Nonlinear effects and generalized additive models
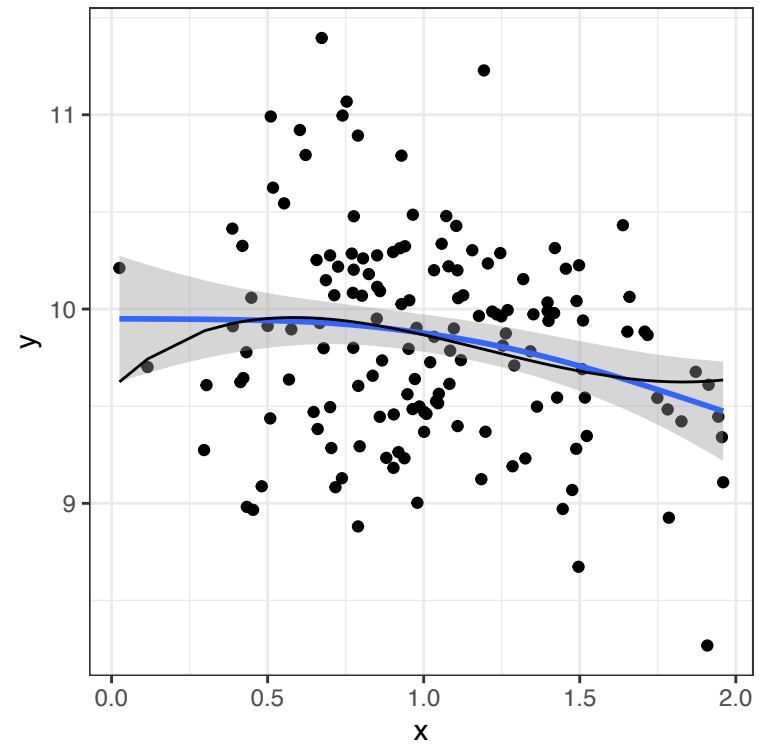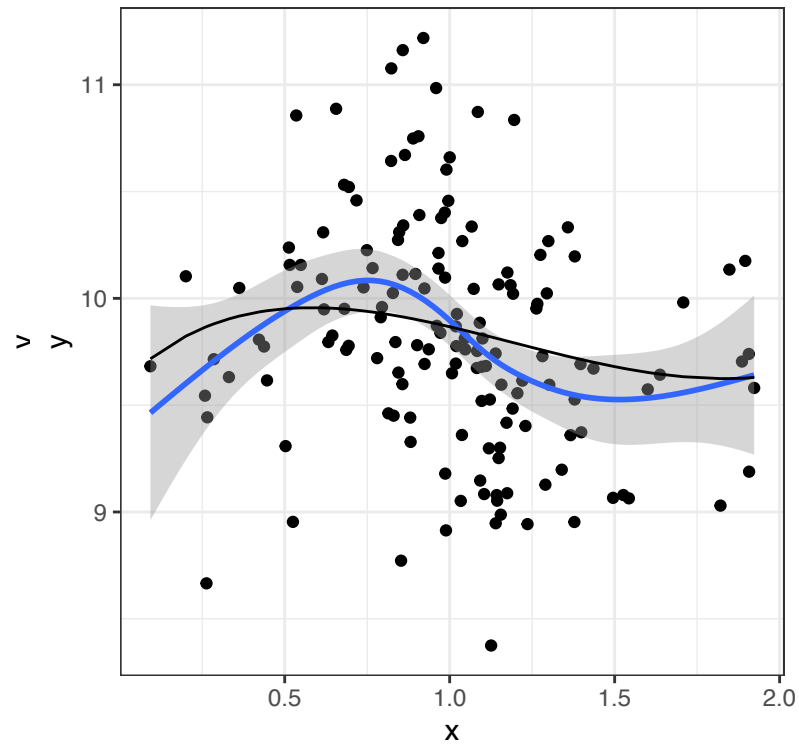
Roger Levy

30 April 2025

Inferring the trend

Generalized Additive Models (GAMs) Linear Models (=other regression) Linear models (=other regression)

Taking on nonlinearity: major ideas

- Use a **basis expansion:** a set of functions $\{h_i\}$ each of which differently transforms your original predictor X

- Learn the $Y \sim X$ as a linear comb. of the basis functions

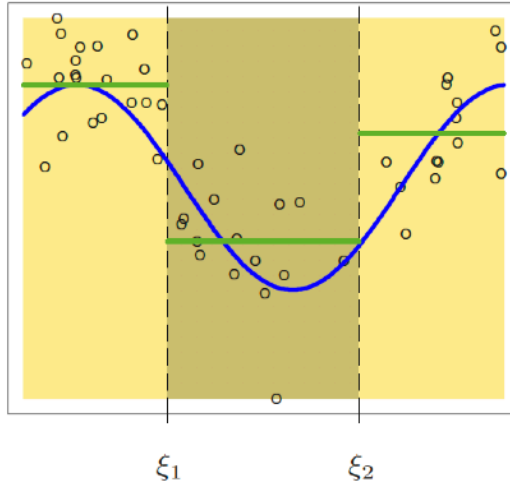$$f(x) = \sum_i \beta_i h_i(x) \qquad\qquad y \sim \mathrm{Exp}(f(x))$$

Some exponential-family function (one that you could use in a generalized linear model)
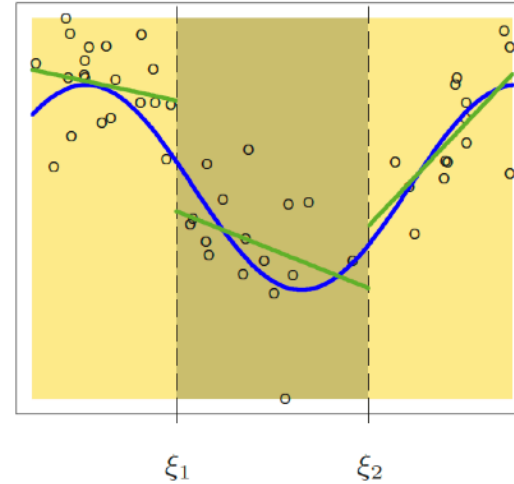
- Use **regularization** and **held-out accuracy evaluation** to guard against overfitting

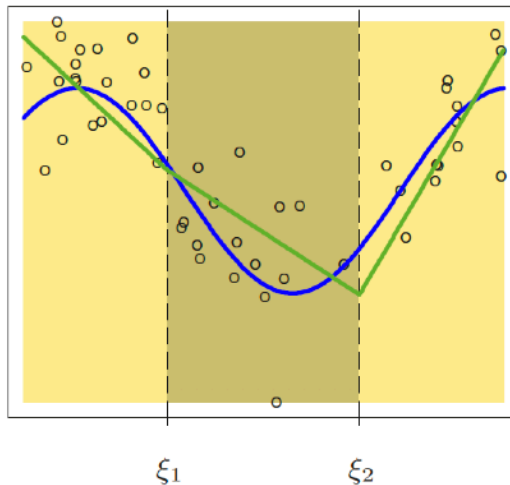# Example basis expansions for *piecewise* fits

Piecewise Constant

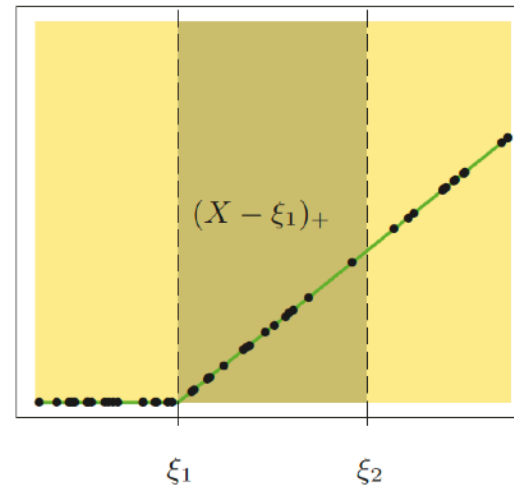$h_1 = I(X < \xi_1)$

$h_2 = I(\xi_1 < X \leq \xi_2)$

$h_3 = I(\xi_2 < X \leq \xi_3)$

Piecewise Linear

$h_1 = I(X < \xi_1)$

$h_2 = I(\xi_1 < X \leq \xi_2)$

$h_3 = I(\xi_2 < X \leq \xi_3)$

$h_4 = h_1(X)X$

$h_5 = h_2(X)X$

$h_6 = h_3(X)X$

Continuous Piecewise Linear

$h_1 = 1$

$h_2 = X$

$h_3 = (X - \xi_1)_+$

$h_4 = (X - \xi_2)_+$

Piecewise-linear Basis Function

$(X - \xi_1)_+$

(From Hastie et al., 2009)

4

Example basis expansions for *piecewise* fits

Piecewise Cubic Polynomials
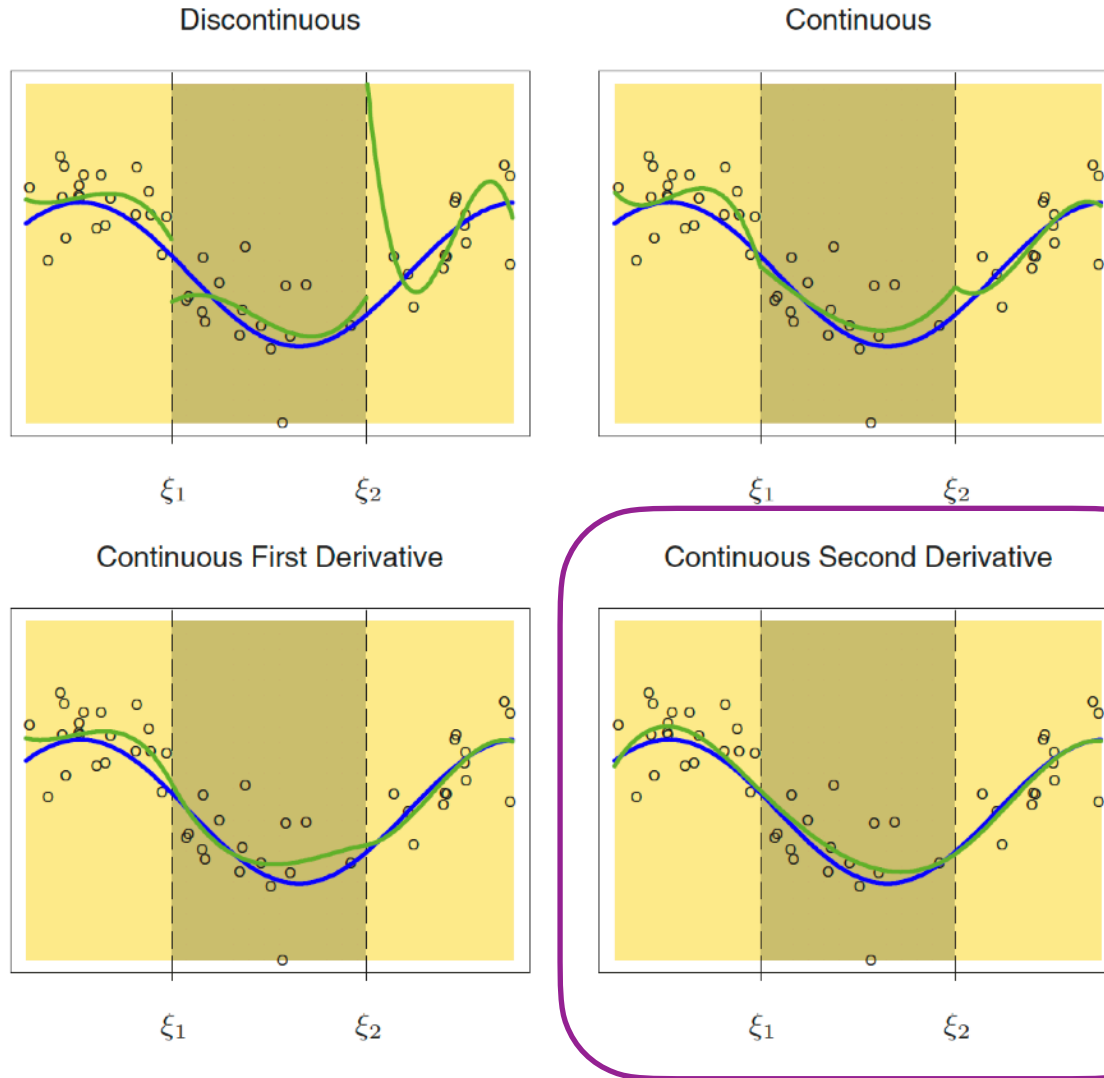
Popular modeling choice

5

Natural, or restricted, cubic splines

- Intuitively: a piecewise cubic function whose 0th, 1st, and second derivatives are continuous at all the **knots**
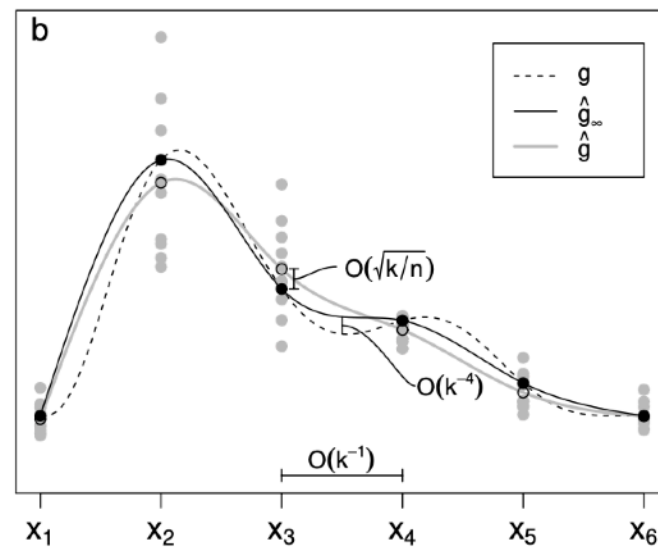
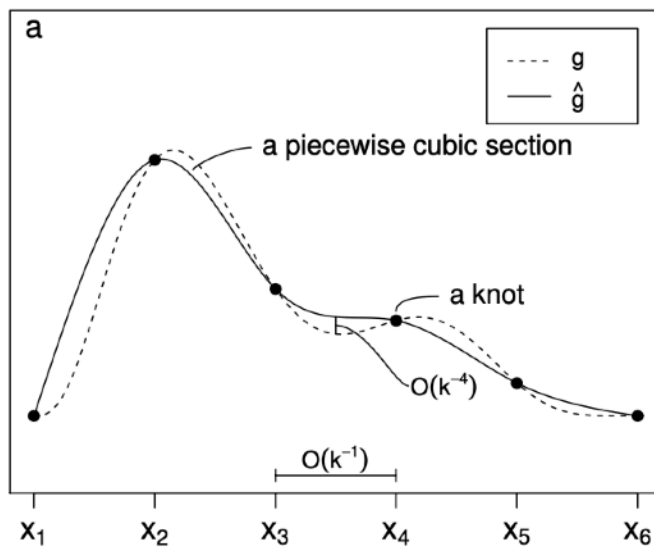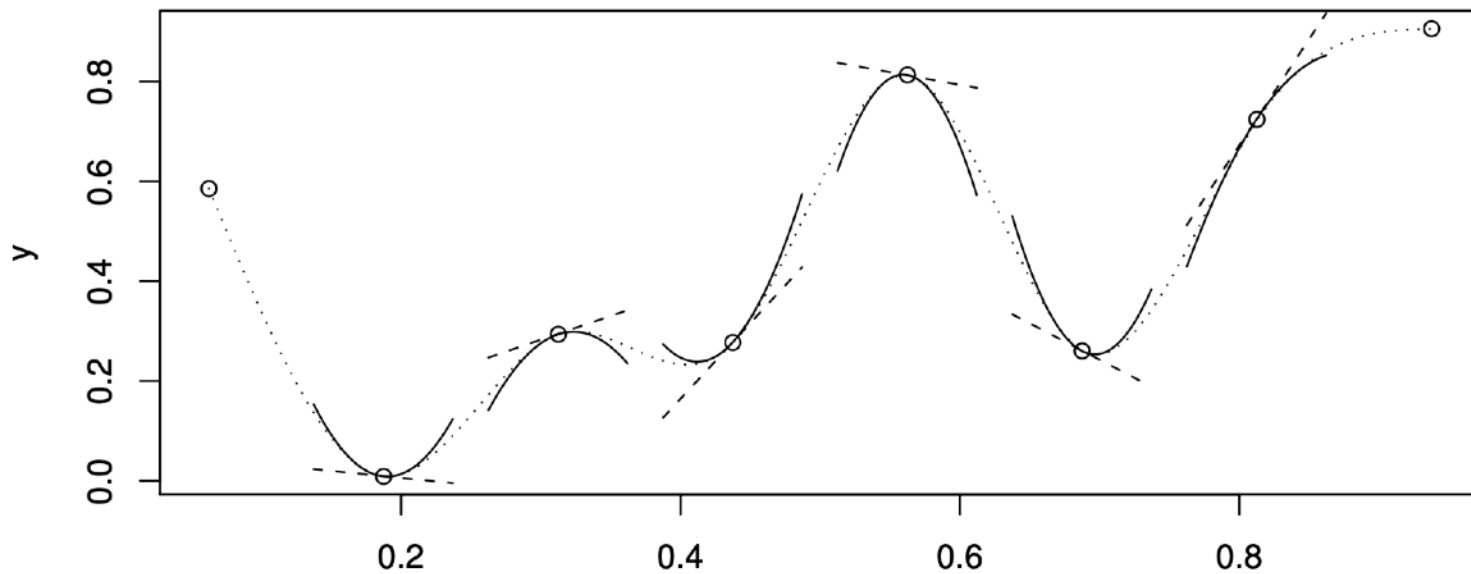- Technically, for n knots:

$$h_1(X) = 1$$
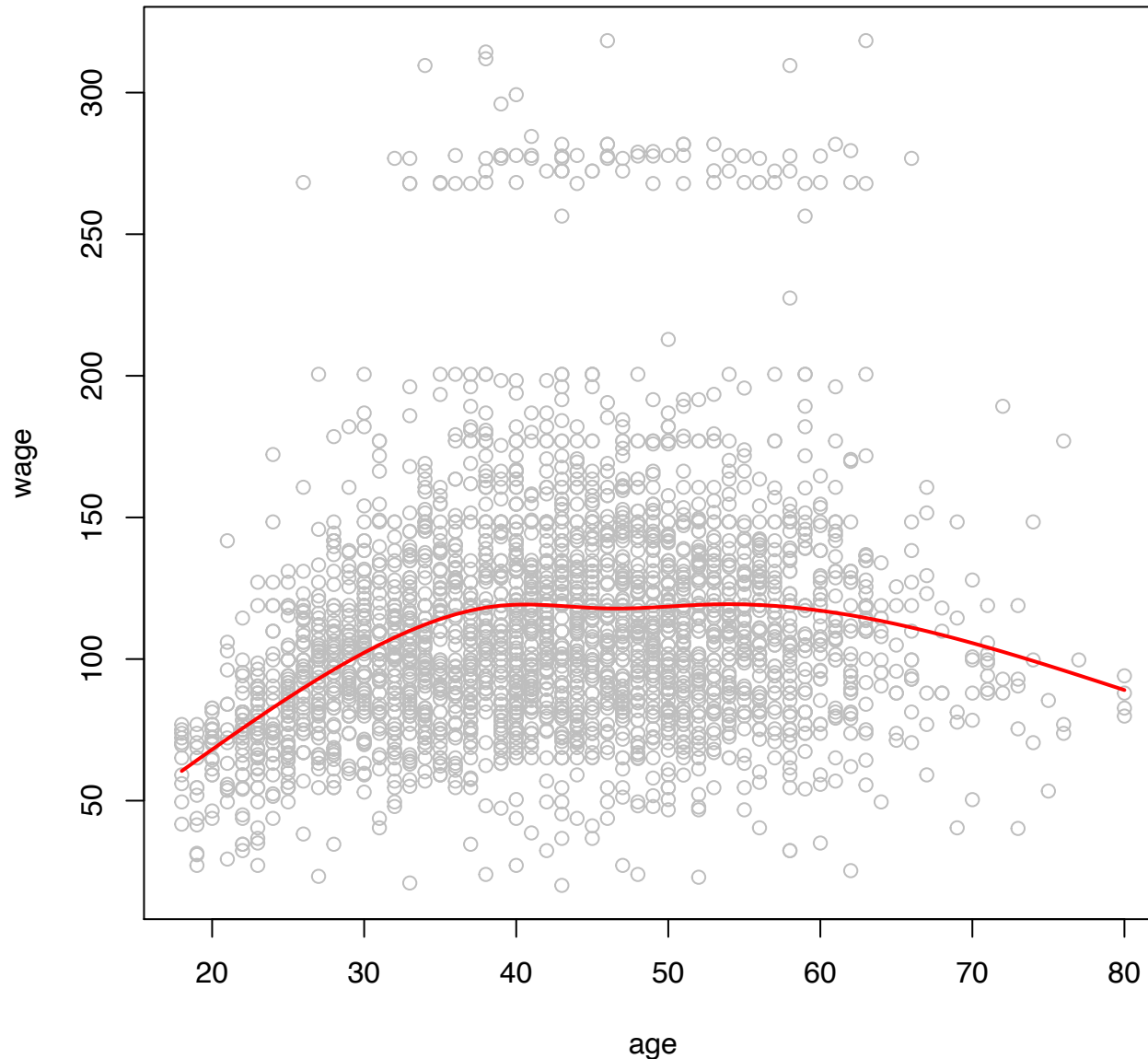$$h_2(X) = X$$
$$h_3(X) = X^2$$
$$h_4(X) = X^3$$
$$h_i(X) = (x - \xi_i)_+^3, \text{ for } i = 5,...,n + 4$$

- Powerful framework, but leaves open:
  - How many knots to use
  - Where to put the knots

Natural cubic splines, visualized

Example of fitted natural splines (5 d.f.)

From smooths to GAMs

- A **generalized additive model** (GAM) involving 2 predictors X and V for response Y looks like this:

$$y_i = \mathbf{A}_i \theta + f_1(x_i) + f_2(v_i) + f_3(x_i, v_i) + \epsilon_i$$

  where each $f_i()$ is a **smooth function** of its arguments

- This is a highly general and expressive formalism, and cannot be fitted via maximum likelihood (why?)

- So in practice, GAMs are fitted by:
  - choosing a **penalization** or **regularization** component for the objective function;
  - choosing a set of **basis functions** by which to represent each smooth function;
  - Searching for a fit that optimizes the objective function

Penalization & cross-validation

- Penalized objective:

$$\|\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\|^2 + \underbrace{\lambda_1 \boldsymbol{\beta}^\mathsf{T} \mathbf{S}_1 \boldsymbol{\beta}}_{} + \underbrace{\lambda_2 \boldsymbol{\beta}^\mathsf{T} \mathbf{S}_2 \boldsymbol{\beta}}_{}$$

Terms penalizing wiggliness for f_1 and f_2

How well data are predicted

- How to choose $\{\lambda_i\}$?

- Here: **cross-validation**

Strategy for getting proficient in GAMs

- Read Jones et al. 2013 Chapter 7, and do the exercises

- Read Wood 2017, starting with Chapter 4 and Chapter 7, and do the exercises

- To build expertise read Wood 2017 Chapter 6 and also Chapter 5

- Both these books are available online via MIT Libraries

- For the rest of this class, we will build a bit of expertise via worked examples

  - Data from Boyce & Levy, 2023

  - Material from the "GAMs in R" interactive course (https://noamross.github.io/gams-in-r-course/)