Module 2: Splines and Kernel Methods

Regression Splines, Smoothing Splines

STAT/BIOSTAT 527, University of Washington
Emily Fox
April 10th, 2014

©Emily Fox 2014

Backtrack a bit…

- Instead of just considering input variables *x* (potentially mult.), augment/replace with transformations = "input features"

- ***Linear basis expansions*** maintain linear form in terms of these transformations

$$f(x) = \sum_{m=1}^{M} \beta_m h_m(x)$$

← trans.

← linear in these transformations

- What transformations should we use?
  - $h_m(x) = x_m$ → linear model
  - $h_m(x) = x_j^2, \quad h_m(x) = x_j x_k$ → polynomial reg.
  - $h_m(x) = I(L_m \leq x_k \leq U_m)$ → piecewise constant
  - …

©Emily Fox 2014

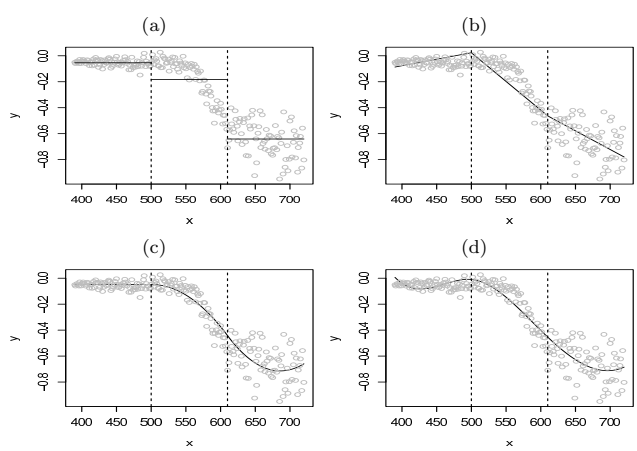
Piecewise Polynomial Fits

- Again, assume *x* univariate (multivariate x later)
- Polynomial fits are often good locally, but not globally
  - Adjusting coefficients to fit one region can make the function go wild in other regions
- Consider ***piecewise polynomial*** fits
  - Local behavior can often be well approximated by low-order polynomials

©Emily Fox 2014

Piecewise Polynomial Fits

LIDAR Data Example



From Wakefield book

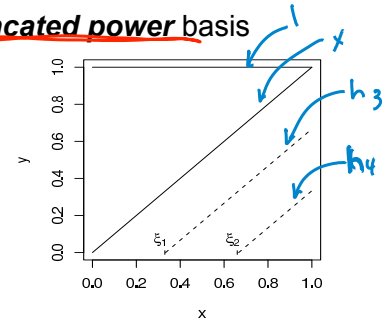
©Emily Fox 2014

Regression Splines – Linear

- More directly, we can use the ***truncated power*** basis

$$h_1(x) = 1$$
$$h_2(x) = x$$
$$h_3(x) = (x - \xi_1)_+$$
$$h_4(x) = (x - \xi_2)_+$$

knots specified by user



From Wakefield book

- Resulting model:

$$f(x) = \beta_0 + \beta_1 x + \beta_2 (x - \xi_1)_+ + \beta_3 (x - \xi_2)_+$$

- Continuous at the knots because all prior basis functions are contributing to the fit up to any single *x*

©Emily Fox 2014

Regression Splines – Cubic

- Naively, extend as

quadratic

$$f(x) = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 (x - \xi_1)_+ + \beta_4 (x - \xi_1)_+^2 + \beta_5 (x - \xi_2)_+ + \beta_6 (x - \xi_2)_+^2$$

before in piecewise linear

- But, 1st derivate is discontinuous (check this)
- Drop the truncated linear basis:

$$f(x) = \beta_0 + \beta_1 x + \beta_2 x^2 + b_1 (x - \xi_1)_+^2 + b_2 (x - \xi_2)_+^2$$

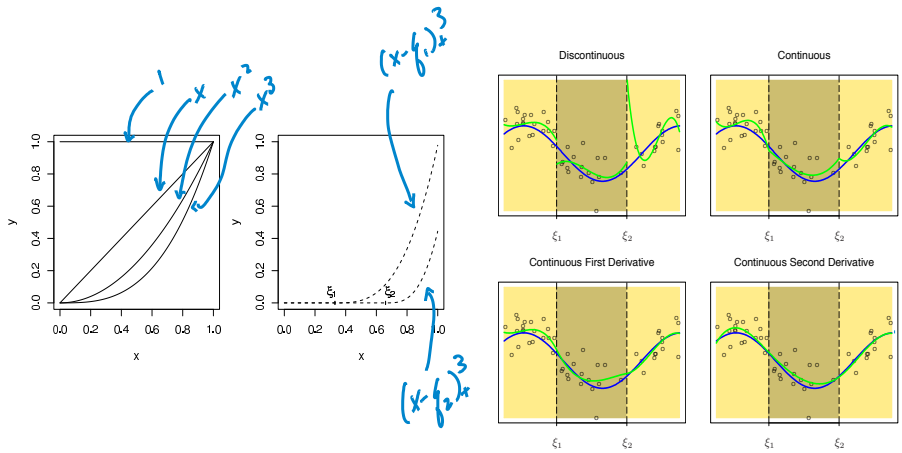
- Has continuous 1st derivative (check), but not 2nd
- Popular to consider ***cubic spline***:

$$\boxed{f(x) = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3 + b_1 (x - \xi_1)_+^3 + b_2 (x - \xi_2)_+^3}$$

- Has continuous 1st and 2nd derivatives
- Typically people stop here … smooth enough

©Emily Fox 2014

Cubic Spline Basis and Fit



Cubic Splines as Linear Smoothers

- Cubic spline function with *K* knots:

$$f(x) = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3 + \sum_{k=1}^{K} b_k (x - \xi_k)_+^3$$

- Simply a linear model $f(x) = E[Y|c] = c\gamma$

$$C = \begin{bmatrix} 1 & x_1 & x_1^2 & x_1^3 & (x_1-\xi_1)_+^3 & \cdots & (x_1-\xi_K)_+^3 \\ 1 & \vdots & & & & & \vdots \\ \vdots & & & & & \ddots & \\ 1 & x_n & x_n^2 & x_n^3 & (x_n-\xi_1)_+^3 & \cdots & (x_n-\xi_K)_+^3 \end{bmatrix} \qquad \gamma = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \\ b_1 \\ \vdots \\ b_K \end{bmatrix}$$

- Estimator:

$$\hat{\gamma} = (C^T C)^{-1} C^T y$$

- Linear smoother: $\hat{f} = \underbrace{C(C^T C)^{-1} C^T}_{L \text{ hat matrix}} y$

Natural Cubic Splines

- For polynomial regression, fit near boundaries is erratic.
  - Problem is worse for splines: each is fit locally so no global constraint
- ***Natural cubic splines*** enforce linearity beyond boundary knots
- Starting from a cubic spline basis, the natural cubic spline basis is

$$N_1(x) = 1 \qquad N_2(x) = x \qquad N_{k+2}(x) = d_k(x) - d_{K-1}(x)$$

$$d_k(x) = \frac{(x-\xi_k)_+^3 - (x-\xi_K)_+^3}{\xi_K - \xi_k}$$

- Derivation HW 3

©Emily Fox 2014

Regression Splines – Summary

- Definition:

  *An **order-M spline** with knots $\xi_1 < \xi_2 < \cdots < \xi_K$ is a piecewise M-1 degree polynomial with M-2 continuous derivatives as the knots*

  *A spline that is linear beyond the boundary knots is called a **natural spline***

- Choices:
  - Order of the spline
  - Number of knots
  - Placement of knots

  requires some thought

©Emily Fox 2014

Return to Smoothing Splines

- Objective:

$$\min_{f} \sum_{i=1}^{n} (y_i - f(x_i))^2 + \lambda \int f''(x)^2 dx$$

all fcns in the world

- Solution:
  - ***Natural cubic spline*** ✓
  - Place knots at every observation location x_i ☆

solves this objective

- Proof: See Green and Silverman (1994, Chapter 2) or Wakefield textbook
- Notes:
  - Would seem to overfit, but penalty term shrinks spline coefficients toward linear fit
  - Will not typically interpolate data, and smoothness is determined by λ

Smoothing Splines

- Model is of the form: $f(x) = \sum_{j=1}^{n} N_j(x)\beta_j$
- Rewrite objective:

$$(y - N\beta)^T (y - N\beta) + \lambda \beta^T \Omega_N \beta$$

$[N]_{ij} = N_j(x_i)$

$[\Omega_N]_{jk} = \int N_j''(t) N_k''(t) dt$

- Solution: $\hat{\beta} = (N^T N + \lambda \Omega_N)^{-1} N^T y$ as in ridge
- Linear smoother:

$$\hat{f} = \underbrace{N(N^T N + \lambda \Omega_N)^{-1} N^T}_{L_\lambda} y$$

$\nu_\lambda = tr(L_\lambda)$ "smoothing matrix"

Splines Intro – Summary

- **Regression splines:**
  Fewer number of knots and no regularization
- **Smoothing splines:**
  Knots at every observation and regularization (smoothness penalty) to avoid interpolators

©Emily Fox 2014

Module 2: Splines and Kernel Methods

B-Splines

STAT/BIOSTAT 527, University of Washington
Emily Fox
April 10th, 2014

©Emily Fox 2014

Cubic Spline Basis and Fit

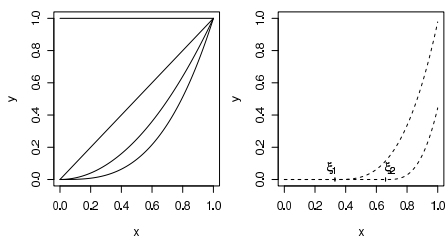
using truncated power basis

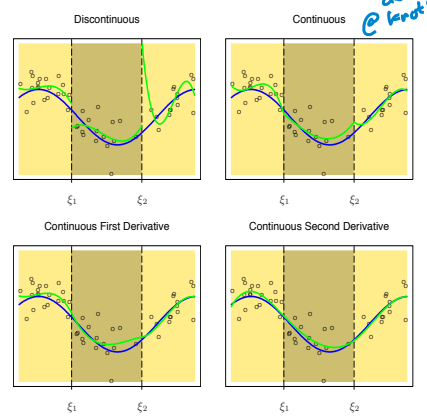
- Cubic spline function with *K* knots:

$$f(x) = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3 + \sum_{k=1}^{K} b_k (x - \xi_k)_+^3$$

M = 4
M-1 deg poly
M-2 cont. der @ knots

basis on (0,1)





B-Splines

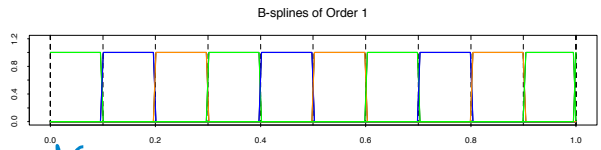
- Alternative basis for representing polynomial splines
- Computationally attractive…Non-zero over limited range
- As before:
  - Knots $\xi_1 < \dots < \xi_K$
  - Domain (a, b)
  - Number of basis functions = $M + K$ ← deg. of poly. + 1
- Step 1: Add knots $\xi_0 = a$ $\xi_{K+1} = b$
- Step 2: Define auxiliary knots τ_j needed to construct basis

choice is arb.

$$\tau_1 \leq \tau_2 \leq \cdots \leq \tau_M \leq \xi_0$$
$$\tau_{j+M} = \xi_j$$
$$\xi_{K+1} \leq \tau_{K+M+1} \leq \cdots \leq \tau_{K+2M}$$

B-Splines

- For 1st order B-spline



From Hastie, Tibshirani, Friedman book

1st order

$$B_j^1(x) = \begin{cases} 1 & \tau_j \le x \le \tau_{j+1} \\ 0 & \text{ow} \end{cases}$$

Haar basis function

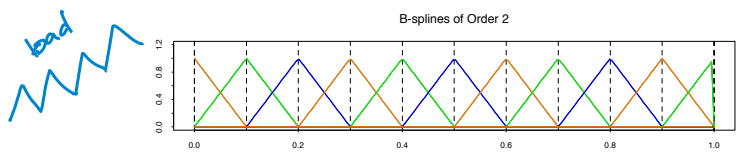
jth basis element

can form any piecewise constant fcn

B-Splines

- For 2nd order B-spline → piecewise linear fcn + cont. @ knots

bad



From Hastie, Tibshirani, Friedman book

- Modify 1st order basis:

neg. slope

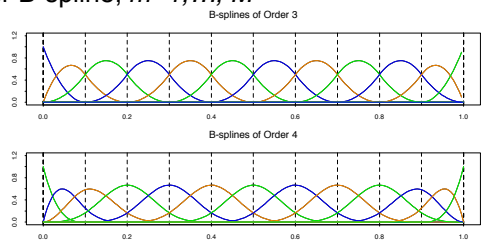
$$B_j^2(x) = \frac{x-\tau_j}{\tau_{j+1}-\tau_j} B_j^1(x) + \frac{\tau_{j+2}-x}{\tau_{j+2}-\tau_{j+1}} B_{j+1}^1(x)$$

pos. slope

- Convention: If divide by 0, set basis element to 0 IF $\tau_j = \tau_{j+1}$

B-Splines

- For m[th] order B-spline, *m=1,…, M*



From Hastie, Tibshirani, Friedman book

- Modify (m-1)[th] order basis:

$$B_j^m(x) = \frac{x - \tau_j}{\tau_{j+m-1} - \tau_j} B_j^{m-1} + \frac{\tau_{j+m} - x}{\tau_{j+m} - \tau_{j+1}} B_{j+1}^{m-1}$$

  - B-spline bases are ~~non-zero over domain spanned by at most M+1~~ knots
  - ~~Only~~ subsets $\{B_i^m \mid i = M - m + 1, \ldots, M + K\}$ are needed for basis of order *m* with knots ξ

for m=M → M+K basis fcns

Cubic Splines as Linear Smoothers

- Cubic spline function with *K* knots:

truncated power basis

$$f(x) = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3 + \sum_{k=1}^{K} b_k (x - \xi_k)_+^3$$

- Simply a linear model

$f(x) = E[Y|c] = c\gamma$

$$\underset{n\times(M+K)}{C} = \begin{bmatrix} 1 & x_1 & x_1^2 & x_1^3 & (x_1-\xi_1)_+^3 & \cdots & (x_1-\xi_K)_+^3 \\ \vdots & & & & & & \\ 1 & x_n & x_n^2 & x_n^3 & (x_n-\xi_1)_+^3 & \cdots & (x_n-\xi_K)_+^3 \end{bmatrix} \quad \gamma = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \\ b_1 \\ \vdots \\ b_K \end{bmatrix}$$

- Estimator: $\hat{\gamma} = (C^T C)^{-1} C^T Y$

- Linear smoother: $\hat{f} = \underbrace{C(C^T C)^{-1} C^T}_{L} Y$

Cubic B-Splines as Linear Smoother

- Cubic B-spline with *K* knots has basis expansion:

$$f(x) = \sum_{j=1}^{K+4} B_j^4(x)\beta_j$$

- Simply a linear model

$$B = \begin{bmatrix} B_1^4(x_1) & \cdots & B_{K+4}^4(x_1) \\ \vdots & & \\ B_1^4(x_n) & \cdots & B_{K+4}^4(x_n) \end{bmatrix} \qquad \gamma = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_{K+4} \end{bmatrix}$$

$$\hat{\gamma} = (B^T B)^{-1} B^T y$$

- Computational gain:

$n \times (K+4)$ matrix B with many 0's

$\rightarrow$ fewer multiplies (sparse inv.)

Return to Smoothing Splines

- Objective:

$$\min_f \sum_{i=1}^n (y_i - f(x_i))^2 + \lambda \int f''(x)^2 dx$$

- Solution:
  - ***Natural cubic spline***
  - Place knots at every observation location x_i
- Proof: See Green and Silverman (1994, Chapter 2) or Wakefield textbook
- Notes:
  - Would seem to overfit, but penalty term shrinks spline coefficients toward linear fit
  - Will not typically interpolate data, and smoothness is determined by λ

Smoothing Splines

of obs.

- Model is of the form: $f(x) = \sum_{j=1}^{n} N_j(x)\beta_j$

natural cubic spline basis

- Rewrite objective:

$$(y - N\beta)^T(y - N\beta) + \lambda\beta^T\Omega_N\beta$$

$n \times n$ matrix

$[N]_{ij} = N_j(x_i)$

$[\Omega_N]_{jk} = \int N_j''(t) N_k''(t)\,dt$

- Solution: $\hat{\beta} = (N^T N + \lambda\Omega_N)^{-1} N^T y$

as in ridge

- Linear smoother:

$$\hat{f} = \underbrace{N(N^T N + \lambda\Omega_N)^{-1} N^T}_{L_\lambda} y$$

$V_\lambda = tr(L_\lambda)$

"smoothing matrix"

Smoothing Splines

Previously,

- Model is of the form: $f(x) = \sum_{j=1}^{n} N_j(x)\beta_j$

Now,

- Using B-spline basis instead: $f(x) = \sum_{j=1}^{n+4} B_j^4(x)\beta_j$

$K = n$ knots

order $M = 4$ spline (cubic)

- Solution: $\hat{\beta} = (B^T B + \lambda\Omega_B)^{-1} B^T y$

$(n \times 4) \times (n \times 4)$

$n \times (n+4)$

lower 4 banded → computational eff.

- Penalty implicitly leads to natural splines
  - Objective gives infinite weight to non-zero derivatives beyond boundary

Spline Overview (so far)

Smoothing Splines

- Knots at data points x_i
- Natural cubic spline
- O(n) parameters
  - Shrunk towards subspace of smoother functions

Regression Splines

- $K < n$ knots chosen
- M^{th} order spline = piecewise M-1 degree polynomial with M-2 continuous derivatives at knots

- Linear smoothers, for example using natural cubic spline basis:

©Emily Fox 2014

Reading

- Hastie, Tibshirani, Friedman: 5.1-5.5 (skipping 5.3), Ch. 5 appendix
- Wakefield: 11.1.1-11.2.6

©Emily Fox 2014

What you should know…

- Regression splines
  - Cubic splines, natural cubic splines, …
  - Interpretation as a linear smoother
  - Degrees of freedom
- Smoothing splines
  - Arising from penalized regression setting with smoothness penalty
  - Cubic spline basis with knots at every data point
- Natural splines
  - Linear beyond boundary points
- B-splines
  - Basis functions with compact support
- Penalized regression splines
  - Choose knots as in regression splines, but penalize associated coefficients

©Emily Fox 2014