Outline
○○

Background
○○○
○○○○○
○○○○○

Dissertation
○○○
○○○○○○○○○○○○○
○○○○○

Conclusion
○
○○○○
○○○
○○○

References
○○

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
00

Background
00000
00000

Dissertation
000
0000000000000
00000

Conclusion
0
0000
000
000

References
00

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

Outline
○●

Background
○○○○○
○○○○○

Dissertation
○○○
○○○○○○○○○○○○○
○○○○○

Conclusion
○
○○○○
○○○
○○○

References
○○

Outline

► Background
  ► Spline Model Development
  ► Bayesian Regularization
  ► Bayesian Variable Selection

► Dissertation
  ► Bayesian Hierarchical Additive Models
  ► Additive Cox Proportional Hazards Model
  ► R package BHAM

► Conclusion
  ► Future Research
  ► Questions & Answers
  ► Closing Statement & Acknowledgment

Outline
○○

Background
●
○○○○○
○○○○○
○○○○○

Dissertation
○○○
○○○○○○○○○○○○○○
○○○○○

Conclusion
○
○○○○
○○
○○○

References
○○

Background

Outline
OO

Background
●○○○○
○○○○
○○○○○

Dissertation
○○○
○○○○○○○○○○○○○
○○○○○

Conclusion
○
○○○○
○○
○○○

References
○○

Spline Model Development

Spline Model Development

Outline
OO

Background
○●○○○
○○○○○
○○○○○

Dissertation
○○○
○○○○○○○○○○○○○
○○○○○

Conclusion
○
○○○○
○○
○○○

References
○○

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?

Outline
○○

Background
○○●○○
○○○○
○○○○○

Dissertation
○○○
○○○○○○○○○○○○○
○○○○○

Conclusion
○
○○○○
○○○
○○○

References
○○

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. (Simon N. Wood 2017)

Functions: — Linear — Spline
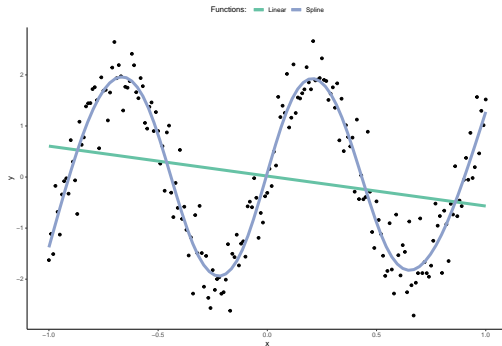
▶ For simplicity, we assume all functions have K basis functions and knots of functions are equidistance.

Outline
○○

Background
○○○●○
○○○○○

Dissertation
○○○○
○○○○○○○○○○○○
○○○○○

Conclusion
○○○○
○○○
○○○

References
○○

Spline Model Development

Generalized Additive Models with Splines

Generalized additive model (Hastie and Tibshirani 1987) is expressed

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

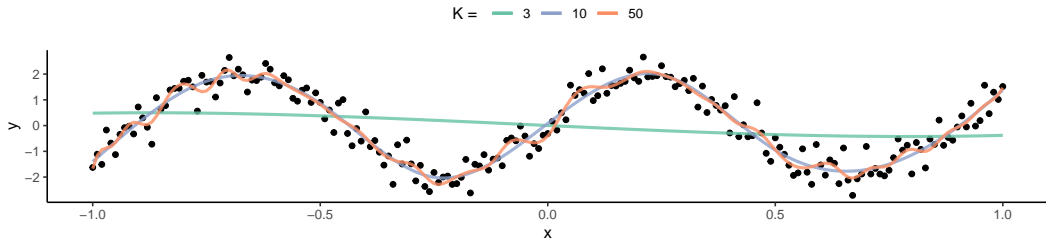where $B(x_i)$ is the spline function, $g(\cdot)$ is a link function, ϕ is the dispersion parameter

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

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

Outline
○○

Background
○○○○●
○○○○○
○○○○○

Dissertation
○○○
○○○○○○○○○○○○○
○○○○○

Conclusion
○○○○
○○○
○○○
○○○

References
○○

Spline Model Development

Problem: Function Smoothness

Question

How to mathematically define and estimate the smoothness of spline functions?

Outline
OO

Background
O
OOOOO
●OOOO
OOOOO

Dissertation
OOO
OOOOOOOOOOOOO
OOOOO

Conclusion
O
OOOO
OO
OOO
OOO

References
OO

Bayesian Regularization

Bayesian Regularization

Outline
OO

Background
OOOOO
OⓋOOO
OOOOO

Dissertation
OOO
OOOOOOOOOOOOO
OOOOO

Conclusion
OOOO
OO
OOO
OOO

References
OO

Bayesian Regularization

Smoothing Spline Model

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

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

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

Outline
○○

Background
○○○○○
○○○●○○
○○○○○

Dissertation
○○○
○○○○○○○○○○○○○
○○○○○

Conclusion
○○○○
○○○
○○○

References
○○

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

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

Question

How to choose λ_i for $i = 1, \ldots, p$?

- Global smoothing: $\lambda_1 = \cdots = \lambda_p$
- Adaptive smoothing: unique λ_i for $i = 1, \ldots, p$

Outline
○○

Background
○
○○○○○
○○○○●
○○○○○

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

Conclusion
○
○○○○
○○
○○○
○○○

References
○○

Bayesian Regularization

Bayesian Regularization

▶ Bayesian regularization is the Bayesian analogy of penalized models by using regularizing priors

$$\text{Bayesian ridge:} \quad \beta \sim N(0, \tau^2) \rightarrow \lambda = \sigma^2/\tau^2$$

▶ 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 (Lang and Brezger 2004)
  ▶ Log-normal spline model for τ_k^2 (Baladandayuthapani, Mallick, and Carroll 2005)

Outline
OO

Background
O
OOOOO
OOOOO
●OOOO

Dissertation
OOO
OOOOOOOOOOOOO
OOOOO

Conclusion
O
OOOO
OOO
OOO

References
OO

Bayesian Variable Selection

Bayesian Variable Selection

Outline
○○

Background
○
○○○○○
○●○○○○

Dissertation
○○○
○○○○○○○○○○○○
○○○○○

Conclusion
○
○○○○
○○○
○○○

References
○○

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 of the outcome.

Question

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
OO

Background
O
OOOOO
OO●OO

Dissertation
OOO
OOOOOOOOOOOOO
OOOOO

Conclusion
O
OOOO
OOO
OOO
OOO

References
OO

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
▶ Slab density $f_{slab}(x)$ is a flat density for large effects
▶ Spike density $f_{spike}(x)$ concentrates around 0 for small 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
○○○
○○○○○○○○○○○○○
○○○○○

Conclusion
○
○○○○
○○
○○○

References
○○

Bayesian Variable Selection

Spike-and-Slab LASSO Priors

▶ Double exponential distributions as the spike and slab distributions

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

  ▶ Computation advantages via Expectation-Maximization (EM) algorithms
  ▶ Seamless variable selection as coefficients shrink to 0
▶ Group spike-and-slab LASSO prior
  ▶ Structure among predictors, e.g. gene pathways, bases of a spline function
  ▶ Structured prior $\gamma_k|\theta_j \overset{\text{iid}}{\sim} Binomial(1, \theta_j), k \in j$

Outline
○○

Background
○○○○○
○○○○○
○○○○●

Dissertation
○○○
○○○○○○○○○○○○○
○○○○○

Conclusion
○○○○
○○○
○○○
○○○

References
○○

Bayesian Variable Selection

Problem: High-dimensional Spline Model

Question

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

▶ Excess shrinkage due to negligence of smooth penalty
  ▶ Group lasso penalty (Ravikumar et al. 2009; Huang, Horowitz, and Wei 2010), group SCAD penalty (Wang, Chen, and Li 2007; Xue 2009)
▶ All-in-all-out selection
  ▶ Failed to select function as a whole, e.g. group spike-and-slab LASSO prior
  ▶ Can not detect if a function is linear, e.g. spike-and-slab grouped LASSO prior (Bai et al. 2020; Bai 2021)

Outline
OO

Background
O
OOOOO
OOOOO

Dissertation
●OO
OOOOOOOOOOOOOO
OOOOO

Conclusion
O
OOOO
OO
OOO

References
OO

Dissertation

Outline
○○

Background
○○○○○
○○○○
○○○○○

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

Conclusion
○
○○○○
○○
○○○

References
○○

Objectives

▶ To develop statistical models that improve curve interpolation and outcome prediction
  ▶ Adaptive regularization that accounts for signal sparsity and function smoothness
  ▶ Bi-level selection for linear and nonlinear effect
▶ To develop a fast and scalable algorithm
▶ To implement a user-friendly statistical software

Outline
○○

Background
○○○
○○○○○
○○○○○

Dissertation
○○●
○○○○○○○○○○○○○
○○○○○

Conclusion
○○○○
○○○
○○○

References
○○

Projects

▶ **Guo, B.**, Jaeger, B. C., Rahman, A. F., Long, D. L., Yi, N. (2022). Spike-and-Slab LASSO 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 hazards 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. *arXiv*. doi: https://doi.org/10.48550/arXiv.2207.02348

Outline
○○

Background
○
○○○○○
○○○○○

Dissertation
○○○
●○○○○○○○○○○○○
○○○○○

Conclusion
○
○○○○
○○
○○○

References
○○

Bayesian Hierarchical Additive Models

Bayesian Hierarchical Additive Models

Outline
OO

Background
OOOOO
OOOOO

Dissertation
OOOOOOOOOOOOO
OOOOO

Conclusion
OOOO
OO
OOO

References
OO

Bayesian Hierarchical Additive Models

Generalized Additive Model

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

$$y_i \stackrel{\text{i.i.d.}}{\sim} EF(\mu_i, \phi), \quad i = 1, \ldots, n.$$

$$g(\mu_i) = \beta_0 + \sum_{j=1}^{p} B_j(x_{ij}) = \beta_0 + \sum_{j=1}^{p} \sum_{k=1}^{K} \beta_{jk} b_{jk}(x_{ij}) = \beta_0 + \sum_{j=1}^{p} \boldsymbol{x}_{ij}^{T} \beta_j$$

▶ Each spline function consists of K bases
▶ Identifiability constraint: $\mathbb{E}[B_j(X)] = 0, j = 1, \ldots, p$

Outline
oo

Background
o
ooooo
ooooo

Dissertation
ooo
oo●oooooooooo
ooooo

Conclusion
oooo
oo
ooo
ooo

References
oo

Bayesian Hierarchical Additive Models

Spline Function Reparameterization

- ▶ Smoothing penalty $\lambda\beta^T \boldsymbol{S}\beta$
    - ▶ $\boldsymbol{S}$ is symmetric and positive semi-definite
    - ▶ $\boldsymbol{S} = \boldsymbol{U}\boldsymbol{D}\boldsymbol{U}^T$ via eigendecomposing
- ▶ Isolate the linear and nonlinear components

$$\boldsymbol{X}^T \beta = (\boldsymbol{X}^T \boldsymbol{U})(\boldsymbol{U}^T \beta) = X^0\beta + \boldsymbol{X}^*\beta^*$$

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

Outline
○○

Background
○
○○○○○
○○○○
○○○○

Dissertation
○○○
○○○●○○○○○○○○○○
○○○○○

Conclusion
○
○○○○
○○
○○○
○○○

References
○○

Bayesian Hierarchical Additive Models

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

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

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

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

   ▶ γ_j controls the inclusion of linear component
   ▶ γ_j^* controls the inclusion of nonlinear component

Outline	Background	Dissertation	Conclusion	References
○○	○○○○○ ○○○○○ ○○○○○	○○○ ○○○○●○○○○○○○○○ ○○○○○	○○○○ ○○ ○○ ○○○	○○

Bayesian Hierarchical Additive Models

Effect Hierarchy

▶ *Effect hierarchy* assumes lower-order effects are more likely to be active than higher-order effects

▶ Structured prior on latent inclusion indicators γ_j and γ_j^*

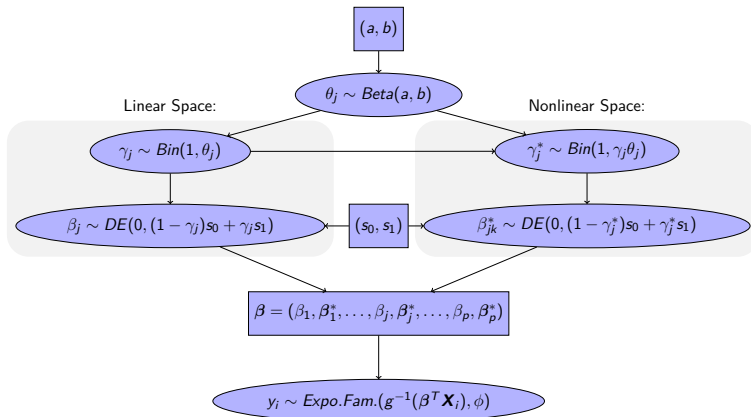$$\gamma_j|\theta_j \sim Bin(\gamma_j|1, \theta_j), \quad \gamma_j^*|\gamma_j, \theta_j \sim Bin(1, \gamma_j\theta_j),$$

  ▶ Simplification via analytic integration

$$\gamma_j^*|\theta_j \sim Bin(1, \theta_j^2),$$

▶ Adaptive shrinkage

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

Bayesian Hierarchical Additive Models

Visual Representation

Outline
○○

Background
○
○○○○○
○○○○○

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

Conclusion
○
○○○○
○○○
○○○

References
○○

Bayesian Hierarchical Additive Models

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
○○○
○○○○○○○●○○○○○○
○○○○○

Conclusion
○○○○
○○○
○○○

References
○○

Bayesian Hierarchical Additive Models

Decomposition of Objective Function

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

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

▶ L_1-penalized likelihood function of $\boldsymbol{\beta}, \phi$

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

▶ Posterior density of θ given data points γs

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

▶ Q_1 and Q_2 are independent conditioning on γs

Outline
○○

Background
○○○
○○○○○
○○○○○

Dissertation
○○○
○○○○○○○○●○○○○○
○○○○○

Conclusion
○○○○
○○○
○○○

References
○○

Bayesian Hierarchical Additive Models

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

Dissertation
oooooooooo●oooo
oooooooo
ooooo

Conclusion
o
oooo
oo
ooo

References
oo

Bayesian Hierarchical Additive Models

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

Outline
○○

Background
○○○○○
○○○○○
○○○○○

Dissertation
○○○
○○○○○○○○○○●○○○
○○○○○

Conclusion
○○○○
○○
○○
○○○

References
○○

Bayesian Hierarchical Additive Models

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
ooo
ooooo
ooooo

Dissertation
ooo
oooooooooo●oo
ooooo

Conclusion
o
oooo
oo
ooo
ooo

References
oo

Bayesian Hierarchical Additive Models

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
○○

Background
○○
○○○○○
○○○○○

Dissertation
○○○○○○○○○○○○●○
○○○○○

Conclusion
○
○○○○
○○
○○○

References
○○

Bayesian Hierarchical Additive Models

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
OOOOO

Dissertation
OOOOOOOOOOOOO●
OOOOO

Conclusion
O
OOOO
OO
OOO

References
OO

Bayesian Hierarchical Additive Models

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
○○

Background
○
○○○○○
○○○○
○○○○○

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

Conclusion
○
○○○○
○○
○○○

References
○○

Additive Cox Proportional Hazards Model

Outline
oo

Background
o
ooooo
ooooo

Dissertation
ooo
oooooooooooooo
ooooo

Conclusion
o
oooo
ooo
ooo

References
oo

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	Conclusion	References
00	00000	000	0000	00
	00000	00●000000000000	000	
	00000	00000	000	

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

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

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

▶ 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
○○○
○○○●○○○○○○○○
○○○○○

Conclusion
○○○○
○○○
○○○

References
○○

Additive Cox Proportional Hazards Model

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	Background	Dissertation	Conclusion	References
○○	○○○○○	○○○	○○○○	○○
	○○○○○	○○○○○○○○○○○○○	○○○	
	○○○○○	○○○○●○○○○	○○○	
			○○○	

Additive Cox Proportional Hazards Model

Simulation Study

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

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

▶ $p = 4, 10, 50, 100, 200$

▶ Survival and censoring time follow Weibull distribution

$$\log \eta = (x_1 + 1)^2/5 + \exp(x_2 + 1)/25 + 3\sin(x_3)/2 + (1.4x_4 + 0.5)/2$$

▶ Censoring rate is controlled at $\{0.15, 0.3, 0.45\}$

▶ Splines are constructed using 10 knots

▶ 50 Iterations

Outline
oo

Background
o
ooooo
ooooo

Dissertation
ooo
oooooooooooooo
ooooo●oo

Conclusion
oooo
ooo
ooo

References
oo

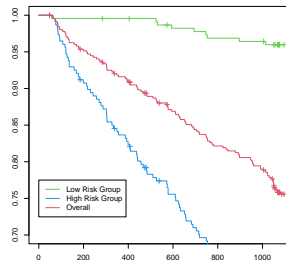Additive Cox Proportional Hazards Model

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

Outline
Background
Dissertation
Conclusion
References

Additive Cox Proportional Hazards Model

Prediction Performance

- ▶ Linear LASSO Model performs bad in general
- ▶ Low dimensional settings:
    - ▶ mgcv performs the best
    - ▶ BHAM performs as good as mgcv
- ▶ High dimensional setting:
    - ▶ BHAM performs better than COSSO models as p increases and more censoring events

Outline
○○

Background
○○○○
○○○○○
○○○○○

Dissertation
○○○
○○○○○○○○○○○○○
○○○○○●

Conclusion
○○○○
○○○
○○○

References
○○

Additive Cox Proportional Hazards Model

Emipirical Performance: Emory Cardiovascular Biobank

▶ predicting all-cause mortality among patents undergoing cardiac catheterization

▶ Sample size N=454 and number of features p=200

▶ 5-knot cubic spline

Outline
OO

Background
O
OOOOO
OOOOO

Dissertation
OOO
OOOOOOOOOOOOOO
●OOOO

Conclusion
O
OOOO
OO
OOO

References
OO

R Package BHAM

R Package BHAM

Outline
○○

Background
○
○○○○○
○○○○○
○○○○○

Dissertation
○○○
○○○○○○○○○○○○○
○●○○○

Conclusion
○
○○○○
○○
○○○

References
○○

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*

Outline
○○

Background
○○○
○○○○○
○○○○○

Dissertation
○○○
○○○○○○○○○○○○○
○○●○○

Conclusion
○○○○
○○○
○○○

References
○○

R Package BHAM

Design Matrix of Spline Fucntions

▶ Flexible spline function formulation for high-dimensional 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
ooo
ooooo
ooooo

Dissertation
ooo
ooooooooooooo
ooooeo

Conclusion
o
oooo
oo
ooo

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
O OOOOO
OOOOO
OOOOO

Dissertation
OOO
OOOOOOOOOOOOOO
OOOO●

Conclusion
O
OOOO
OO
OOO

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
○○

Background
○
○○○○○
○○○○○
○○○○○

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

Conclusion
●○○○
○○○○
○○
○○○
○○○

References
○○

Conclusion

Outline
○○

Background
○
○○○○○
○○○○
○○○○○

Dissertation
○○○
○○○○○○○○○○○○○
○○○○○

Conclusion
●○○○
○○
○○
○○○

References
○○

Future Research

Future Research

Outline
○○

Background
○
○○○○○
○○○○○

Dissertation
○○○
○○○○○○○○○○○○○
○○○○○

Conclusion
○●○○
○○○
○○○

References
○○

Future Research

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 spline function $B(z_{ij})$ with $B(z_{ij})x_{ij} \equiv (x_{ij}\mathbf{Z}_{ij}^T)\beta_j$
- Model fitting with EM-Coordinate Descent
- Nonlinear interaction of a continuous variable and a categorical variable

Question

How to model nonlinear interaction of two continuous variables?

Outline	Background	Dissertation	Conclusion	References
oo	ooooo	ooo	oooo	oo
	ooooo	oooooooooooooo	ooo	
	ooooo	ooooo	ooo	

Future Research

Smooth Surface Fitting

▶ Tensor product of spline functions

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

▶ Smooth Surface

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

Question

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

Outline
○○

Background
○○○○○
○○○○○
○○○○○

Dissertation
○○○○
○○○○○○○○○○○○
○○○○○

Conclusion
○○○●
○○○
○○○

References
○○

Future Research

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
○○○
○○○○○○○○○○○○○○
○○○○○○
○○○○○

Conclusion
○
○○○○
●○
●●○○
○○○
○○○

References
○○

Q & A

Q & A

Outline
OO

Background
O OOOO
OOOOO
OOOOO

Dissertation
OOO
OOOOOOOOOOOOOOO
OOOOO

Conclusion
O OOO
OO
O●O
OOO

References
OO

Q & A

Q & A

TODO (Audience): Ask questions here

Outline
oo

Background
o
ooooo
ooooo

Dissertation
ooo
oooooooooooooo
ooooo

Conclusion
o
oooo
oo
●oo
ooo

References
oo

Conclusion

Conclusion

Outline
○○

Background
○○○
○○○○○
○○○○○

Dissertation
○○○
○○○○○○○○○○○○
○○○○○

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 spline 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
OOOOO
OOOOO

Dissertation
OOO
OOOOOOOOOOOO
OOOOO

Conclusion
OOOO
OO
OOO●
OOO

References
OO

Conclusion

Acknowledgement

I would like to thank

- ▶ My parents and girlfriend
- ▶ Dissertation Committee
- ▶ Biostatistics Department
  - ▶ Faculty
  - ▶ Staff
  - ▶ Students

- ▶ REGARDS Research Group
  - ▶ PIs
  - ▶ Analytic Team
  - ▶ Collaborators
- ▶ School of Public Health
  - ▶ Friends
- ▶ Public Health Student Association
- ▶ Graduate Student Government

Outline
○○

Background
○
○○○○○
○○○○○

Dissertation
○○○
○○○○○○○○○○○○○○
○○○○○

Conclusion
○
○○○○
○○○
●○○

References
○○

Advocacy

Advocacy
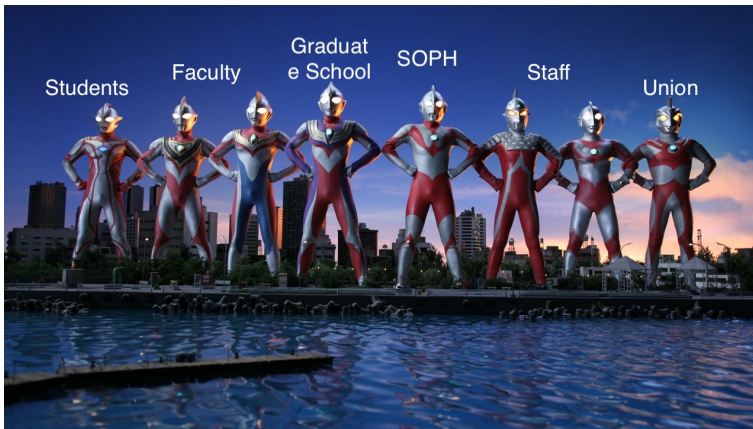
Outline
Background
Dissertation
Conclusion
References

Advocacy

Advocacy

Outline
○○

Background
○○○
○○○○○
○○○○○

Dissertation
○○○○
○○○○○○○○○○○○○
○○○○○

Conclusion
○○○○
○○○○
○○○
○○●

References
○○

Advocacy

Advocacy

Outline
OO

Background
O
OOOOO
OOOO
OOOOO

Dissertation
OOO
OOOOOOOOOOOOOO
OOOOO

Conclusion
O
OOOO
OO
OOO

References
●O

References

Outline
00

Background
00000
00000
00000

Dissertation
000
0000000000000
00000

Conclusion
0000
000
000

References
0●

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.

Baladandayuthapani, Veerabhadran, Bani K. Mallick, and Raymond J. Carroll. 2005. "Spatially Adaptive Bayesian Penalized Regression Splines (p-Splines)." *Journal of Computational and Graphical Statistics* 14 (2): 378–94.

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.

Outline
○○

Background
○○○○○
○○○○○

Dissertation
○○○
○○○○○○○○○○○○
○○○○○

Conclusion
○○○○
○○○
○○○

References
○●

References II

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

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

Lang, Stefan, and Andreas Brezger. 2004. "Bayesian p-Splines." *Journal of Computational and Graphical Statistics* 13 (1): 183–212. https://doi.org/10.1198/1061860043010.

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.

Outline
○○

Background
○○○○○
○○○○○

Dissertation
○○○
○○○○○○○○○○○○○
○○○○○

Conclusion
○○○○
○○○
○○○

References
○●

References III

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.

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.

Outline
○○

Background
○○○○○
○○○○○

Dissertation
○○○
○○○○○○○○○○○○○
○○○○○

Conclusion
○○○○
○○○
○○○

References
○●

References IV

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

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.

Outline
○○

Background
○○○○○
○○○○○

Dissertation
○○○
○○○○○○○○○○○○○
○○○○○

Conclusion
○○○○
○○○
○○○

References
○●

References V

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.

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