Spline Basics - ECNU Seminar Notes

Bu Zhou

2011,6,21

1 Three different kinds of splines

$y = u(t) + \epsilon,$
$\epsilon \sim (N)(0, \sigma^2 I_n);$
$y_i = u(t_i) + \epsilon_i,$
$\epsilon_i \sim (N)(0, \sigma^2);$

1.1 Interpolating<->Smoothing:

1.2 Different Splines

Linear Regression-> Polynomial Regression-> Orthogonal Polynomial Regression-> Piecewise Polynomial Regression

Linear Regression	Polynomial Regression
$Y = X\beta + \epsilon$;$y = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}$,$\beta = \begin{pmatrix} \beta_0 \\ \vdots \\ \beta_t \end{pmatrix}$	
$X_L = \begin{pmatrix} 1 & X_{11} & \cdots & X_{1t} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & X_{n1} & \cdots & X_{nt} \end{pmatrix}$	$X_P = \begin{pmatrix} 1 & x_1 & x_1^2 & \cdots & x_1^t \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_n & x_n^2 & \cdots & x_n^t \end{pmatrix}$
$M(\beta) = RSS(\beta) = \Sigma_{i=1}^n (y_i - x_i\beta)^2$	$M(\beta) = RSS(\beta)$

Orthogonal Polynomial Regression
$X_O = \begin{pmatrix} 1 & \varphi_1(x_1) & \varphi_2(x_1) & \cdots & \varphi_t(x_1) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & \varphi_1(x_n) & \varphi_2(x_n) & \cdots & \varphi_t(x_n) \end{pmatrix}$
$X'X$ is a diagonal matrix, for computation convenience
$M(\beta) = RSS(\beta)$

Piecewise Polynomial Regression
partition the domain of t into consequential intervals(add knots), in each interval use a polynomial to approximate $f(()_+)$
$M = (X_P Z)$;$Z = \begin{pmatrix} (x_1 - \xi_1)_+^{m-1} \; 1 & \cdots & (x_1 - \xi_k)_+^{m-1} \\ \vdots & \ddots & \vdots \\ (x_n - \xi_1)_+^{m-1} & \cdots & (x_n - \xi_k)_+^{m-1} \end{pmatrix}$;$x_+^r = \{max(0, x)\}^r$
need to select knots(number&location)
$M(\beta) = RSS(\beta)$

Smoothing Spline
$Y = X\beta_P + Zu + \epsilon = \beta B(x) + \epsilon$
$B(x) = [B_1(x), \cdots, B_N(x)]^T$, vector of spline basis functions
knots are data points, need to select α
$M(\beta) = \Sigma_{i=1}^n \{y_i - \beta^T B(x_i)\}^2 + \alpha\beta^T D\beta; D = \int_a^b B^{(q+1)}(x)\{B^{(q+1)}(x)\}^T dx$

Penalized Spline
$Y = X\beta_P + Zu + \epsilon = \beta B(x) + \epsilon$
need to select knots and α
$M(\beta) = \Sigma_{i=1}^n \{y_i - \beta^T B(x_i)\}^2 + \alpha\beta^T D\beta$

2 Regression Spline/Least-square Spline

$u(t) \sim S(t) = \Sigma_{i=1}^m \theta_j t^{j-1} + \Sigma_{j=1}^k \delta_j (t - \xi_j)_+^{m-1}$
order: m
knots: $\xi_1, \cdots, \xi_k$: $\xi = \{\xi_1, \cdots, \xi_k\}$
coefficients: $\delta_1, \cdots \delta_k$

2.1 least square spline estimator of u

$u = \Sigma_{j=1}^{m+k} b_{\xi j} x_j$, where
$x_j(t) = t^{j-1}$, $j = 1, \cdots, m$
$x_{m+j}(t) = (t - \xi_j)_+^{m-1}$, $j = 1, \cdots, k b_\xi$: the estimator of
$\beta = (\theta_1, \cdots, \theta_m, \delta_1, \cdots, \delta_k)^T$,
obtained by minimizing $RSS(c; \xi) = \Sigma_{i=1}^n (y_i - \Sigma_{j=1}^{m+k} c_j x_j(t))^2$,
with respect to $c = (c_1, \cdots, c_{m+k})^T$
$X_\xi^T X_\xi c = X_\xi^T Y$,$X_\xi = \{x_j(t_i)\} i = 1, \cdots, n_j; j = 1, \cdots, m + k;$
$b_\xi = (X_\xi^T X_\xi)^{-1} X_\xi Y$

2.2 selecting ξ:

number and location of the knots for the estimator(m=2,3,(4))

1. visual inspection of the data: more knots will be needed where u seems to change more rapidly

2. variable projection method: minimizing $RSS(\xi) = \Sigma_{i=1}^n (y_i - \Sigma_{j=1}^{m+k} b_{\xi j} x_j(t_j))^2$, $\xi = \{\xi_1, \cdots, \xi_k\}$arbitrary choice for knots set

3. GCV: data driven choice for both location and number of knots
$\xi(K) = \{\xi_1, \cdots, \xi_K\}$, $GCV[\xi(K)] = \frac{n^{-1} RSS[\xi(K)]}{[1-(m+K)/n]^2}$
fixed K ->$\hat{\xi}(K)$ ->$\hat{\xi}(\hat{K})$ -> ...

4. MDL

2.3 Do not admit kernel or series representations

2.4 Large sample properties:

[Shanggang Zhou, Douglas A. Wolfe, On derivative estimation in spline regression]

3 1.2 Smoothing Spline

$y_i = u(x_i) + \epsilon_i$
$M(\lambda) = \Sigma[y_i - \hat{u_n}(x_i)]^2 + \lambda J(u)$
$J(u)$: roughness penalty(regulation), $= \int[u''(x)]^2 dx$

3.1 Definition:

Knots: $\xi_1 < \xi_2 < \cdots < \xi_k$ be a set of ordered points contained in some interval (a, b).

Mth-order spline: a piecewise$M-1$ degree(=order-1) polynomial with $M-2$ continuous derivatives at the knots.

Cubic spline($M = 4$): a continuous function u such that (i) u is a cubic polynomial over $(\xi_1, \xi_2), ...$ and (ii) u has continuous first and second derivatives at the knots.

Natural Spline: a spline that is linear beyond the boundary knots.

3.2 Theorem:

The function $u_n(x)$ that minimizes $M(\lambda)$ with penalty $\int[u''(x)]^2 dx$ is a natural cubic spline with knots at the data points(so there is no need to select the knots, all knots are data points, number of knots are the number of observations, this is the difference from regression splines). The estimator u_n is called a smoothing spline.

3.3 Basis:

To give an explicit form for $u_n$we need to construct a basis for the set of splines.

[Proofs that they are basis: de Boor, A Practical Guide to Splines (revised edition) (2001)]

Truncated power basis: (convenience for representation)

A basis for the set of cubic splines at knots ξ: $h_1, ..., h_{k+4}$: $h_1(x) = 1, h_2(x) = x, h_3(x) = x^2, h_4(x) = x^3, h_j(x) = (x - \xi_j - 4)^3_+$ for $j = 5, ..., k+4$.

$u(x) = \Sigma_{j=1}^{k+4} \beta_j h_j(x)$.

B-spline basis: (convenience for computation)

3.4 Estimation:

$\hat{u_n}(x) = \Sigma_{i=1}^{N} \hat{\beta}_j B_j(x)$, $B_1, \cdots, B_N$: natural spline basis
$M(\beta) = (Y - B\beta)^T (Y - B\beta) + \lambda \beta^T \Omega \beta$,
$\beta = (\hat{\beta_1}, \cdots, \hat{\beta_N})^T$
$B_{ij} = B_j(x_i)$
$\Omega_{jk} = \int B_j^{''}(x) B_k^{''}(x) dx$
we have
$\hat{\beta} = (B^T B + \lambda \Omega)^{-1} B^T Y$

3.5 Selection of λ

[Randall L. Eubank, Nonparametric Regression and Spline Smoothing, P239]

1. minimization of unbiased risk criterion $\hat{P}(\lambda) = \frac{1}{n} RSS(u_\lambda) + \frac{2\hat{\sigma}}{n} tr S_\lambda$, $S_\lambda = X(B^T B + \lambda \Omega)^{-1} B^T$, $\hat{\sigma}$:GSJS estimator

2. $CV = n^{-1} \Sigma_{i=1}^{n} [y_i - u_{\lambda(i)} t(i)]^2$, $u_{\lambda(i)}$: estimator when ith observation (t_i, y_i) deleted from data

3. $GCV = \frac{nRSS(u_\lambda)}{tr(I - S_\lambda)}$

4. MDL

3.6 Other properties

Admit kernel/series representations,

can be interpreted as minimum variance unbiased linear prediction pf a class of stochastic process[George S. Kimeldorf, Grace Wahba, Spline functions and stochastic processes, 1970].

3.7 Large sample properties:

[Randall L. Eubank, Nonparametric Regression and Spline Smoothing, P247]
$n \to \infty$, $\lambda \to 0$; $\lambda$is similar to the bandwidth h in kernel estimation
use a global, kernel type approximation for linear smoothing spline

3.8 Interpretation as splines with penalty:

[Randall L. Eubank, Nonparametric Regression and Spline Smoothing, P227]

3.9 Interpretation using Taylor's theorem:

[Randall L. Eubank, Nonparametric Regression and Spline Smoothing, P228]

4 1.3 Penalized Spline

[Eilers, Marx, Flexible smoothing with B-splines and penalties]: relatively large number of knots and a difference penalty on coefficients of adjacent B-splines.

differences from Smoothing Spline: (i) less knots (ii) general penalties

4.1 General Definition of a Penalized Spline:

is $\hat{\beta}^T B(x)$, where $\hat{\beta}$is the minimizer of $\Sigma_{i=1}^n \{y_i - \beta^T B(x_i)\}^2 + \alpha \beta^T D \beta$, for some symmetric positive semidefinite matrix D and scalar $\alpha > 0$.

4.2 Steps for applying penalized splines:

(1) Spline Model: degree, knot locations, constrains (such as natural spline boundary constrains);

(2) Penalty: form of penalty up to a nonnegtive smoothing parameter

(3) Basis functions: (truncated power functions or B-splines) used to represent the model, used in the computations; (do not affect the fitted curve)

(2+3) Penalty matrix D: automatically determined once the Penalty(form) and Basis functions have been determined.

4.3 Penalized splines in use:

1. Large sample properties:

[Gerda Claeskens, Tatyana Krivobokova, Jean D. Opsomer, Asymptotic properties of penalized spline estimators]

or

[Hulin Wu, Hongqi Xue, Arun Kumar, Numerical discretization-based estimation methods for ordinary differential equation models via penalized spline smoothing]

2. Working assumption: nonparameter estimation->parameter estimation

Reason: [Yan Yu, David Ruppert, Penalized Spline Estimation for Partially Linear Single-Index Models]

[Yang Bai, Wing K. Fung, Zhong Yi Zhu, Penalized quadratic inference functions for single-index models with longitudinal data]

η_0:($R \to R$) unknown univariate function

suppose η_0 is a pth degree spline function with knots $\xi_1, \cdots, \xi_K$, spline basis $B(x) = (1, x, \cdots, x^p, (x - \xi_1)_+^p, \cdots, (x - \xi_K)_+^p)^T$

$\hat{\eta}_0 \approx B\delta_0$, is the best projection of the true smoothing function$\eta_0$on the space$B\delta$. δ_0 are spline coefficients, spline parameter. We call δ_0 the "true" parameter.

If there are other parameters(β) to estimate, we represent all the parameters as $\theta = (\beta^T \delta_0^T)^T$

4.4 selection of knots and λ

1. [David Ruppert, Selecting the Number of Knots For Penalized Splines]

[David Ruppert, M. P. Wand, R. J. Carroll, Semiparametric Regression]

2. MDL

4.5 Penalized spline and shrinkage:

[David Ruppert, M. P. Wand, R. J. Carroll, Semiparametric Regression, P74]

5 References:

5.1 Books:

Semiparametric Regression by David Ruppert, M. P. Wand, R. J. Carroll, 2003

Nonparametric Regression and Spline Smoothing, Second Edition, by Randall L. Eubank, 1999

All of Nonparametric Statistics, by Larry Wasserman, 2005

A Practical Guide to Splines (revised edition), by Carl de Boor, 2001

The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Second Edition, by Trevor Hastie, Robert Tibshirani, Jerome Friedman, 2009

5.2 Papers:

Flexible smoothing with B-splines and penalties, Eilers, Marx, Statistical Science, 1996

Spline functions and stochastic processes, George S. Kimeldorf, Grace Wahba, The Indian Journal of Statistics, 1970

On derivative estimation in spline regression, Shanggang Zhou, Douglas A. Wolfe, Statistica Sinica, 2000

Asymptotic properties of penalized spline estimators, Gerda Claeskens, Tatyana Krivobokova, Jean D. Opsomer, Biometrika, 2009,

Numerical discretization-based estimation methods for ordinary differential equation models via penalized spline smoothing, Hulin Wu, Hongqi Xue, Arun Kumar, submitted to the Annals of Statistics

Penalized Spline Estimation for Partially Linear Single-Index Models, Yan Yu, David Ruppert, Journal of the American Statistical Association, 2002

Penalized quadratic inference functions for single-index models with longitudinal data, Yang Bai, Wing K. Fung, Zhong Yi Zhu, Multivariate Analysis, 2009

Selecting the Number of Knots For Penalized Splines, David Ruppert, 2000