再生核空间中的函数型岭回归及其收敛率

万润哲

数学与应用数学
数学科学学院
复旦大学

December 24, 2016

Outline

Functional Data

- 来源: 信息化时代(空间统计, 医学图像)
- 特点：函数视为数据单位
- 思想：重复性 + 正则性
- 应用：降维, 分类, 聚类, 回归

Functional Data

- 来源: 信息化时代（空间统计，医学图像）
- 特点：函数视为数据单位
- 思想：重复性 + 正则性
- 应用：降维, 分类, 聚类, 回归

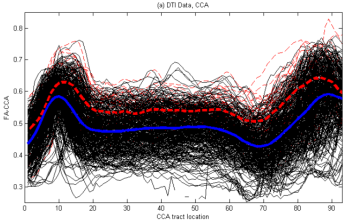
Functional Data

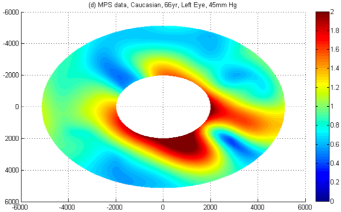
- 来源: 信息化时代（空间统计，医学图像）
- 特点：函数视为数据单位
- 思想：重复性 + 正则性
- 应用：降维, 分类, 聚类, 回归

Functional Data

- 来源: 信息化时代 (空间统计, 医学图像)
- 特点 : 函数视为数据单位
- 思想 : 重复性 + 正则性
- 应用 : 降维, 分类, 聚类, 回归

Examples


(a) DTI Data, CCA
FA-CCA
CCA tract location


(d) MPS data, Caucasian, 66yr, Left Eye, 45mm Hg

Outline

Outline

Model&Objective

$$Y_i = \alpha_0 + \int_T X_i(t) b(t) dt + \epsilon_i \tag{1}$$

Y_i: 实值响应变量,$X_i(t)$:T 上的随机过程

- 预测函数:$F_0(X(t)) = \alpha_0 + \int_T X_i(t) b(t) dt$
- 溢出风险:$\epsilon(\hat{F}_0) := E[F_0(X(t)) - \hat{F}_0(X(t))]^2$
- 经验溢出风险 : $\hat{\epsilon}(\hat{F}_0) := 1/n * \sum_{i=1}^{n} (y_i - \hat{F}_0(x_i))^2$

Model&Objective

$$Y_i = \alpha_0 + \int_T X_i(t)b(t)dt + \epsilon_i \tag{1}$$

Y_i: 实值响应变量,$X_i(t)$:T 上的随机过程

- 预测函数:$F_0(X(t)) = \alpha_0 + \int_T X_i(t)b(t)dt$
- 溢出风险:$\epsilon(\hat{F}_0) := E[F_0(X(t)) - \hat{F}_0(X(t))]^2$
- 经验溢出风险 : $\hat{\epsilon}(\hat{F}_0) := 1/n * \sum_{i=1}^n (y_i - \hat{F}_0(x_i))^2$

Model&Objective

$$Y_i = \alpha_0 + \int_T X_i(t)b(t)dt + \epsilon_i \tag{1}$$

Y_i: 实值响应变量,$X_i(t)$:T 上的随机过程

- 预测函数:$F_0(X(t)) = \alpha_0 + \int_T X_i(t)b(t)dt$
- **溢出风险**:$\epsilon(\hat{F}_0) := E[F_0(X(t)) - \hat{F}_0(X(t))]^2$
- 经验溢出风险：$\hat{\epsilon}(\hat{F}_0) := 1/n * \sum_{i=1}^n (y_i - \hat{F}_0(x_i))^2$

Model&Objective

$$Y_i = \alpha_0 + \int_T X_i(t) b(t) dt + \epsilon_i \tag{1}$$

Y_i: 实值响应变量,$X_i(t)$:T 上的随机过程

- 预测函数:$F_0(X(t)) = \alpha_0 + \int_T X_i(t) b(t) dt$
- 溢出风险:$\epsilon(\hat{F}_0) := E[F_0(X(t)) - \hat{F}_0(X(t))]^2$
- **经验溢出风险**:$\hat{\epsilon}(\hat{F}_0) := 1/n * \sum_{i=1}^{n} (y_i - \hat{F}_0(x_i))^2$

Outline

1. 函数型数据
2. 函数型线性回归
   - 问题与模型
   - 方法综述
   - 再生核 Hilbert 空间及相关算子
   - 再生核空间中的函数型岭回归
3. 收敛率推导
   - 结论
   - 证明纲要
4. 算法与实验
   - 算法推导
   - 数值实验

- 正则性中的信息:$x_i(t), b(t)$
- 基函数的选取：样条、主成分、小波、傅里叶、核函数等
- 惩罚项的添加方法: 截断、粗糙惩罚项和稀疏惩罚项
- FPCR, PSR,FKRR 等

Methodology Review

- 正则性中的信息:$x_i(t), b(t)$
- 基函数的选取：样条、主成分、小波、傅里叶、核函数等
- 惩罚项的添加方法: 截断、粗糙惩罚项和稀疏惩罚项
- FPCR, PSR,FKRR 等

Methodology Review

- 正则性中的信息:$x_i(t), b(t)$
- 基函数的选取：样条、主成分、小波、傅里叶、核函数等
- 惩罚项的添加方法: 截断、粗糙惩罚项和稀疏惩罚项
- FPCR, PSR,FKRR 等

Methodology Review

- 正则性中的信息:$x_i(t), b(t)$
- 基函数的选取：样条、主成分、小波、傅里叶、核函数等
- 惩罚项的添加方法: 截断、粗糙惩罚项和稀疏惩罚项
- FPCR, PSR,FKRR 等

Outline

One Definition

Evaluation Functional

称 H 上的泛函 f 为 x 点处的取值泛函，如果对于所有 $g \in H, f(g) = g(x)$，记为 f_x

RKHS

称 $\mathcal{X}$ 为 RKHS，如果 $\forall x, f_x$ 在 H 中连续。

Reproducing Kernel

$$\forall f_x, \exists g_x \in H, s.t. \forall g \in H, f_x(g) =< g_x, g >_H$$

从而 $g_x(y) = g_y(x)$. 我们诱导出了 X 上的一个对称二元函数 $K(x, y) := g_x(y)$，称为 H 对应的再生核

One Definition

Evaluation Functional

称 H 上的泛函 f 为 x 点处的取值泛函，如果对于所有 $g \in H, f(g) = g(x)$，记为 f_x

RKHS

称 $\mathcal{X}$ 为 RKHS，如果 $\forall x, f_x$ 在 H 中连续。

Reproducing Kernel

$$\forall f_x, \exists g_x \in H, s.t. \forall g \in H, f_x(g) =< g_x, g >_H$$

从而 $g_x(y) = g_y(x)$. 我们诱导出了 X 上的一个对称二元函数 $K(x, y) := g_x(y)$，称为 H 对应的再生核

Operators

直观理解

外延广泛，性质良好，再生性，结构性的 Hilbert 空间

Integral Operator

$L_K : H \to H, g \mapsto \int_X g(x)K(x,*)d\rho x =< g, K_x >_2$

Properties

毕业论文报告

万润哲

再生性及相关性质

$< K(x,*), g >_H = g(x);$

$\forall f \in L_2, L_K(f) \in H;$

$\forall f, g \in H, < f, L_K(g) >_2 = < f, g >_H$

Representer Theorem

if $\hat{f} := Argmin_{f \in H} E_f((x_1, y_1), (x_2, y_2), ..., (x_n, y_n)) + J(\|f\|_H)$

then $\exists \vec{l} \in R^n, s.t. \tilde{f} = \sum_{i=1}^{n} l_i K(*, x_i)$

Outline

Estimator

$$(\hat{\alpha_0}, \hat{b}(t)) := Argmin_{\alpha \in R, b \in H}\{\hat{\epsilon}(\alpha, b) + \lambda * \|b\|_H^2\} \tag{2}$$

- 外延广泛：各种斜率函数和惩罚项
- 性质优良：无穷维, 非参数问题-> 有限维
- 岭惩罚项: 光滑性与可解释性

Estimator

$$(\hat{\alpha_0}, \hat{b}(t)) := Argmin_{\alpha \in R, b \in H}\{\hat{\epsilon}(\alpha, b) + \lambda * \|b\|_H^2\} \qquad (2)$$

- 外延广泛：各种斜率函数和惩罚项
- 性质优良：无穷维, 非参数问题-> 有限维
- 岭惩罚项: 光滑性与可解释性

Estimator

$$(\hat{\alpha_0}, \hat{b}(t)) := Argmin_{\alpha \in R, b \in H}\{\hat{\epsilon}(\alpha, b) + \lambda * \|b\|_H^2\} \qquad (2)$$

- 外延广泛：各种斜率函数和惩罚项
- 性质优良：无穷维, 非参数问题-> 有限维
- 岭惩罚项: 光滑性与可解释性

Properties

目标函数约化

$1: \hat{\alpha}_0 = \bar{y} - \int_T \bar{x}\hat{b}dt$

$2: \hat{b}(t) := Argmin_{b \in H}\{\frac{1}{n}\sum_{i=1}^{n}[\tilde{y}_i - \int_T \tilde{x}_i b dt]^2 + \lambda\|b\|_H^2\}$

斜率函数估计量表示定理

对于以上估计量，存在 $c_i \in R, i = 1,,,n$ 使得

$$\hat{b}(t) = \sum_{i=1}^{n} c_i L_K(x_i)$$

Outline

Outline

Aim

$$P(\epsilon(\hat{b}) \leq A[2\frac{Bln(2/\delta)[\frac{k}{\sqrt{n\lambda}} + \sqrt{\Psi}]}{\sqrt{n}} + C]^2) \geq 1 - \delta \tag{3}$$

$$A_n : (1 + \Lambda_n + \Lambda_n^2) \tag{4}$$

$$B := (\|Y\|_\infty + \|f_\lambda\|_\infty) \tag{5}$$

$$C := \sqrt{E(F_0^2)} \tag{6}$$

Outline

算子表示

$$\epsilon(\hat{b}) = \|L_C^{\frac{1}{2}}(\hat{b} - b_0)\|_2^2 \tag{7}$$

$$b_\lambda = L_k^{\frac{1}{2}}(L_k^{\frac{1}{2}} L_C L_k^{\frac{1}{2}} + \lambda I)^{-1} L_k^{\frac{1}{2}} L_C b_0 \tag{8}$$

$$\hat{b}_\lambda = L_k^{\frac{1}{2}}(L_k^{\frac{1}{2}} L_{C_n} L_k^{\frac{1}{2}} + \lambda I)^{-1} L_k^{\frac{1}{2}}\left(\frac{\sum_{i=1}^n y_i x_i(t)}{n}\right) \tag{9}$$

偏差与方差

$$\epsilon(\hat{b}) \leq \left(\|L_C^{\frac{1}{2}}(\hat{b} - b_\lambda)\|_2 + \|L_C^{\frac{1}{2}}(b_0 - b_\lambda)\|_2\right)^2 \tag{10}$$

Theorem (偏差部分的上界)

$$\|L_C^{\frac{1}{2}}(b_0 - b_\lambda)\|_2 \leq \sqrt{E(F_0^2)} \tag{11}$$

符号与方差部分变形

符号

$S_n := L_K L_{C_n}$

$S := L_K L_C$

$\xi_\lambda(z_i) := (y_i - f_\lambda(x_i))L_K(x_i)$

$\bar{\xi}_\lambda := \frac{\sum_{i=1}^n \xi_\lambda(z_i)}{n}$

方差部分变形

$$\|L_C^{\frac{1}{2}}(\hat{b} - b_\lambda)\|_2 = \|S_n^{\frac{1}{2}}\{(S_n + \lambda I)^{-1}[\bar{\xi}_\lambda - E(\xi_\lambda)]\}\|_H$$

一些引理

毕业论文报告
万润哲
函数型数据
函数型线性回归
问题与模型
方法综述
再生核 Hilbert 空间及相关算子
再生核空间中的函数型岭回归
收敛率推导
结论
证明纲要
算法与实验
算法推导
数值实验

算子二阶分解

$$(S_n+\lambda I)^{-1}-(S+\lambda I)^{-1}=(S+\lambda I)^{-1}(S-S_n)(S+\lambda I)^{-1}+(S+\lambda I)^{-1}(S-S_n)(S_n+\lambda I)^{-1}(S-S_n)(S+\lambda I)^{-1}$$

算子范数不等式

$$\|L_K^{\frac{1}{2}}(L_K+\lambda I)^{-\frac{1}{2}}\|\leq 1$$
$$\|(L_K+\lambda I)^{-1}\|\leq\frac{1}{\lambda}$$
$$\|(L_K+\lambda I)^{-\frac{1}{2}}\|\leq\frac{1}{\sqrt{\lambda}}$$

概率不等式

$$P(\|\bar{\xi}-E(\xi)\|\leq\frac{2M\ln(2/\delta)}{n}+\sqrt{\frac{2c\ln(2/\delta)}{n}})\geq 1-\delta$$

主定理符号与思路

符号

$\xi_0(z_i) := (y_i - f_0(x_i))L_K(x_i)$

$\Delta_\lambda := \bar{\xi}_\lambda - E(\xi_\lambda); \quad \Delta_0 := \bar{\xi}_0 - E(\xi_0) = \bar{\xi}_0; \; \Delta := \Delta_\lambda - \Delta_0$

$\theta_n := \|(S + \lambda I)^{-1/2}(S - S_n)\|; \; \Theta_n := 1 + \frac{\theta_n}{\sqrt{\lambda}} + \frac{\theta_n^2}{\lambda}$

$\Psi(\lambda) := Tr((L_K L_C + \lambda I)^{-1} L_K L_C); \qquad \Omega_n := \frac{2k^2}{n\sqrt{\lambda}} + \frac{2k\sqrt{\Psi}}{\sqrt{\lambda}}$

$Q_1 := (S_n + \lambda I)^{-1} - (S + \lambda I)^{-1}; \qquad Q_2 := S - S_n$

$S_\lambda := S + \lambda I; \qquad S_{n\lambda} := S_n + \lambda I$

主定理证明思路

对算子进行二阶分解后，利用放缩进行简化，处理至算子差和随机变量差两部分，分别运用引理估计概率界。

Outline

Outline

Transform Optimization Problem

回顾

$\hat{b}(t) := Argmin_{b\in H}\{\frac{1}{n}\sum_{i=1}^{n}[\tilde{y}_i - \int_T \tilde{x}_i b dt]^2 + \lambda\|b\|_H^2\}$

$\hat{b}(t) = \sum_{i=1}^{n} c_i L_K(x_i)$

$$\vec{c} = Argmin_{\vec{l}\in R^n}\{\frac{1}{n}\sum_{i=1}^{n}[\tilde{y}_i - \int_T \tilde{x}_i(\sum_{j=1}^{n} l_j L_K(\tilde{x}_j))dt]^2 + \lambda\|\sum_{i=1}^{n} l_i L_K(\tilde{x}_i)|$$

$$= \ldots$$

$$= Argmin_{\vec{l}\in R^n}\{\frac{\|\vec{y}\|_2}{n} - \frac{2\vec{y}^T\Phi}{n}\vec{l} + \vec{l}^T(\frac{\Phi^T\Phi}{n} + \lambda\Gamma)\vec{l}\} \tag{12}$$

无约束二次优化问题。成熟的算法。类似样条光滑。

Algorithm

step0（预处理）：对 $X_i(t)$ 的离散测量值 $X_i(t_j)$，选取基进行光滑化，得到 $X_i(t)$
step1（中心化）：计算样本平均 $\bar{x}_i(t)$ 和 $\bar{y}_i$，对数据中心化 $\tilde{x}_i(t) = x_i(t) - \bar{x}_i(t)$,$\tilde{y}_i$ 同理
step2(积分算子变换)：通过数值积分，计算 $L_K(\tilde{x}_i(t))$，向量化
step3(计算矩阵): 利用数值积分，即 H 上具体的内积定义，计算 Φ 和 Γ
step4(整理): 利用矩阵乘法，计算
$a_1 = \frac{\|\vec{y}\|_2}{n}, \vec{a_2} = \frac{2\vec{y}^T\Phi}{n}, A = \frac{\Phi^T\Phi}{n} + \lambda\Gamma$
step5(求解无约束二次优化)：利用现有算法，求解 $Argmin_{\vec{l}\in R^n}\{a_1 - \vec{a_2}^T\vec{l} + \vec{l}^T A\vec{l}\}$
step6(回带)：分别回带，得到 $\hat{\alpha_0}$,$\hat{b}(t)$ 与 $\hat{F}_0(X(t))$

Outline

TTI Data Set

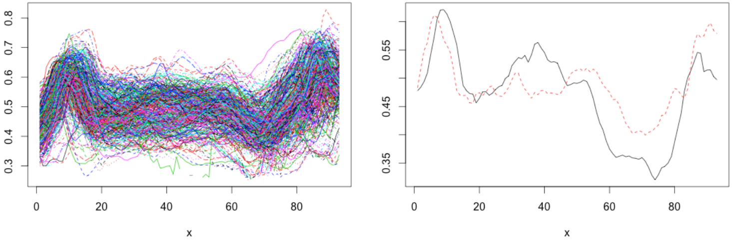


Figure 1: Functional Data

Choose RKHS

由于没有特定的周期性，我们利用 DTI 自身具有的光滑性，采用以下的 RKHS。$\hat{H} = \mathcal{W}_2^2$ 为二阶 Sobolev 空间，赋以常见的范 $\forall f \in \hat{H} : \|f\| = \|f'\|_2^2$ 记 $\hat{H}$ 商去使 $\|f\| = 0$ 的子空间 $H1$，得到 RKHS H，范数保持。记 $B_m(t)$ 为第 m 个 Bernoulli 多项式，则其对应的再生核为

$$K(s,t) = \frac{1}{4}B_2(s)B_2(t) - \frac{1}{24}B_4(|s-t|) \tag{13}$$

Result

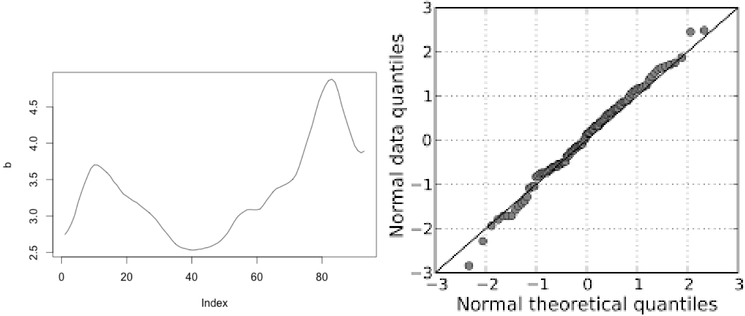


Figure 2: Estimate and Test

Thanks for your attention!