- ▶ Metrics
    - ▶ Prediction: R^2 for continuous outcomes, out-of-sample AUC for binary outcomes
    - ▶ Variable Selection: positive predictive value (precision), true positive rate (recall), and Matthews correlation coefficient (MCC)

Outline
oo

Background
o
oooooo
ooooo
ooooo

Dissertation
ooo
oooooo
oooooo
ooo●o
oooooo
oooooo

Future Research
o
oooo

Conclusion
oo

References
oo

Simulation Study

Prediction Performance

- ▶ Linear LASSO Model performs bad and mgcv performs well
- ▶ BHAM performs better than COSSO, adaptive COSSO and spikeSlabGAM
- ▶ BHAM performs better than SB-GAM in low-dimensional case but slightly worse in the high-dimensional setting
- ▶ BHAM is much faster than SB-GAM in fitting models

Outline
OO

Background
O
OOOOO
OOOO
OOOOO

Dissertation
OOO
OOOOO
OOOOO
OOOO●
OOOOO
OOOOO

Future Research
O
OOOO

Conclusion
OO

References
OO

Simulation Study

Variable Selection Performance

- ▶ SB-GAM has the best variable selection performance
- ▶ BHAM has conservative selection
- ▶ BHAM and spikeSlabGAM have trade-offs for bi-level selection
  - ▶ spikeSlabGAM tends to select either linear or nonlinear components of the function
  - ▶ BHAM is more likely to select both parts

Outline
OO

Background
O
OOOOO
OOOOO
OOOOO

Dissertation
OOO
OOOOO
OOOOO
OOOOO
●OOOO
OOOOO
OOOOO

Future Research
O
OOOO

Conclusion
OO

References
OO

Additive Cox Proportional Hazards Model

Additive Cox Proportional Hazards Model

Outline
○○

Background
○
○○○○○
○○○○○
○○○○○

Dissertation
○○○
○○○○○
○○○○○
○●○○○
○○○○○

Future Research
○
○○○○

Conclusion
○○

References
○○

Additive Cox Proportional Hazards Model

Model & Objective Functions

▶ Cox proportional hazard model with event time t_i

$$h(t_i) = h_0(t_i) \exp(\sum_{j=1}^{p} B_j(x_{ij})), \quad i = 1, \ldots, n.$$

  ▶ No intercept term because of the baseline hazard function
▶ Model fitting
  ▶ Replace likelihood function with partial likelihood function

$$\hat{h}_0(t_i|\beta) = d_i / \sum_{i' \in R(t_i)} exp(X_{i'}\beta).$$

Outline
○○

Background
○
○○○○○
○○○○○

Dissertation
○○○
○○○○○
○○○○○
○○○●○○
○○○○○

Future Research
○○○○

Conclusion
○○

References
○○

Additive Cox Proportional Hazards Model

Two-part Spike-and-slab LASSO (SSL) Prior

▶ SSL prior for the linear coefficient and group SSL priors for nonlinear coefficients

$$\beta_j | \gamma_j, s_0, s_1 \sim DE(0, (1 - \gamma_j)s_0 + \gamma_j s_1)$$

$$\beta_{jk}^* | \gamma_j^*, s_0, s_1 \overset{\text{iid}}{\sim} DE(0, (1 - \gamma_j^*)s_0 + \gamma_j^* s_1), k = 1, \dots, K_j$$

▶ Effect hierarchy enforced latent inclusion indicators γ_j and γ_j^* for bi-level selection

$$\gamma_j | \theta_j \sim Bin(\gamma_j | 1, \theta_j), \quad \gamma_j^* | \gamma_j, \theta_j \sim Bin(1, \gamma_j \theta_j),$$

▶ Local adaptivity of signal sparsity and function smoothness

$$\theta_j \sim \text{Beta}(a, b)$$

Summary of EM-Coordinate Descent Algorithm

- ▶ E-step
    - ▶ Formulate $E_{\gamma|\Theta^{(t)}}[Q(\Theta, \gamma)] = E(Q_1) + E(Q_2)$
        - ▶ $E(Q_1)$ is a penalized likelihood function of β, ϕ
        - ▶ $E(Q_2)$ is a posterior density of θ given $E(\gamma)$
        - ▶ $E(Q_1)$ and $E(Q_2)$ are conditionally independent
    - ▶ Calculate $E(\gamma_j)$, $E(\gamma_j^*)$ and the penalties parameters by Bayes' theorem
- ▶ M-step:
    - ▶ Use Coordinate Descent to fit the penalized model in $E(Q_1)$ to update β, ϕ
    - ▶ Closed form calculation via $E(Q_2)$ to update θ

Outline
OO

Background
O
OOOOO
OOOOO
OOOOO

Dissertation
OOO
OOOOO
OOOOO
OOOOO
OOOOO
●OOOO
OOOOO

Future Research
O
OOOO

Conclusion
OO

References
OO

Simulation Study

Simulation Study

Outline	Background	Dissertation	Future Research	Conclusion	References
○○	○	○○○	○	○○	○○
	○○○○○	○○○○○	○○○○		
	○○○○○	○○○○○			
	○○○○○	○○○○○			
		○●○○○			
		○○○○○			

Simulation Study

Simulation Study

▶ Follow the data generating process introduced in Bai et al. (2020).

▶ $n_{train} = 500$, $n_{test} = 1000$

▶ $p = 4, 10, 50, 200$

$$\mu = 5\sin(2\pi x_1) - 4\cos(2\pi x_2 - 0.5) + 6(x_3 - 0.5) - 5(x_4^2 - 0.3),$$

▶ $f_j(x_j) = 0$ for $j = 5, \ldots, p$.

▶ 2 types of outcome: Gaussian ($\phi = 1$), Binomial

▶ Splines are constructed using 10 knots

▶ 50 Iterations

Outline
OO

Background
O
OOOOO
OOOOO
OOOOO

Dissertation
OOO
OOOOO
OOOOO
OOOOO
OO●OO
OOOOO

Future Research
O
OOOO

Conclusion
OO

References
OO

Simulation Study

Comparison & Metircs

▶ Methods of comparison
  ▶ Proposed model BHAM
  ▶ Linear LASSO model as the benchmark
  ▶ mgcv (S. N. Wood 2004)
  ▶ COSSO (Zhang and Lin 2006) and adaptive COSSO(Storlie et al. 2011)
▶ Metrics
  ▶ Out-of-sample deviance & Concordance
  ▶ Variable Selection: positive predictive value (precision), true positive rate (recall), and
    Matthews correlation coefficient (MCC)

Outline
00

Background
0
00000
00000
00000

Dissertation
000
00000
00000
00000
000●0
00000

Future Research
0
0000

Conclusion
00

References
00

Simulation Study

Prediction Performance

▶ Linear LASSO Model performs bad and mgcv performs well
▶ BHAM performs better than COSSO, adaptive COSSO and spikeSlabGAM

Outline
○○

Background
○
○○○○○
○○○○○

Dissertation
○○○
○○○○○
○○○○○
○○○○○
○○○○○
○○●○○
○○○○○

Future Research
○
○○○○

Conclusion
○○

References
○○

Simulation Study

Emipirical Performance: Emory Cardiovascular Biobank

Outline
○○

Background
○
○○○○○
○○○○○
○○○○○

Dissertation
○○○
○○○○○
○○○○○
○○○○○
○○○○○
●○○○○○

Future Research
○
○○○○

Conclusion
○○

References
○○

R Package BHAM

R Package BHAM

Outline
OO

Background
O
OOOOO
OOOOO
OOOOO

Dissertation
OOO
O
OOOOO
OOOOO
OOOOO
OOOOO
OO●OO

Future Research
O
OOOO

Conclusion
OO

References
OO

R Package BHAM

R Package BHAM

- ▶ Model formulation for high-dimensional data
- ▶ Model fitting and tuning
- ▶ Model summary and variable selection
- ▶ Spline function visualization
- ▶ Website via *boyiguo1.github.io/BHAM*

Design Matrix of Spline Fucntions

▶ Flexible spline function formulation for high-dimensional data via `construct_smooth_data`

```
spline_df <- dplyr::tribble(
    ~Var, ~Func, ~Args,
    "X1", "s", "bs='cr', k=5",
    "X2", "s", NA,
    "X3", "s", "")
spline_df <- data.frame(
    Var = setdiff(names(dat), "y"),
    Func = "s",
    Args ="bs='cr', k=7")
train_sm_dat <- BHAM::construct_smooth_data(spline_df, dat)
```

Outline
OO

Background
O
OOOOO
OOOOO
OOOOO

Dissertation
OOO
OOOOO
OOOOO
OOOOO
OOOOO
OOOO●O

Future Research
O
OOOO

Conclusion
OO

References
OO

R Package BHAM

Model Fitting Functions

- ▶ Model fitting via `bamlasso`
  - ▶ Argument `family` for generalized and survival outcomes
  - ▶ Argument `ss` for spike-and-slab LASSO scale parameters
  - ▶ Argument `group` for group structures among predictors
- ▶ Model tuning via `tune`
  - ▶ Argument `nfolds`, `ncv` for nested cross-validation
  - ▶ Argument `s0` for tuning candidates

Outline
oo

Background
oooooo
ooooo
ooooo

Dissertation
oooo
ooooo
ooooo
ooooo
ooooo
oooooo

Future Research
o
oooo

Conclusion
oo

References
oo

R Package BHAM

Post Fitting Functions

▶ Bi-level selection via bamlasso_var_selection
▶ Make prediction data for splines make_predict_dat
▶ Plot spline functions via plot_smooth_term

Outline
OO

Background
O
OOOOO
OOOOO

Dissertation
OOO
OOOOO
OOOOO
OOOOO
OOOOO
OOOOO

Future Research
●
OOOO

Conclusion
OO

References
OO

Future Research

Outline
OO

Background
O
OOOOO
OOOOO

Dissertation
OOO
OOOOO
OOOOO
OOOOO
OOOOO

Future Research
O
●OOO

Conclusion
OO

References
OO

Modeling Interactions

Modeling Interactions

Outline   Background   Dissertation   Future Research   Conclusion   References
○○   ○   ○○○   ○   ○○   ○○
                 ○○○○○   ○○○○○   ○●○○
                 ○○○○○   ○○○○○
                 ○○○○○   ○○○○○
                                     ○○○○○
                                     ○○○○○

Modeling Interactions

Varying coefficient models

▶ Assume the coefficient of a variable X_j is a function of a covariate Z_j
  ▶ linear model: $\beta(Z_j) = \beta$
  ▶ VC model: $\beta(Z_j) = B(Z_j)$
▶ Replace each smooth function $B(z_{ij})$ with $B(z_{ij})x_{ij} \equiv (x_{ij}\boldsymbol{Z}_{ij}^T)\beta_j$
▶ Model fitting with EM-Coordinate Descent

Question: Can Z_j be continuous? Can we have a more flexible model?

Outline
oo

Background
o
ooooo
ooooo

Dissertation
ooo
ooooo
ooooo
ooooo
ooooo

Future Research
o
oooo

Conclusion
oo

References
oo

Modeling Interactions

Smooth Surface Fitting

▶ Tensor product of spline functions

$$B_j(x_{ij}) + B_s(x_{is}) + B_{js}(x_{ij}, x_{is}),$$

where

$$B_{js}(x_{ij}, x_{is}) = \sum_{P\rho=1}^{K} \sum_{v=1}^{K} \beta_{jspv} b_{j\rho}(x_{ij}) b_{sv}(x_{is})$$

Question: Can we have a generalized model that accounts fixed effects, nonlinear curves, smooth surfaces, and random effects?

Modeling Interactions

Structural Additive Model

High-dimensional structural additive model can be formulated as

$$g(\mathbb{E}(y_i)) = \boldsymbol{x}_i^T \boldsymbol{\theta} + \boldsymbol{u}_i^T \boldsymbol{\gamma} + B(z_{i1}) + B(z_{i2}, z_{i3}) + B_{spat}(s_i)$$

▶ Un-regularized predictors $\boldsymbol{x}_i$
▶ Regularized predictors $\boldsymbol{u}_i$
▶ Predictors with nonlinear effects $\boldsymbol{z}_i$
▶ Spatial random effects with coordinates s_i

Spike-and-slab LASSO prior motivates a seamless process of variable/functional selection and a scalable optimization-based model fitting algorithm

Outline
○○

Background
○
○○○○○
○○○○○
○○○○○

Dissertation
○○○
○○○○○
○○○○○
○○○○○
○○○○○
○○○○○

Future Research
○
○○○○

Conclusion
●○

References
○○

Conclusion

Conclusion

- ▶ Identify challenges in high-dimensional GAM with spline functions
  - ▶ Balance between signal sparsity and function smoothness
  - ▶ Bi-level selection to automatically detect linear and nonlinear effects
- ▶ Statistical contribution
  - ▶ Two-part spike-and-slab LASSO prior for smooth functions
  - ▶ Scalable EM-Coordinate Descent algorithms for generalized and survival outcomes
  - ▶ R package BHAM
- ▶ Future Research
  - ▶ Extension of spike-and-slab LASSO prior in structured additive model

Outline
OO

Background
O
OOOOO
OOOOO

Dissertation
OOO
OOOOO
OOOOO
OOOOO
OOOOO
OOOOO

Future Research
O
OOOO

Conclusion
OO

References
●O

References

Outline
oo

Background
o
ooooo
ooooo

Dissertation
ooo
ooooo
oooo
ooooo
ooooo

Future Research
oooo

Conclusion
oo

References
o●

References I

Bai, Ray. 2021. "Spike-and-Slab Group Lasso for Consistent Estimation and Variable Selection in Non-Gaussian Generalized Additive Models." *arXiv:2007.07021v5.*

Bai, Ray, Gemma E Moran, Joseph L Antonelli, Yong Chen, and Mary R Boland. 2020. "Spike-and-Slab Group Lassos for Grouped Regression and Sparse Generalized Additive Models." *Journal of the American Statistical Association*, 1–14.

Hastie, Trevor, and Robert Tibshirani. 1987. "Generalized additive models: Some applications." *Journal of the American Statistical Association* 82 (398): 371–86. https://doi.org/10.1080/01621459.1987.10478440.

Hastie, Trevor, Robert Tibshirani, and Jerome Friedman. 2009. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction.* Springer Science & Business Media.

Outline
○○

Background
○
○○○○○
○○○○○

Dissertation
○○○○
○○○○○
○○○○○
○○○○○
○○○○○

Future Research
○
○○○○

Conclusion
○○

References
○●

References II

Huang, Jian, Joel L Horowitz, and Fengrong Wei. 2010. "Variable Selection in
Nonparametric Additive Models." *Annals of Statistics* 38 (4): 2282.

Ravikumar, Pradeep, John Lafferty, Han Liu, and Larry Wasserman. 2009. "Sparse
additive models." *Journal of the Royal Statistical Society: Series B (Statistical
Methodology)* 71 (5): 1009–30. https://doi.org/10.1111/j.1467-9868.2009.00718.x.

Ročková, Veronika, and Edward I. George. 2014. "EMVS: The EM approach to
Bayesian variable selection." *Journal of the American Statistical Association* 109
(506): 828–46. https://doi.org/10.1080/01621459.2013.869223.

———. 2018. "The Spike-and-Slab LASSO." *Journal of the American Statistical
Association* 113 (521): 431–44. https://doi.org/10.1080/01621459.2016.1260469.

Outline
○○

Background
○
○○○○○
○○○○○

Dissertation
○
○○○○
○○○○○
○○○○
○○○○○
○○○○○

Future Research
○
○○○○

Conclusion
○○

References
○●

References III

Scheipl, Fabian, Ludwig Fahrmeir, and Thomas Kneib. 2012. "Spike-and-slab priors for function selection in structured additive regression models." *Journal of the American Statistical Association* 107 (500): 1518–32. https://doi.org/10.1080/01621459.2012.737742.

Storlie, Curtis B, Howard D Bondell, Brian J Reich, and Hao Helen Zhang. 2011. "Surface Estimation, Variable Selection, and the Nonparametric Oracle Property." *Statistica Sinica* 21 (2): 679.

Wang, Lifeng, Guang Chen, and Hongzhe Li. 2007. "Group SCAD Regression Analysis for Microarray Time Course Gene Expression Data." *Bioinformatics* 23 (12): 1486–94.

Outline
○○

Background
○○○○○
○○○○○
○○○○○

Dissertation
○○○
○○○○○
○○○
○○○○○
○○○○○
○○○○○

Future Research
○
○○○○

Conclusion
○○

References
○●

References IV

Wood, S. N. 2004. "Stable and Efficient Multiple Smoothing Parameter Estimation for Generalized Additive Models." *Journal of the American Statistical Association* 99 (467): 673–86.

Wood, Simon N. 2017. *Generalized additive models: An introduction with R, second edition.* https://doi.org/10.1201/9781315370279.

Xue, Lan. 2009. "Consistent Variable Selection in Additive Models." *Statistica Sinica*, 1281–96.

Yi, Nengjun, Zaixiang Tang, Xinyan Zhang, and Boyi Guo. 2019. "BhGLM: Bayesian Hierarchical GLMs and Survival Models, with Applications to Genomics and Epidemiology." *Bioinformatics* 35 (8): 1419–21.

Outline
○○

Background
○
○○○○○
○○○○○
○○○○○

Dissertation
○○○
○○○○○
○○○○○
○○○○○
○○○○○
○○○○○

Future Research
○
○○○○

Conclusion
○○

References
○●

References V

Zhang, Hao Helen, and Yi Lin. 2006. "Component Selection and Smoothing for Nonparametric Regression in Exponential Families." *Statistica Sinica*, 1021–41.