LECTURE 5, PART I: BASICS OF SPLINES

Text references: Chapter 8 in Shalizi

Big Picture: Like kernel and kNN regression, splines provide a flexible way of estimating the underlying regression function $r(x) = \mathbb{E}[Y|X = x]$. As we shall see, splines can be seen as extensions of linear models.

Consider the regression model

$$Y_i = r(X_i) + \epsilon_i$$

and assume for simplicity that the covariates X_i ($i = 1, \ldots, n$) are one-dimensional. Suppose we estimate r by choosing $\widehat{r}(x)$ to minimize the sums of squares

$$\sum_{i=1}^{n} (Y_i - \widehat{r}(X_i))^2,$$

over a class of functions. Consider two extreme cases:

1. Minimizing over all linear functions (i.e., functions of the form $\beta_0 + \beta_1 x$) yields the *least squares estimator*,
2. Minimizing over all functions yields a function that *interpolates* the data.

In previous lectures, we avoided these two extreme solutions by controlling how smooth we made $\widehat{r}(x)$ indirectly through the bandwidth of a kernel smoother or the number of nearest neighbors in kNN-regression. Alternatively, we can get a solution in between these extremes by controlling smoothness itself. More specifically, by minimizing the **spline objective function** (which is a **penalized sum of squares**):

The penalty term leads to a solution that favors smoother functions. Adding a penalty term to the criterion we are optimizing is sometimes called **regularization**. The parameter λ controls the trade-off between fit (the first term of the equation above) and the penalty. Let $\widehat{r}$ denote the function that minimizes $\mathcal{L}(\lambda)$. The parameter λ controls the amount of smoothing. What is the solution when $\lambda = 0$?

What is the solution when $\lambda \to \infty$? (How is this different from a kernel smoother with $h \to \infty$?)

What does $\widehat{r}$ look like for $0 < \lambda < \infty$? To answer the last question, we need to define *splines*, which are special piece-wise polynomials.

Piecewise polynomials.

Consider first fitting a piecewise cubic polynomial with a single *knot* (point where the coefficients change) at a point c; such a model take the following form:

Each of these polynomial functions can be fit using least squares, and adding more knots leads to a more flexible piecewise polynomial. However, from the top left panel of Figure 1, we immediately see a problem: the function is discontinuous and lead to a poor fit of the data! How many degrees of freedom has this piecewise polynomial model in total?

Constraints and Splines.

A spline is a piece-wise polynomial with additional constraints. In words, a *kth order spline* is a piecewise polynomial function of degree k, that is continuous with continuous derivatives of orders $1, \ldots k-1$ at its knot points. Formally, a function $f : \mathbb{R} \to \mathbb{R}$ is a kth order spline with knot points at $t_1 < \ldots < t_m$, if

- f is a polynomial of degree k on each of the intervals $(-\infty, t_1], [t_1, t_2], \ldots [t_m, \infty)$, and
- $f^{(j)}$, the jth derivative of f, is continuous at $t_1, \ldots t_m$, for each $j = 0, 1, \ldots k-1$.

The most commonly considered splines are cubic splines (that is, the case $k = 3$). A **cubic spline** is a piecewise polynomial function r such that

How many degrees of freedom has a cubic spline in total?

Note that the continuity in all of their lower order derivatives makes splines very smooth (see Figure 1, bottom left). Indeed, it is usually hard to detect the locations of the knots of a cubic spline by eye!

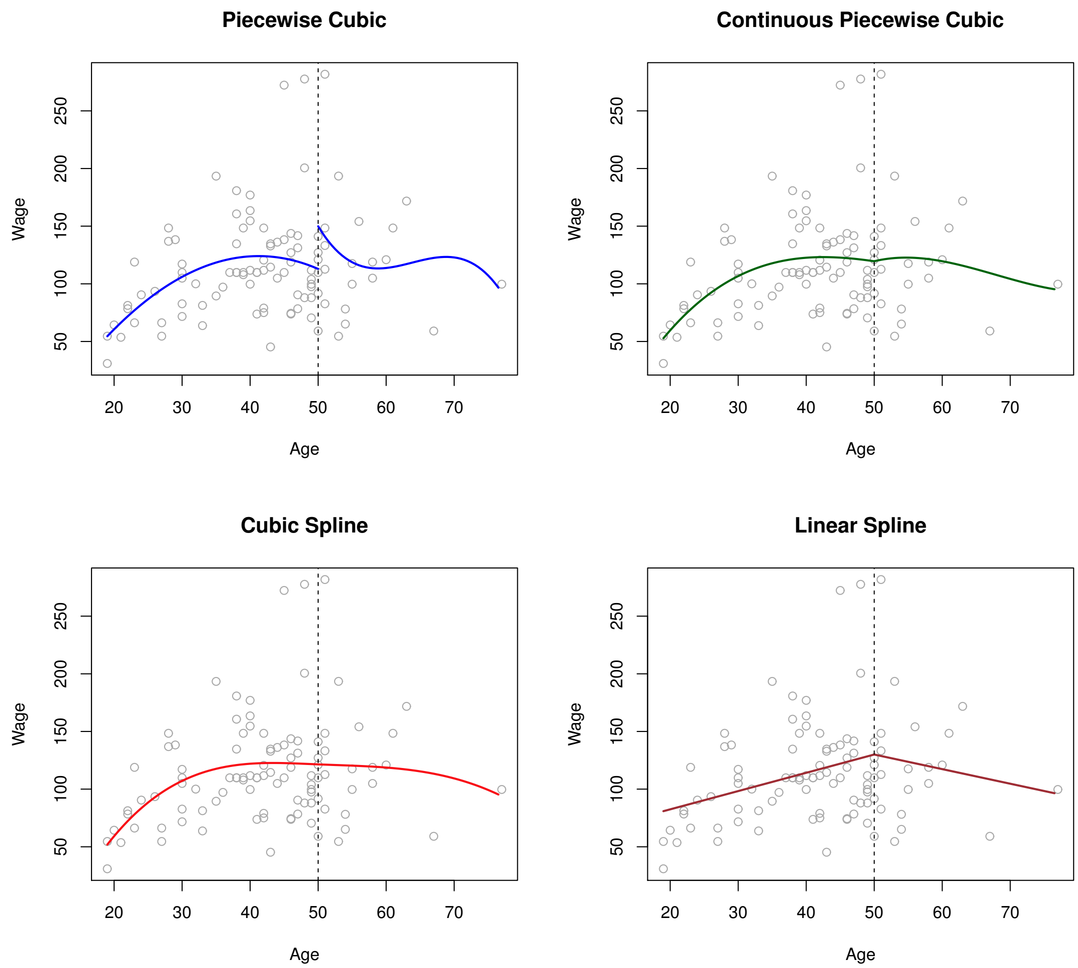

Figure 1: Various piecewise polynomials are fit to a subset of the `Wage` data, with a knot at `age=50`. *Top Left:* The cubic polynomials are unconstrained. *Top Right:* The cubic polynomials are constrained to be continuous at `age=50`. *Bottom Left:* The cubic polynomials are constrained to be continuous, and to have continuous first and second derivatives. *Bottom Right:* A linear spline is shown, which is constrained to be continuous. [Ref: ISL]

Natural Splines.

One problem with regression splines is that the estimates tend to display erractic behavior, i.e., they have high variance, at the boundaries of the domain of $x_1, \dots x_n$; see Figure 2. This gets worse as the order k gets larger.

A way to remedy this problem is to add *boundary constraints*, and force the piecewise polynomial function to have a lower degree to the left of the leftmost knot, and to the right of the rightmost knot—this is exactly what *natural splines* do. A natural spline of order k, with knots at $t_1 < \dots < t_m$, is a piecewise polynomial function f such that

- f is a polynomial of degree k on each of $[t_1, t_2], \dots [t_{m-1}, t_m]$,
- f is a polynomial of degree $(k-1)/2$ on $(-\infty, t_1]$ and $[t_m, \infty)$,
- f is continuous and has continuous derivatives of orders $1, \dots k-1$ at its knots $t_1, \dots t_m$.

It is implicit here that natural splines are only defined for odd orders k. **Natural cubic splines** ($k = 3$) are cubic splines which are *linear beyond the boundary knots*; these are the most common splines used in practice.

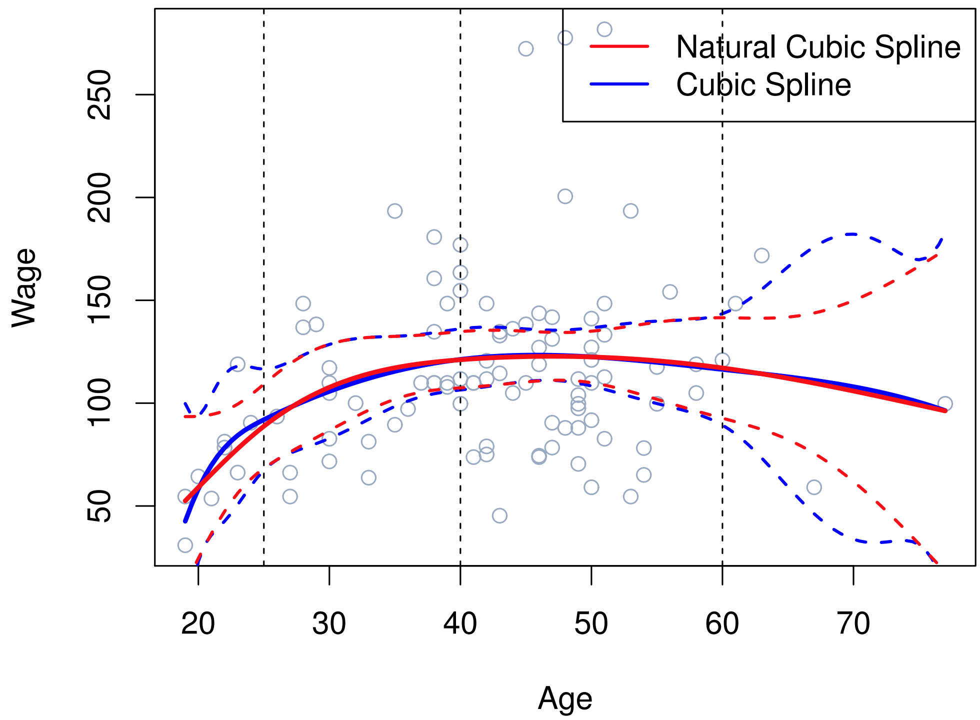

Figure 2: A cubic spline and a natural cubic spline, with three knots, fit to a subset of the `Wage` data. Splines can have high variance at the outer range of the predictors, which is indicated by the wide confidence bands (blue dashed lines). The additional boundary constraint on natural splines produces more stable estimates at the boundaries, which is here indicated by the narrower confidence bands (red dashed lines). [Ref: ISL]

<u>Follow-up question:</u> How do you use splines for regression in practice? In Part II of Lecture 5 we will discuss how to fit a "regression spline" (that is, a piecewise degree-d polynomial under mentioned constraints) to data. We will then also introduce so-called **smoothing splines**, which are *natural cubic splines with knots at the data points*; the latter splines arise naturally in the penalized regression framework.