Low Rank Functional Analysis-of-Variance Modeling

Libin Liang

Department of Statistics
Rutgers University

April 9, 2021

Overview

- Background
- Model Introduction
- Fitting Method
- Sub-problem Fitting Algorithms
- Preliminary Numerical Results
- Future Work

Background

- Additive Model(Stone, 1986)
  y_i is a response and p-dimensional convariate vector $x_i = (x_{i1}, ..., x_{ip})$

$$y_i = \mu + \sum_{j=1}^{p} f_j(x_{ij}) + \epsilon_i$$

- Functional analysis of variance (ANOVA) modeling

$$Y_i = \mu + \sum_{1 \le j \le p} f_j(x_{ij}) + \sum_{1 \le j1 \le j_2} f_{j_1, j_2}(x_{ij_1}, x_{ij_2}) + ... + \sum_{1 \le j_1 \le ... \le j_K \le p} f_{j_1, ..., j_K}(x_{ij_1}, ..., x_{ij_K}) + \epsilon_i$$

Background

Problem of Increasing Dimensionality

Basic: $\phi_{j_k}^{m}(x_{j_k}) \quad m = 1, ..., p$

$$f_{j_1,...,j_K}(x_{ij_1}, ..., x_{ij_K}) = \sum_{1 \le m_1 \le ... \le m_K \le p} \beta_{m_1,...,m_K} \prod_{k=1}^{K} \phi_{j_k}^{m_k}(x_{j_k})$$

$$|\{\beta_{m_1,...,m_K}\}| = p^K.$$

Model Introduction

- Low Rank ANOVA modeling(Suppose we have 3 variables x_1, x_2, x_3)

$$Y_i = \mu + \sum_{j=1}^{3} f_j(x_{ij}) + f_1^{(1,2)}(x_{i1}) f_2^{(1,2)}(x_{i2}) + f_1^{(1,3)}(x_{i1}) f_3^{(1,3)}(x_{i3}) + f_2^{(2,3)}(x_{i2}) f_3^{(2,3)}(x_{i3})$$
$$+ f_1^{(1,2,3)}(x_{i1}) f_2^{(1,2,3)}(x_{i2}) f_3^{(1,2,3)}(x_{i3})$$

Dimensionality

$$\phi_{j_k}^{m}(x_{j_k}), \quad m = 1, ..., p$$

$$\prod_{k=1}^{3} f_k^{(1,2,3)}(x_{ik}) = \sum_{1 \le m_1 \le ... \le m_K \le p} \prod_{k=1}^{3} \beta_{m_k} \phi_{j_k}^{m_k}(x_{j_k})$$

$$|\{\beta_{m_k}\}| = p * 3$$

Model Introduction

- Penalty
  - Total Variation: $TV(f^{(m-1)})$
  - Empirical Norm: $||f||_n = \sqrt{\frac{\sum_{i=1}^n f^2(x_i)}{n}}$

Truncated Power Basics:

$$\phi_k(x) = x^k, k = 1, 2, ..., m-1, \quad \phi_k(x) = (z - t_{(j)})_+^{(m-1)}, \forall j = 1, ..., p-m+1$$

$\Phi_j = (\phi_{1j}, ..., \phi_{pj}) \in \mathcal{R}^{n \times p}$, $f_j(x_j) = \Phi_j \beta_j \in \mathcal{R}^n$. and

$$TV(f_j^{(m-1)}) = ||D\beta_j||_1$$

and

$$||f_j||_n = \frac{1}{\sqrt{n}}||\Phi_j \beta_j||_2$$

Model Fitting

$$Y = \mu + \sum_{j=1}^{3}(\Phi_j\beta_j) + \sum_{1\leq j_1\leq j_2\leq 3}(\Phi_{j_1}\beta_{j_1}^{(j_1,j_2)}) * (\Phi_{j_2}\beta_{j_2}^{(j_1,j_2)}) + (\Phi_{j_1}\beta_1^{(1,2,3)}) * (\Phi_{j_1}\beta_2^{(1,2,3)}) * (\Phi_{j_1}\beta_3^{(1,2,3)}) + \epsilon_i$$

where $\Phi_j, \Phi_{j_k} \in \mathbb{R}^{n\times p}$. $*$ represent the Hadamard product $\begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix} * \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix} = \begin{pmatrix} a_1 b_1 \\ a_2 b_2 \\ \vdots \\ a_n b_n \end{pmatrix}$

Squared loss:

$$||Y - (\mu + \sum_{j=1}^{3}(\Phi_j\beta_j) + \sum_{1\leq j_1\leq j_2\leq 3}(\Phi_{j_1}\beta_{j_1}^{(j_1,j_2)}) * (\Phi_{j_2}\beta_{j_2}^{(j_1,j_2)}) + (\Phi_{j_1}\beta_1^{(1,2,3)}) * (\Phi_{j_1}\beta_2^{(1,2,3)}) * (\Phi_{j_1}\beta_3^{(1,2,3)}))||^2$$

Model Fitting

With truncated power basic, if we fixed others paramters and try to update a specific component , such as $\beta_1^{(12)}$, then we are trying to solve the following subproblem.

$$\frac{1}{2}||Residual - \big(Diag(\Phi\beta_2^{(1,2)})\Phi_1\big)\beta||^2 + \rho||D\beta||_1 + \lambda\frac{1}{\sqrt{n}}||\Phi_1\beta||_2 \tag{1}$$

is convex w.r.t. β.

Initialization: $\beta_1, \beta_2, ..., \beta_1^{(123)}, ..., \beta_3^{(123)}$
while *Not converge* **do**
  1.Iteratively choosing a component β *from*$\{\beta_1, \beta_2, ..., \beta_1^{(123)}, ..., \beta_3^{(123)}\}$
  2.Fixed other parameters except β, then minimize the objective function w.r.t. β by solving problem that is similar as (1).
end

Sub-problem Fitting Algorithms

Objective function:

$$\frac{1}{2}||Y - Kx||^2 + \rho||Dx||_1 + \lambda||Mx||_2 \qquad (2)$$

Chambolle-Pock(2010)

$$G(x) + H(Lx) \qquad (3)$$

where $x \in \mathcal{R}^n$ and $L \in \mathcal{R}^{m \times n}$

Conjugated version:

$$G(x) + < y, Lx > -H^*(y) \qquad (4)$$

where $H^*(y) = min_x < y, x > -H(x)$
(x^*, y^*) is a mimizer of (3) iff (x^*, y^*) is a saddle point of (4).

Sub-problem Fitting Algorithms

Initialization: $x_0 \in \mathcal{R}^n, y_0 \in \mathcal{R}^{n\times m}$, t=0, $\tau, \sigma > 0$

while *Not converge* **do**

1. $x_{t+1} = prox_{\tau G}(x_t - \tau L^T y_t)$

2.$y_{t+1} = prox_{\sigma H^*}(y_t + \sigma(2Lx_{t+1} - Lx_t))$

end

Algorithm 1: Chambolle-Pock

where proximal operator: $Prox_{\sigma F}(x) = argmin_a \; F(a) + \frac{1}{2\sigma}||a - x||^2$.

Let $L = \begin{pmatrix} K \\ M \end{pmatrix} \in \mathcal{R}^{2m\times n}$ and $G(x) = \rho||Dx||_1$, $H(Lx) = \frac{1}{2}||Y - Kx||^2 + \lambda||Mx||_2$.

1. $x_{t+1} = argmin_x \; \frac{1}{2\tau}||x - (x_t - \tau L^T y_t)||^2 + \rho||Dx||_1$

2. $y_{t+1} = \begin{pmatrix} \frac{-Y}{1+1/\sigma} + \frac{1/\sigma y_{1t}}{1+1/\sigma} \\ Proj(y_{2t}, \lambda) \end{pmatrix}$

Where $Proj(x, \lambda)$ means project x on spherical $B(0, \lambda)$.

Preliminary Numerical Result

$$Y_i = \mu + \sum_{j=1}^{3} f_j(x_{ij}) + \sum_{1 \le j_1 \le j_2 \le 3} f_{j_1}^{(j_1,j_2)}(x_{ij_1}) f_{j_2}^{(j_1,j_2)}(x_{ij_2}) + \epsilon_i \tag{5}$$

Sample size $n = 500$.
The true functional components: Truncated Power Basic; Knots=3 or 4, m=2.
Probability for parameters is set as zero: 0.2
Noise Level: SD[Y]*0.05

Model 1:

$$y_i = \mu + \sum_{j=1}^{3} (\Phi_{ij}\beta_j) + \sum_{1 \le j_1 \le j_2 \le 3} (\Phi_{ij_1}\beta_{j_1}^{(j_1,j_2)}) * (\Phi_{ij_2}\beta_{j_2}^{(j_1,j_2)}) + \epsilon_i$$

With total variation of the m-1 order derivation penalty.
Model 2:

$$y_i = \mu + \sum_{j=1}^{3} (\Phi_{ij}\beta_j) + \sum_{1 \le j_1 \le j_2 \le 3} \Phi_{i,(j_1,j_2)}\beta_{(j_1,j_2)} + \epsilon_i$$

where $\Phi_{i,(j_1,j_2)} \in \mathbb{R}^{1 \times p^2}$. With lasso penalty.

Preliminary Numerical Result

Estimate Error: $\sum_{j=1}^{3} ||\beta_j - \beta_j^{true}||^2 + \sum_{1\le j_1,j_2\le 3} ||\beta_{(j_1,j_2)} - \beta_{(j_1,j_2)}^{true}||^2$

where $\beta_{(j_1,j_2)}^{true} = \beta_{j_1}^{(j_1,j_2)true} \diamond \beta_{j_2}^{(j_1,j_2)true} \in \mathbb{R}^{p\times p}$

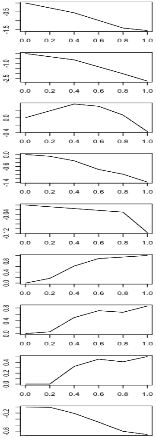
And $\beta_{(j_1,j_2)} = \beta_{j_1}^{(j_1,j_2)} \diamond \beta_{j_2}^{(j_1,j_2)} \in \mathbb{R}^{p\times p}$ for low rank model.

Error:Mean(SD)	3 knots	4 knots
Low rank model(Repeat 10 times)	0.761(0.409)	22.277(4.513)
Full Model	0.978	43.256

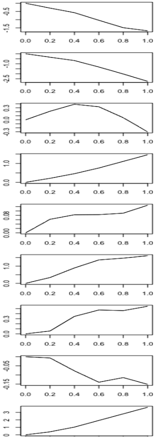
Table: Estimate Error comparison

Preliminary Numerical Result

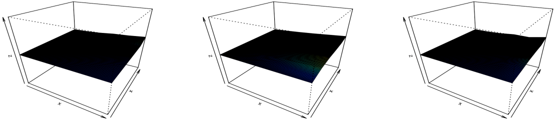
Number of Knots = 4, Estimate Error is around the medium in the 10 times repetition.



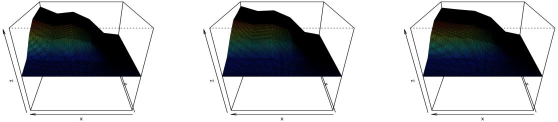
(a) True Components



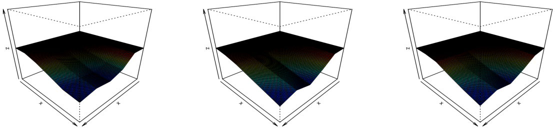
(b) Output Component(Low rank)



(a) x1 and x2



(b) x1 and x3



(c) x2 and x3

Figure: Compare of Interaction effect(From right to left: True, Low rank output, Full model output)

Future Work

- Improve the fitting algorithm
- Theoretical analysis of the model
- Extended from Rank-1 to Rank-R $\quad f_1(x_1)f_2(x_2) \Rightarrow \sum_{r=1}^{R} f_1^r(x_1)f_2^r(x_2)$