STAT462/862 Unit 3c
Piecewise Polynomial and Spline

Devon Lin

Department of Mathematics and Statistics, Queen's University

Outline

1. Piecewise Polynomials
2. Regression Splines
3. Smoothing Splines

Reference: Sections 5.1, 5.2 and 5.4 of HTF

Introduction

Linear models

- Linear in X (input vector)
- Augment/replace X with additional variables. Consider the model of the form

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

where $h_m(X) : \mathbb{R}^p \to \mathbb{R}$ and $h_m(X)$ can be: X_j^2, $X_i X_j$, $\log(X_j)$, or $I(l_m \leq X_k \leq u_m)$.

- Fit by least squares,

$$\min_{\boldsymbol{\beta}} \sum_{i=1}^{N} \left(y_i - \sum_{m=1}^{M} \beta_m h_m(\boldsymbol{x}_i)\right)^2.$$

Polynomial model

- A second-order polynomial model

$$f(X) = \beta_0 + \sum_{j=1}^{p} \beta_j X_j + \sum_{i,j} \beta_{ij} X_i X_j$$

- A cubic model

$$f(X) = \beta_0 + \sum_{j=1}^{p} \beta_j X_j + \sum_{i,j} \beta_{ij} X_i X_j + \sum_{i,j,k} \beta_{ijk} X_i X_j X_k$$

Introduction

Build a dictionary $\mathcal{D}$ consisting typically a very large number $|\mathcal{D}|$ of basis functions.

- Restriction methods: e.g., additivity

$$f(X) = \sum_{j=1}^{p} f_j(X_j) = \sum_{j=1}^{p} \sum_{m=1}^{M_j} \beta_j h_{jm}(X_j).$$

- Selection methods: methods like CART (Classification and regression trees), MARS (Multivariate adaptive regression splines), etc.
- Regularization methods: e.g., ridge regression.

Regularization

- Sometimes we use a large expansion or an over-represented basis:

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

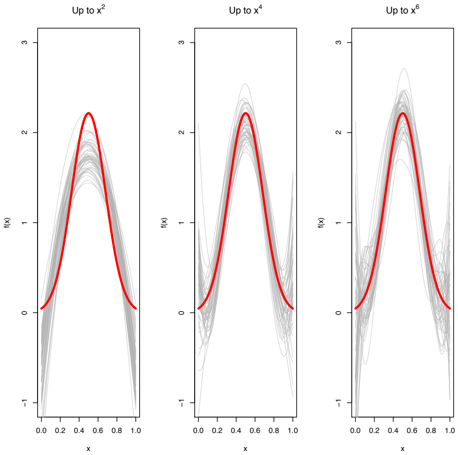
- Control the model complexity by regularization:

$$\min_{\boldsymbol{\beta}} \sum_{i=1}^{N} \left(y_i - \sum_{m=1}^{M} \beta_m h_m(\boldsymbol{x}_i)\right)^2 + \lambda J(f),$$

  where $J(f)$ is a *roughness penalty* or other *regularization*, and λ is tuning parameter.

  - For example: L_2 regularization $J(f) = \|\boldsymbol{\beta}\|^2$, or $J(f) = \int [f''(t)]^2 dt$.

Problems with polynomial regression



Gray lines are models fit to 100 observations arising from the true f, colored red

Global v.s. local basis functions

- Polynomial regression uses global basis functions.
- Now consider *local* basis functions, thereby ensuring that a given observation affects only the nearby fit, not the fit of the entire line.
- We focus on piecewise polynomials.

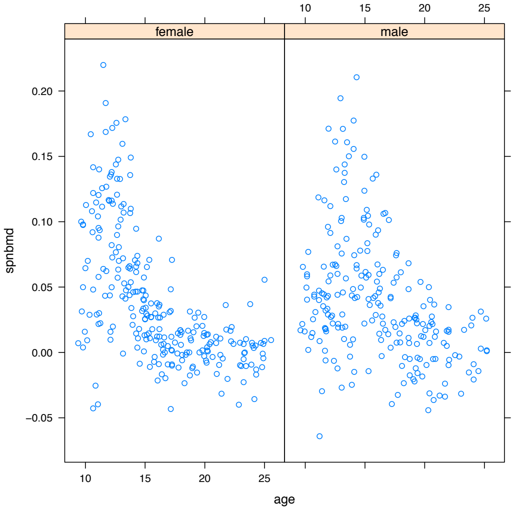
Piecewise Polynomials and Splines

- Piecewise polynomial: $f(X)$ is obtained by dividing the domain of X into contiguous intervals (when X is one-dimensional), and representing f by a separate polynomial in each interval.

- *Splines* are piecewise polynomials joined together to make a single smooth curve.

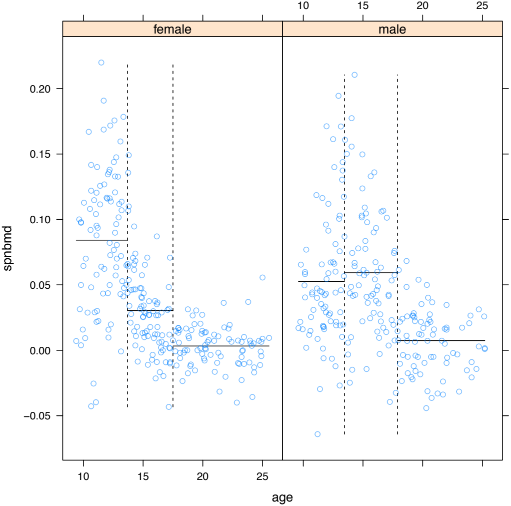
- Consider a study of changes in bone mineral density in adolescents, taken on two consecutive visits
- The outcome of the study is the difference in spinal bone mineral density divided by the average of the two measurements
- Age is a predictor, which is the average age over the two visits

Data

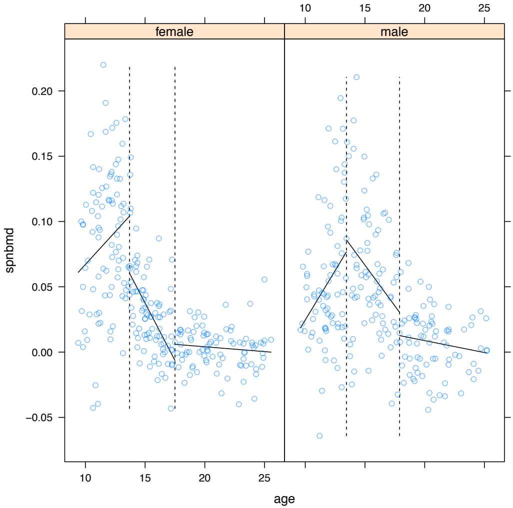


female
male
spnbmd
age
0.20
0.15
0.10
0.05
0.00
−0.05
10
15
20
25
10
15
20
25

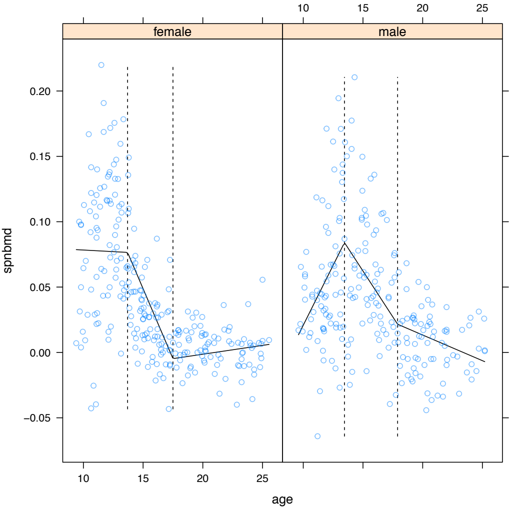
Piecewise Constant Model



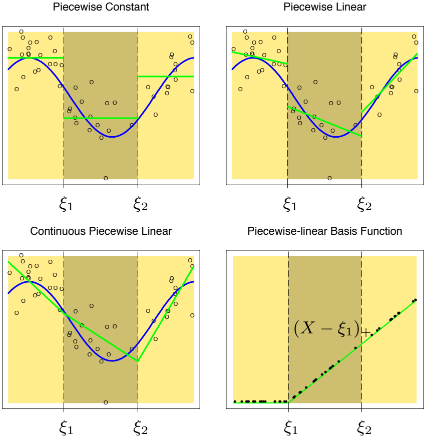
Piecewise linear Model



Piecewise linear Model



Piecewise Polynomial and Order 2 Splines



Basis for the model in the bottom right panel: $h_1(X) = 1$, $h_2(X) = X$, $h_3(X) = (X - \xi_1)_+$, $h_4(X) = (X - \xi_2)_+$.

Cubic Splines, order-4

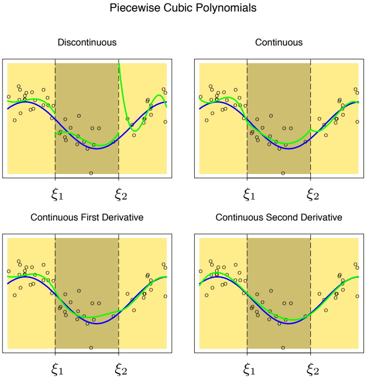
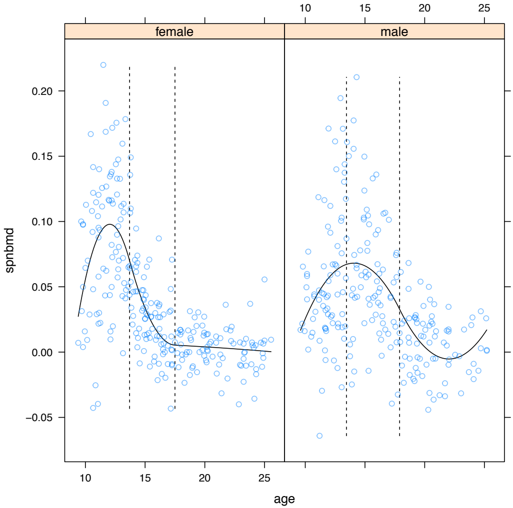


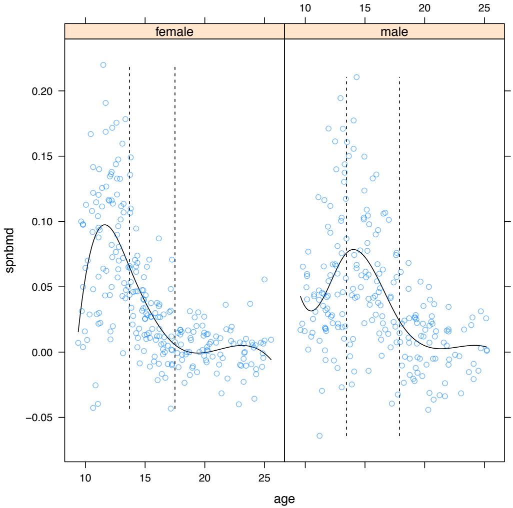
FIGURE 5.2. *A series of piecewise-cubic polynomials, with increasing orders of continuity.*

$$
\begin{aligned}
h_1(X) &= 1, & h_3(X) &= X^2, & h_5(X) &= (X - \xi_1)^3_+ \\
h_2(X) &= X, & h_4(X) &= X^3, & h_6(X) &= (X - \xi_2)^3_+ .
\end{aligned}
$$

Quadratic Spline



Cubic Spline



General order M splines

- Order-M spline with knots ξ_j, $j = 1, \ldots, K$ is a piecewise-polynomial of order M (different from order of polynomials), and has continuous derivatives up to order $M - 2$.

$$h_j(X) = X^{j-1}, \quad j = 1, \ldots, M,$$
$$h_{M+l}(X) = (X - \xi_l)_+^{M-1}, \quad l = 1, \ldots, K$$

  - Second-order spline (M=2)

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

- By requiring continuous derivatives, we ensure that the resulting function is as smooth as possible.

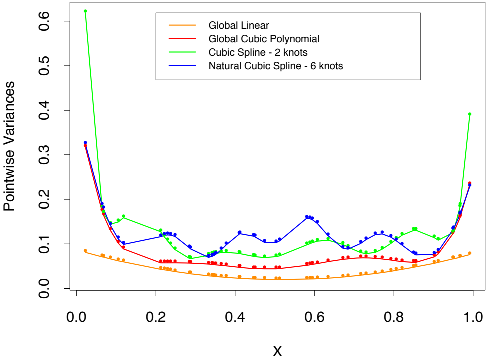
General order M splines

- Within each interval $[a, \xi_1]$, $[\xi_i, \xi_{i+1}]$, $i = 1, \ldots, K-1$, $[\xi_K, b]$ the piecewise function $s(X) = s_i(X)$, a polynomial of order $M-1$.
  - The number of unknown coefficients: $M \times (K+1)$;
  - Continuity to the order $M-2$ derivative: $(M-1) \times K$;
  - The degree of freedom (the number of coefficients that can be chosen freely): $M \times (K+1) - (M-1) \times K = M + K$.

General order M splines

- We can obtain more flexible curves by increasing the degree of the spline and/or by adding knots.
- However, there is a tradeoff:
  - Few knots/low degree: Resulting class of functions may be too restrictive (bias)
  - Many knots/high degree: running the risk of overfitting (variance)

- **Regression splines**: fixed-knot splines, i.e., specify M, the number of knots and their placement.
- **Cubic spline** $(M = 4)$ is the lowest-order spline where knot discontinuity is invisible.



Pointwise variance curves for four different models, with X consisting of 50 points randomly drawn from $U(0,1)$.

$$y_i = f(x_i) + \epsilon_i, \epsilon_i \sim iid\ N(0, \sigma^2)$$

$$Var(\hat{f}(x)) = \sigma^2 \boldsymbol{h}(x)^T (\boldsymbol{H}^T \boldsymbol{H})^{-1}) \boldsymbol{h}(x)$$

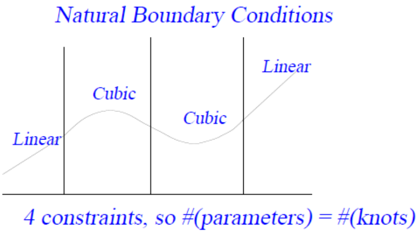
Natural Cubic Splines

Natural cubic spline:

- Linear near the boundary intervals: $[a, \xi_1]$ and $[\xi_K, b]$.
- Cubic polynomial in $[\xi_1, \xi_K]$.
- Continuous up to the 2nd derivative in $(-\infty, \infty)$.

Natural Cubic Splines

- Add additional constraint: linear functions beyond the boundary knots.



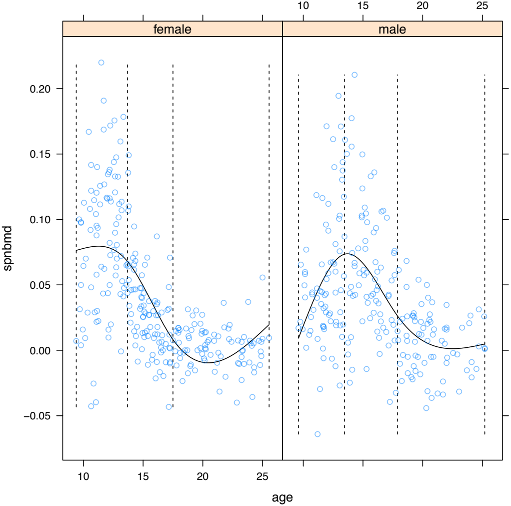
- Free up 4 degrees of freedom (two constraints each in both boundaries).
- Then $df = M + K - 4$, where $M = 4$.

Natural Splines in R

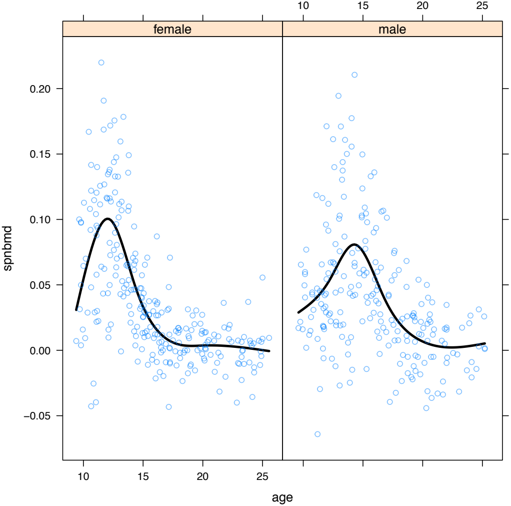
- Use the R package *splines* and the function ns.

```
X = ns(x,knots=quantile(x,p=c(1/3,2/3)))
X = ns(x,df=5)
X2 = predict(X,newx=x2)
```

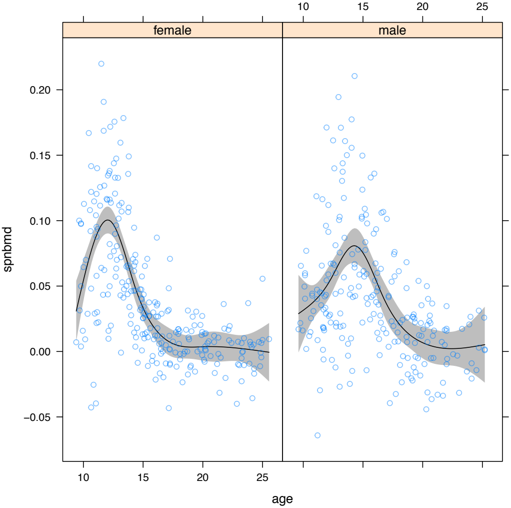
Natural Cubic Spline



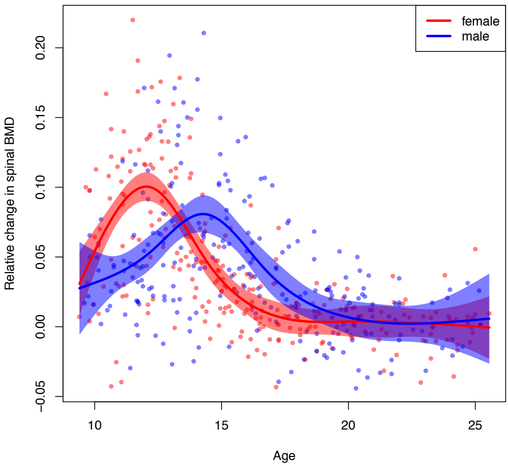
Natural Cubic Spline with $df = 5$



Confidence Band in Natural Cubic Spline with $df = 5$



Confidence Band in Natural Cubic Spline with $df = 5$



Spline in SAS

- ods graphics on;
  proc transreg data = bone plots;
  model identity(Spnbmd) = class(Gender)
    spline(Age/ nknots=2 evenly degree = 2) ;
  output out = spnbmdout predicted cli clm coefficients;
  run;
  ods graphics off;
- use $NATURALCUBIC$ option in $EFFECT$ statement for natural cubic splines.

Smoothing Splines

- Assume using the maximal number of knots;
- Control the complexity of the fit by regularization:

$$RSS(f, \lambda) = \sum_{i=1}^{N} \{y_i - f(x_i)\}^2 + \lambda \int \{f''(t)\}^2 dt.$$

  A trade off between
  - $\lambda = 0$: f can be any functions that interpolates the data;
  - $\lambda = \infty$: the simple least square line fit, since no second derivative can be tolerated.
- A spline that describes and smooths noisy data by passing close to the data (x_i, y_i) without the requirement of passing through them is called a smoothing spline.
- In general, RSS has an explicit minimizer which is a natural cubic spline with knots values of x_i, $i = 1, \dots, N$.

Natural cubic splines for the regularization problem

Theorem: Out of all twice-differentiable functions, the one that minimizes

$$\sum_{i=1}^{N}\{y_i - f(x_i)\}^2 + \lambda \int \{f''(t)\}^2 dt,$$

is a natural cubic spline with knots at every unique value of x_i.

Smoothing Splines

- Consider one-dimensional case:

$$f(x) = \sum_{j=1}^{N} G_j(x)\theta_j$$

- Write RSS into

$$RSS(\theta, \lambda) = (\mathbf{y} - \mathbf{G}\boldsymbol{\theta})^T(\mathbf{y} - \mathbf{G}\boldsymbol{\theta}) + \lambda\boldsymbol{\theta}^T\boldsymbol{\Omega}\boldsymbol{\theta},$$

  where $\{\mathbf{G}\}_{ij} = G_j(x_i)$ and $\{\boldsymbol{\Omega}\}_{jk} = \int G_j''(t)G_k''(t)dt$.
  Solution

$$\hat{\boldsymbol{\theta}} = (\mathbf{G}^T\mathbf{G} + \lambda\boldsymbol{\Omega})^{-1}\mathbf{G}^T\mathbf{y}.$$

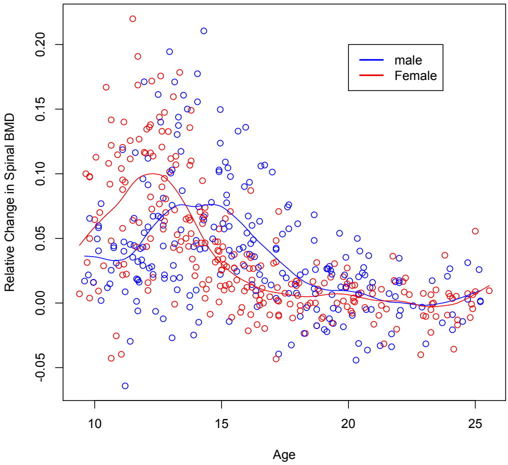
- Smoothing Spline:

$$\hat{f}(x) = \sum_{j=1}^{N} G_j(x)\hat{\theta}_j.$$

Smoothing spline in R: Bone Example

```
plot(spnbmd ~ age, data=bone, col =
ifelse(gender=="male", "blue", "red2"),
xlab="Age", ylab="Relative Change in Spinal BMD")
bone.spline.male <- with(subset(bone,gender=="male"),
smooth.spline(age, spnbmd,df=12))
bone.spline.female <- with(subset(bone, gender=="female"),
smooth.spline(age, spnbmd, df=12))
lines(bone.spline.male, col="blue")
lines(bone.spline.female, col="red2")
legend(20,0.20,legend=c("male", "Female"),
col=c("blue", "red2"),lwd=2)
```

Bone Example



The tuning parameter

- Smoothing splines: the knots are at all the unique training X's, and the cubic degree is always used.
- One way to choose the tuning parameter λ is to use cross-validation.

Choosing λ

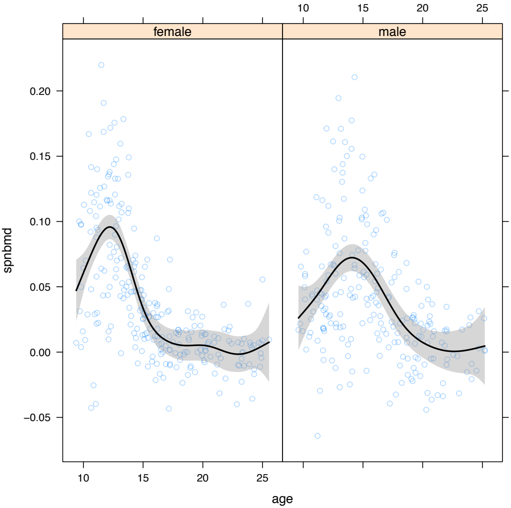
- Use K-fold $(C_1, \dots, C_K)$ cross-validation to select λ

$$CV(\hat{f}_\lambda) = \frac{1}{N} \sum_{k=1}^{K} \sum_{i \in C_k} \left(y_i - \hat{f}_\lambda^{(-C_k)}(x_i) \right)^2 .$$

- Leave-one-out cross-validation (GCV)

$$CV(\hat{f}_\lambda) = \frac{1}{N} \sum_{i=1}^{N} \left(y_i - \hat{f}_\lambda^{(-i)}(x_i) \right)^2 .$$

Bone Example



Smoothing Spline in SAS

```
* smoothing spline;
PROC TRANSREG  ss2 DATA=bone;
 MODEL identity(Spnbmd)=class(Gender / zero=none) *
    smooth(Age / sm=50);
RUN;
```

Summary

- Piecewise polynomial
- Regression Splines
- Natural splines
- Smoothing splines