12 Nonparametric Regression

Regression setting

$$Y = \underbrace{f(\underset{\sim}{x})}_{\text{what we have learned so far}} + \varepsilon$$

linear regression

trees

polynomial regression

bagged trees

random forest or boosted trees

neural networks

deep models

nonparametric fits in here

model complexity

Setup: single quantitative predictor & response.

Goal: Get a flexible & "smooth" model relating X to Y.

Problem with polynomial regression $f(x) = \beta_0 + \beta_1 x + \beta_2 x^2 + \dots + \beta_p x^p$

- often need $p >> 0$ for reasonable fit
- Extrapolated values behave poorly
  "fit is good locally, but not globally"

This chapter:

$$f(x) = \beta_0 + \sum_{j=1}^{p} \beta_j K_j(x)$$

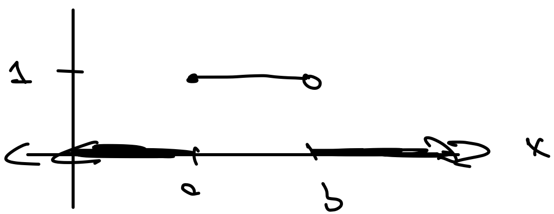
for some functions $K_1, \dots, K_p$

- regression splines
- penalized splines
- smoothing splines
- local regression
- generalized additive models

12.1 Regression Splines

Piecewise constant

$$K(x) = \mathbb{1}_{[a,b)}(x) = \begin{cases} 1, & x \in [a,b) \\ 0, & x \notin [a,b) \end{cases}$$



okay b/c can rescale feature space

Assume $x \in [0,1]$. Define knots $0 < a_1 < a_2 < \dots < a_p < 1$

and

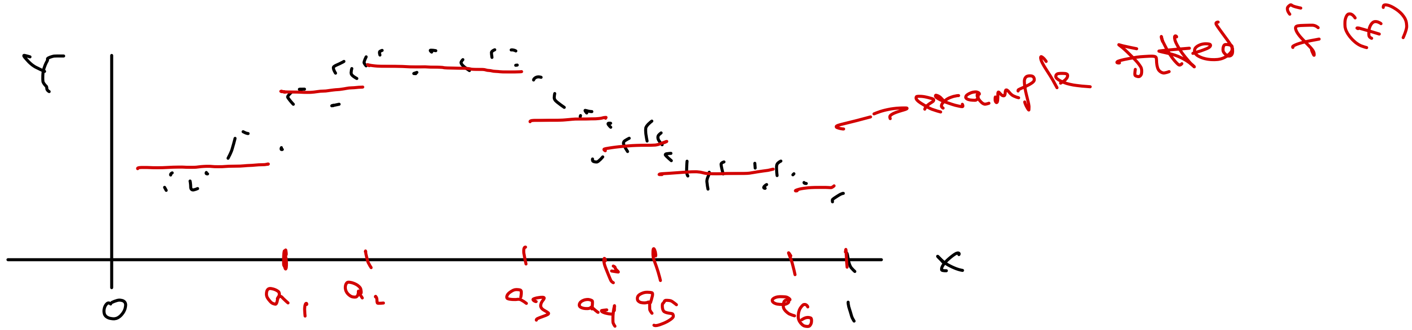
$$K_1(x) = \mathbb{1}_{[a_1, a_2)}(x)$$

$$K_2(x) = \mathbb{1}_{[a_2, a_3)}(x)$$

$$\vdots$$

$$K_p(x) = \mathbb{1}_{[a_p, 1]}(x)$$

$$Y = \beta_0 + \sum_{j=1}^{p} \beta_j K_j(x) + \varepsilon$$



$\boxed{\text{Note}}$ Given data & $K_1, \ldots, K_p$, $\underline{\hat{\beta}}$ can be estimated via OLS using design matrix

$$X = \begin{pmatrix} 1 & K_1(x_1) & K_2(x_1) & \cdots & K_p(x_1) \\ 1 & K_1(x_2) & \vdots & & \vdots \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & K_1(x_n) & K_2(x_n) & \cdots & K_p(x_n) \end{pmatrix}$$

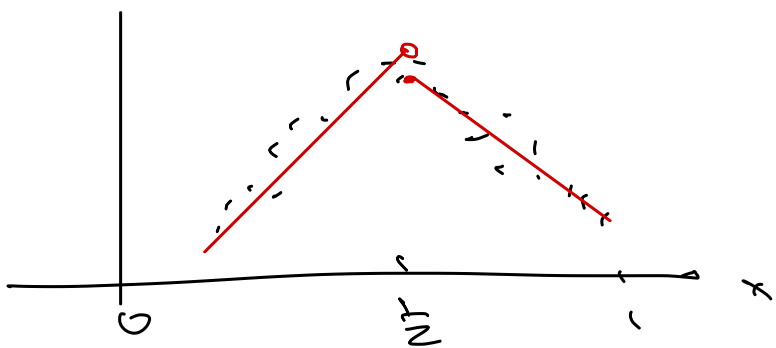
Remark A piecewise linear function is linear btwn knots.

Ex $p=1$ $\quad a_1 = 1/2$

$$K_1(x) = x\,\mathbb{1}_{[0,1/2)}(x)$$

$$K_2(x) = x\,\mathbb{1}_{[1/2,1]}(x)$$

$$\Rightarrow f(x) = \beta_0 + \beta_1 K_1 + \beta_2 K_2 = \begin{cases} \beta_0 + \beta_1 x & x \in [0, \tfrac{1}{2}) \\ \beta_0 + \beta_2 x & x \in [\tfrac{1}{2}, 1] \end{cases}$$



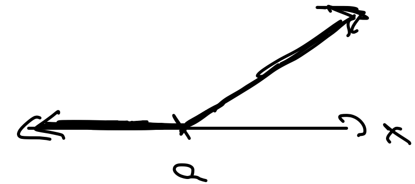
at $\frac{1}{2}$, ...

$$\beta_0 + \beta_1 \tfrac{1}{2} \overset{?}{=} \beta_0 + \beta_2 \tfrac{1}{2}$$

$$\beta_1 \overset{?}{=} \beta_2$$

DEF <u>Truncated linear spline</u> enforces continuity @ knots.

$$(x-a)_+ = \begin{cases} x-a & x \geq a \\ 0 & x < a \end{cases}$$



Model $K_i = (x-a_i)_+$ for knots $0 < a_1 < a_2 < \dots < a_p < 1$

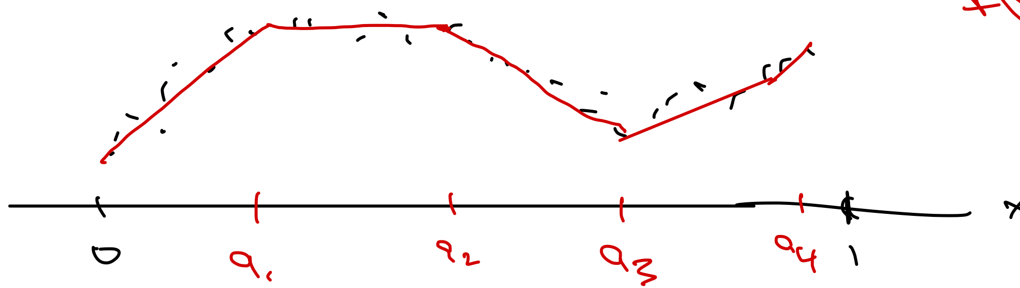
"piecewise linear spline"

$$f(x) = \beta_0 + \beta_1 x + \sum_{j=1}^{p} \beta_{1j} K_j(x)$$

$x \in (a_1, a_2)$

$$f(x) = \beta_0 + \beta_1 x + \beta_{11}(x-a_1)$$

$$= (\beta_0 - \beta_{11} a_1) + (\beta_1 + \beta_{11}) x$$



$\boxed{\text{DEF}}$ A <u>cubic spline</u> with knots $0 < a_1 < a_2 < \cdots < a_p < 1$

is a cubic polynomial on $(0, a_1), (a_1, a_2), \ldots, (a_p, 1)$

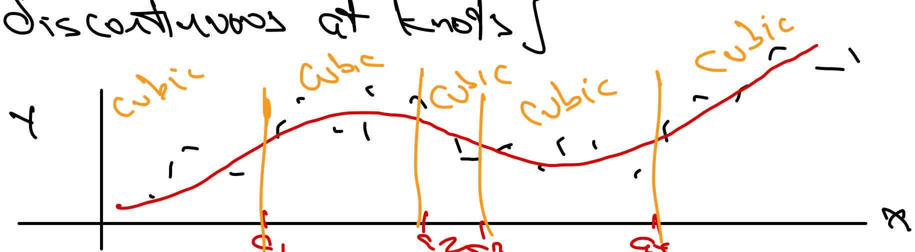
with f, f', f'' continuous at knots.

[cubic spline is <u>not</u> a cubic function]

Representation using truncated basis:

$$f(x) = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3 + \sum_{j=1}^{p} \beta_{3j} (x - a_i)_+^3$$

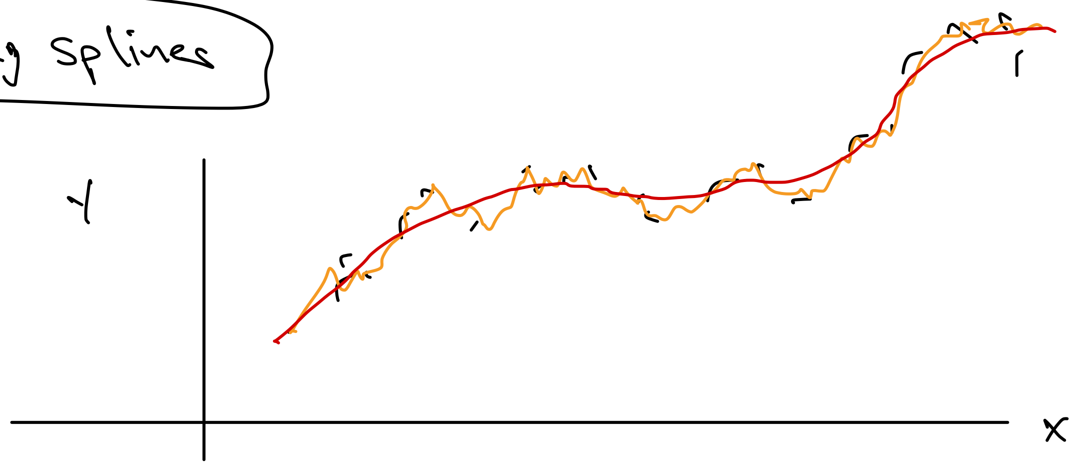
has continuous 0th, 1st, 2nd derivs [3rd deriv is discontinuous at knots]



In practice we use a B-spline representation that spans the same class of functions.

DEF A natural cubic spline with knots $0 < a_1 < a_2 < \dots < a_p < 1$ is a cubic spline that is linear on $(0, a_1)$ + $(a_p, 1)$

Smoothing splines



Cubic smoothing spline f, minimizes

$$\underbrace{\sum_{i=1}^{n}(y_i - f(x_i))^2}_{\text{get close to data}} + \lambda \underbrace{\int_0^1 (f''(x))^2\,dx}_{\text{penalizes wiggliness of } f}$$