

Chapter 7: Moving Beyond Linearity

So far we have mainly focused on linear models.

Previously, we have seen we can improve upon least squares using ridge regression, the lasso, principal components regression, and more.

Through simple and more sophisticated extensions of the linear model, we can relax the linearity assumption while still maintaining as much interpretability as possible.

1 Step Functions

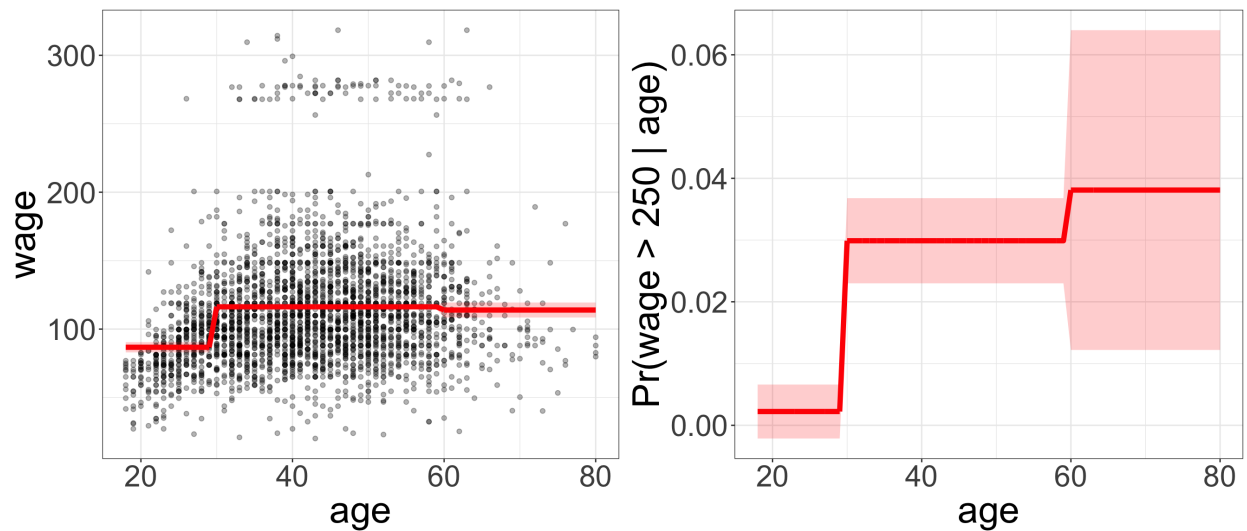
Using polynomial functions of the features as predictors imposes a *global* structure on the non-linear function of X .

We can instead use *step-functions* to avoid imposing a global structure.

For a given value of X , at most one of C_1, \dots, C_K can be non-zero.

Example: Wage data.

year	age	maritl	race	edu- cation	region	job- class	health	health_ins	logwage	wage
2006	18	1. Never Mar- ried	1. White	1. < HS Grad	2. Mid- dle At- lantic	1. Indus- trial	1. <=Good	2. No	4.318063	75.04315
2004	24	1. Never Mar- ried	1. White	4. Col- lege Grad	2. Mid- dle At- lantic	2. Infor- ma- tion	2. >=Very Good	2. No	4.255273	70.47602
2003	45	2. Mar- ried	1. White	3. Some Col- lege	2. Mid- dle At- lantic	1. Indus- trial	1. <=Good	1. Yes	4.875061	130.98218
2003	43	2. Mar- ried	3. Asian	4. Col- lege Grad	2. Mid- dle At- lantic	2. Infor- ma- tion	2. >=Very Good	1. Yes	5.041393	154.68529



2 Basis Functions

Polynomial and piecewise-constant regression models are in fact special cases of a *basis function approach*.

Idea:

Instead of fitting the linear model in X , we fit the model

Note that the basis functions are fixed and known.

We can think of this model as a standard linear model with predictors defined by the basis functions and use least squares to estimate the unknown regression coefficients.

3 Regression Splines

Regression splines are a very common choice for basis function because they are quite flexible, but still interpretable. Regression splines extend upon polynomial regression and piecewise constant approaches seen previously.

3.1 Piecewise Polynomials

Instead of fitting a high degree polynomial over the entire range of X , piecewise polynomial regression involves fitting separate low-degree polynomials over different regions of X .

For example, a piecewise cubic with no knots is just a standard cubic polynomial.

A piecewise cubic with a single knot at point c takes the form

Using more knots leads to a more flexible piecewise polynomial.

In general, we place K knots throughout the range of X and fit $K + 1$ polynomial regression models.

3.2 Constraints and Splines

To avoid having too much flexibility, we can *constrain* the piecewise polynomial so that the fitted curve must be continuous.

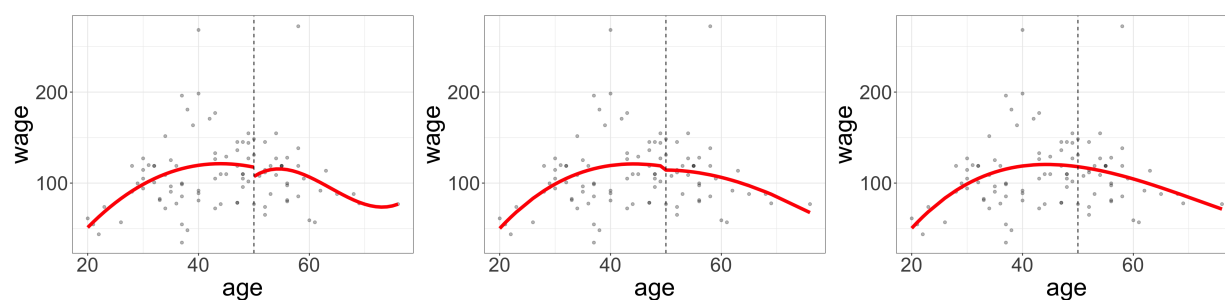
To go further, we could add two more constraints

In other words, we are requiring the piecewise polynomials to be *smooth*.

Each constraint that we impose on the piecewise cubic polynomials effectively frees up one degree of freedom, but reducing the complexity of the resulting fit.

The fit with continuity and 2 smoothness constraints is called a *spline*.

A degree- d spline is



3.3 Spline Basis Representation

Fitting the spline regression model is more complex than the piecewise polynomial regression. We need to fit a degree d piecewise polynomial and also constrain it and its $d - 1$ derivatives to be continuous at the knots.

The most direct way to represent a cubic spline is to start with the basis for a cubic polynomial and add one *truncated power basis* function per knot.

Unfortunately, splines can have high variance at the outer range of the predictors. One solution is to add *boundary constraints*.

3.4 Choosing the Knots

When we fit a spline, where should we place the knots?

How many knots should we use?

3.5 Comparison to Polynomial Regression

