Chp. 7: Beyond Linearity

Outline (non-linearity in a single predictor)

- Polynomial regression
- Step functions
- Regression splines
- Smoothing splines

Basis functions

A family of functions or transformations that can be applied to a variable X : $b_1(X), b_2(X), ..., b_K(X)$.

Once the basis functions have been determined, the models are linear in these new variables and the fitting proceeds as before:

$$y_i = \beta_0 + \beta_1 b_1(x_i) + \beta_2 b_2(x_i) + ... + \beta_K b_K(x_i) + \epsilon_i$$

Polynomial Regression

One way to provide a non-linear fit to data is just to include extra predictors, obtained by raising each of the original predictors to a power. The coefficients can be estimated using least squares.

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 {x_i}^2 + \beta_3 {x_i}^3 + ... + \beta_d {x_i}^d + \epsilon_i$$

Here, the basis function is $b_j(x_i) = x_i^j$

Typically, $d \leq 4$, since the polynomial curve can take some very strange shapes if d is large.

Polynomial Regression

Degree-4 polynomial of wage as a function of age, with wage as a quantitative (left) or binary (right) response:

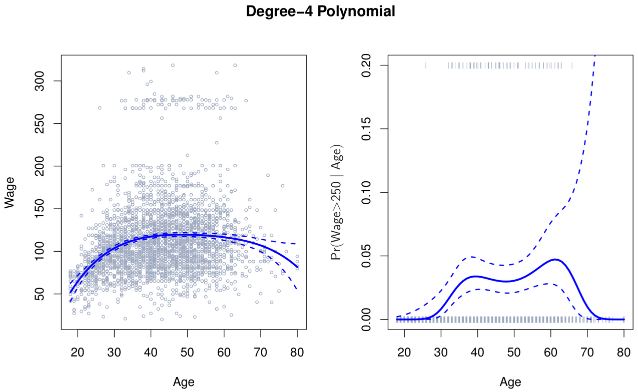

Figure 1: Fig 7.1

Step Functions

Polynomial regression imposes *global* structure on the non-linear function of X.

Alternatively, we can break the range of X into bins and fit a different constant in each bin.

$$
\begin{aligned}
C_0(X) &= I(X < c_1), \\
C_1(X) &= I(c_1 \leq X < c_2) \\
C_2(X) &= I(c_2 \leq X < c_3) \\
&\vdots \\
C_K(X) &= I(c_K \leq X)
\end{aligned}
$$

$I(\cdot)$ is an indicator function that returns 1 if the condition is true and 0 otherwise. For example, $I(c_K \leq X)$ returns 1 if $c_K \leq X$.

Step Functions

We can then use least squares to fit a linear model using $C_1(X), C_2(X), ..., C_K(X)$ as predictors:

$$y_i = \beta_0 + \beta_1 C_1(x_i) + \beta_2 C_2(x_i) + ... + \beta_K C_K(x_i) + \epsilon_i$$

Step Functions

Piecewise-constant functions can miss the action:

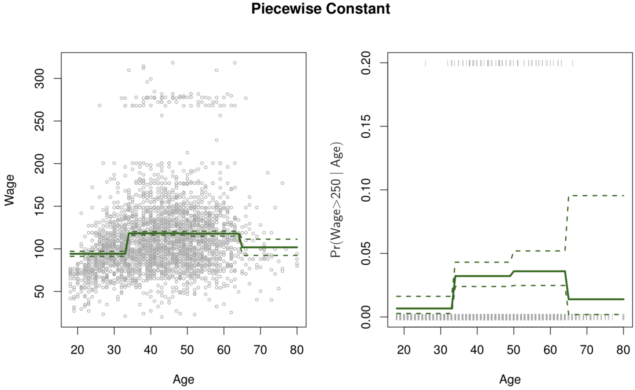

Figure 2: Fig 7.2

Regression Splines

Combining these approaches, we have *piecewise polynomial regression*, in which we fit separate low-degree polynomials over different regions of X.

$$y_i = \begin{cases} \beta_{01} + \beta_{11}x_i + \beta_{21}x_i^2 + \beta_{31}x_i^3 + \epsilon_i & \text{if } x_i < c; \\ \beta_{02} + \beta_{12}x_i + \beta_{22}x_i^2 + \beta_{32}x_i^3 + \epsilon_i & \text{if } x_i \geq c. \end{cases}$$

The point where the coefficient changes is a *knot*. We can use more knots and different degree polynomials.

Regression Splines

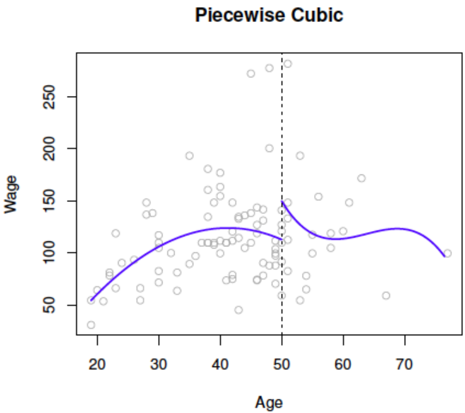

Figure 3: Fig. 7.3a

Regression Splines

Joining fitted curves by using constraints

Figure 4: butterfly splines in a coffee table

Regression Splines

We can join the curves by imposing the constraint that the fitted curve must be continuous (top right, bottom right).

Additional smoothness can be introduced by adding additional constraints: the first and second derivatives of the piecewise polynomials are also continuous at the knots (bottom left).

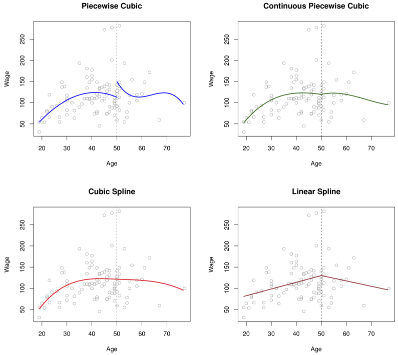

Figure 5: Fig. 7.3

Regression Splines

To reduce variance in the regions before the first knot and after the last knot, we can add an additional constraint that the function is required to be linear at the boundaries. This is known as a *natural spline*.

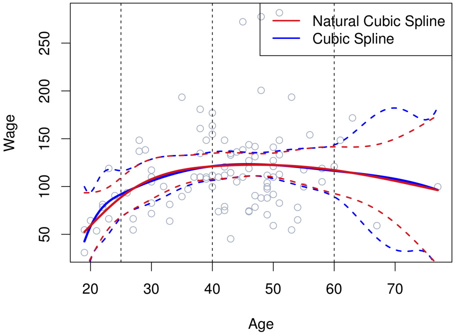

Figure 6: Fig. 7.4

Smoothing Splines

An alternative way to fit a smooth curve to a set of data is to fit a function that minimizes the "Loss + Penalty" formulation we discussed in Chp. 6:

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

Smoothing Splines

An alternative way to fit a smooth curve to a set of data is to fit a function that minimizes the "Loss + Penalty" formulation we discussed in Chp. 6:

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

Here, λ is a nonnegative tuning parameter, g is a smoothing spline, and $g''(t)$ measures the amount by which the slope of g is changing at t.

The $g(x)$ that minimizes the formulation above is a shrunken version of a natural cubic spline with knots at every x.

Smoothing Splines

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

What does g look like when $\lambda = 0$? When $\lambda \to \infty$?

Summary

To provide a non-linear fit to data, we can use **polynomial regression** (global) or break up the range of a variable into different regions, which can be fit locally with constant (**step function**), linear, or other power functions. We can introduce additional constraints to connect and smooth these functions (**regression splines**).

Alternatively, we can optimize the “Loss + Penalty” objective of machine learning (**smoothing splines**) or perform weighted regression within the neighborhood of a point of interest (**local regression**).

Next class:

- Today we have investigated non-linear relationships of a response with a single predictor.
- GAMs are a multivariate extension of these approaches.
  - Please read Chp. 7.6 on GAMs
  - Deleo *et al.* (2019) on GAMs for spatio-temporal analysis of trait variation data