Splines

BY LANCE DAVIS

Overview

What are splines?

Types of splines

Using splines in R

References

What are Splines?

Splines are a way to add flexibility to a model when dealing with data that is not linear.

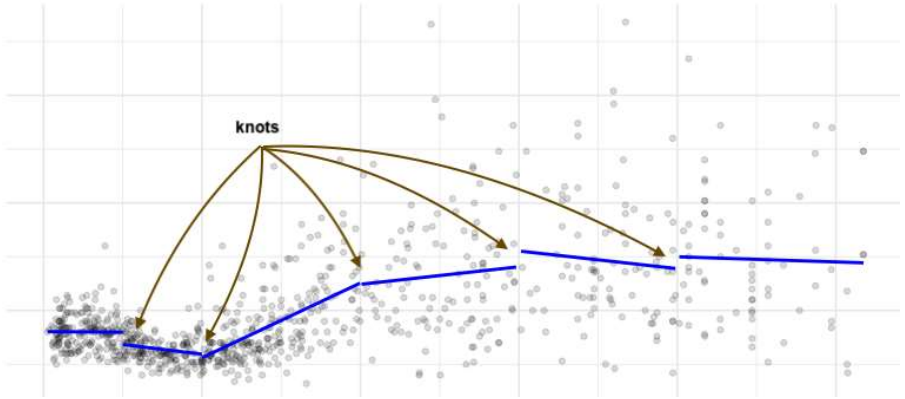
Can utilize knots to break up a data set into multiple parts that can create a better fit.

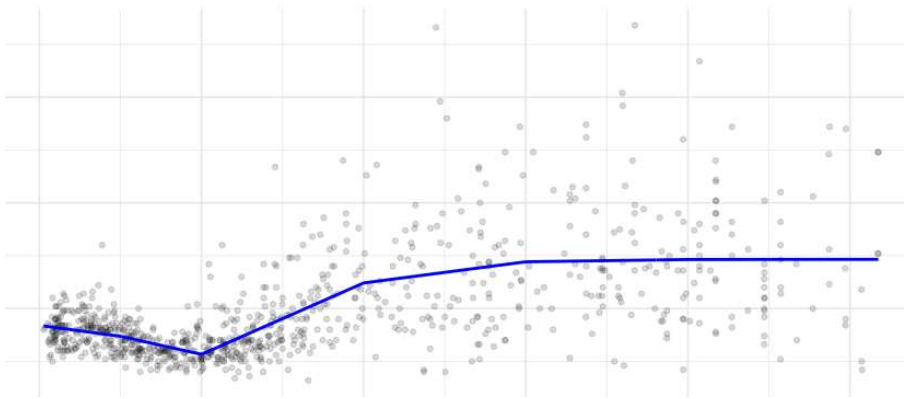
Example of a cubic spline model with K knots.

$$y_i = \beta_0 + \beta_1 b_1(x_i) + \beta_2 b_2(x_i) + \cdots + \beta_{K+3} b_{K+3}(x_i) + \epsilon_i,$$

- b_1 , b_2, and so on are the spline basis functions.

What are Splines?

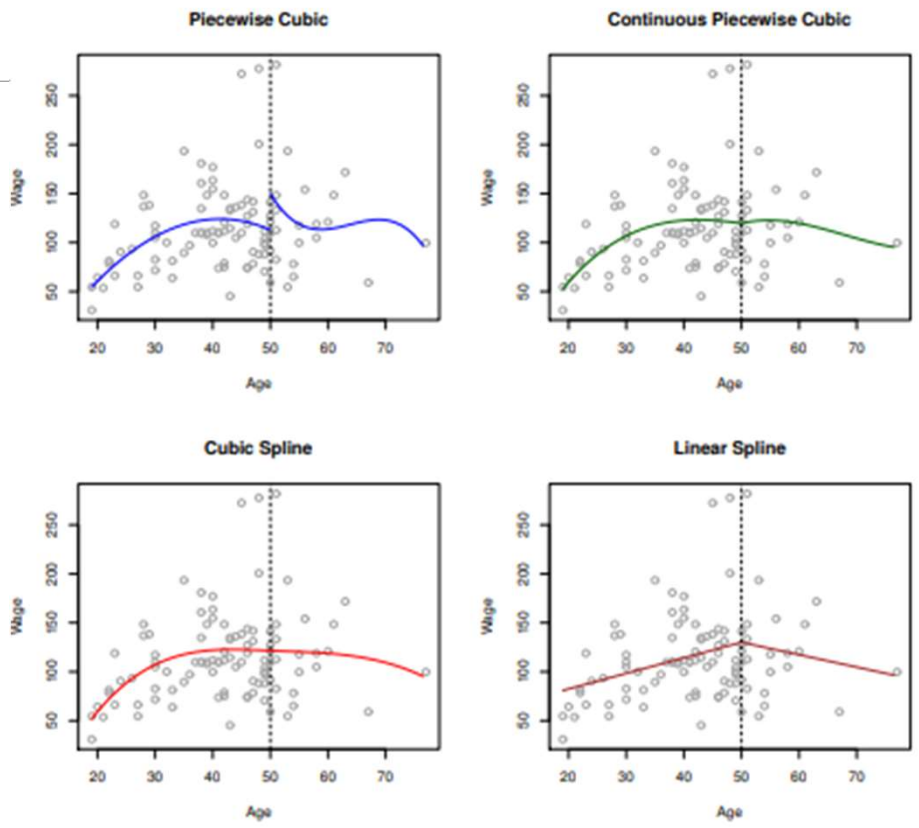




Types of Splines

Piecewise Polynomials

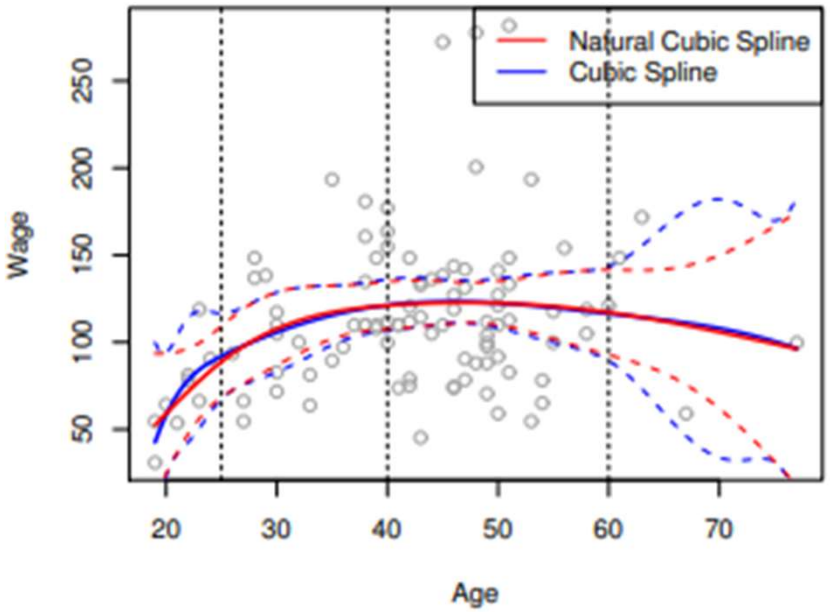
- Involves fitting separate low-degree polynomials over different regions
- Knot at age 50.



Types of Splines

Piecewise polynomial cont.

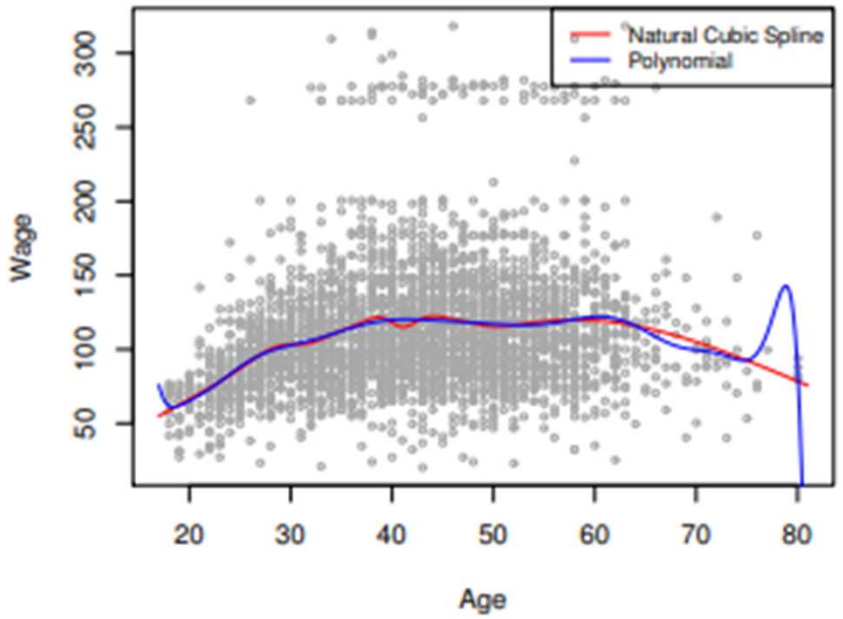
- Multiple knots
- Cubic and natural cubic
  - Natural cubic splines have additional constraints – required to be linear at the boundary
- How many knots and where to put them?
  - Cross Validation



Quick comparison with polynomial regression

Natural cubic spline with 15 df compared to degree-15 polynomial

- Polynomials can have wild swings at the tails.

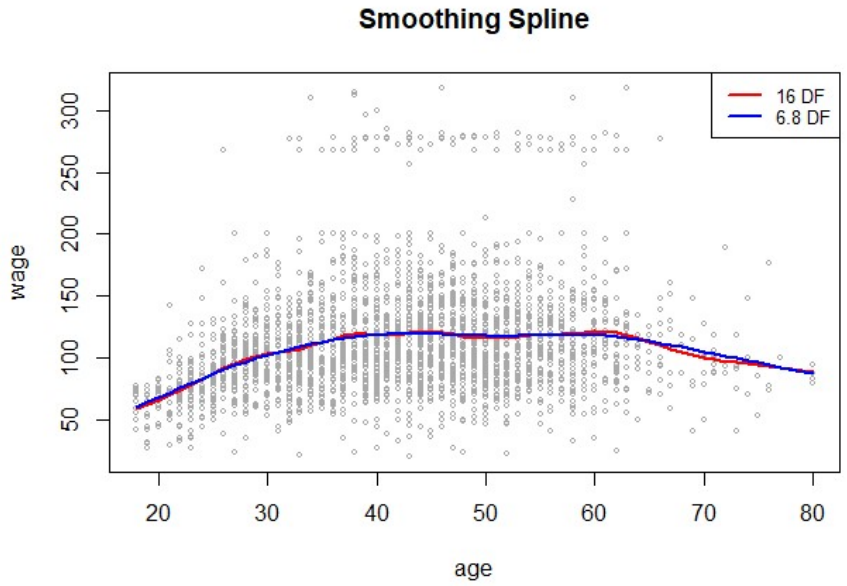


Smoothing Splines

Uses a tuning parameter to help smooth.

How to choose the right tuning parameter?

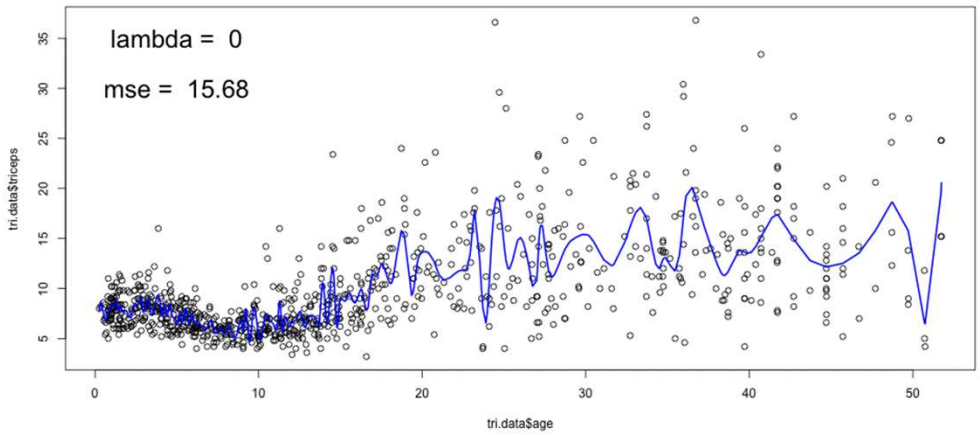
- Cross validation



Smoothing Splines

You can see how changing the tuning parameter (lambda) changes the fit

- Overfit at lambda of 0
- Underfit at lambda of 100

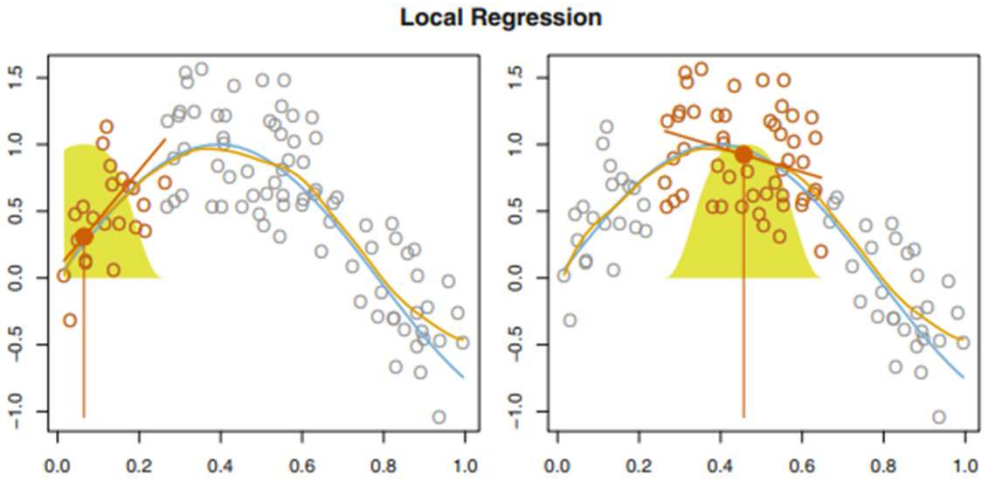


Local Regression

Fit a weighted least squares regression over a specified span with points closest to x_0 having the highest weighting.

How can we choose a span?

- You guessed it – Cross Validation



Splines in R

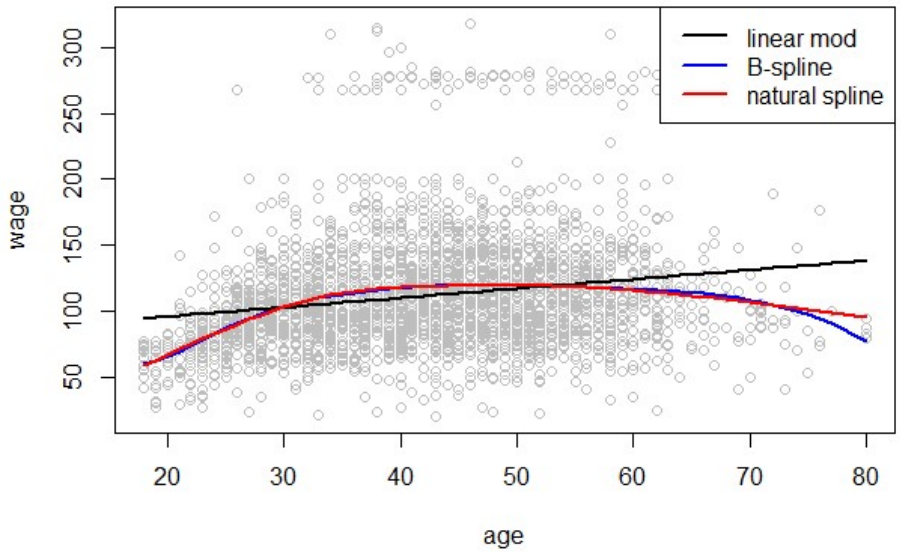
Regression splines

- Bs() function generates basis functions for splines with specified knots

```
fit <- lm(wage ~ bs(age , knots = c(25, 40, 60)), data = Wage)
```

- Ns() function fits a natural spline

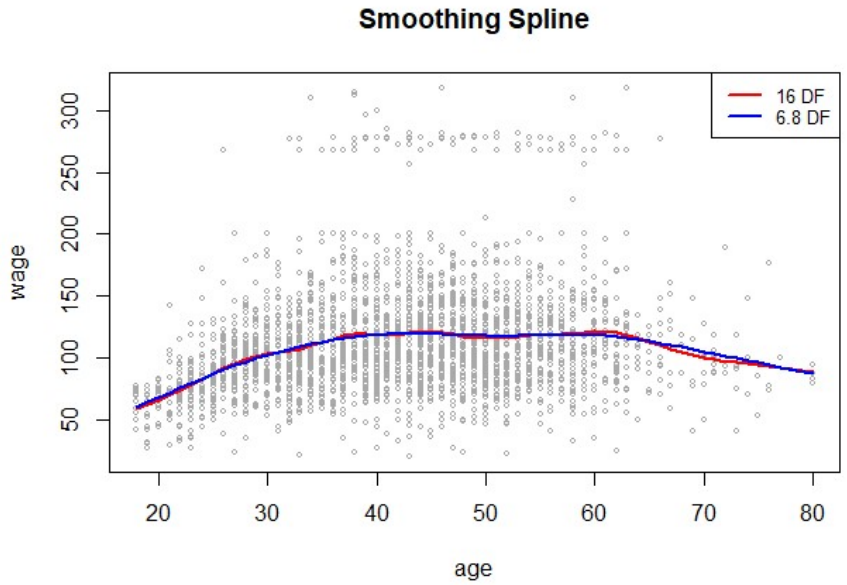
```
fit2 <- lm(wage ~ ns(age , df = 4), data = Wage)
```



Splines in R

For Smoothing splines use the smooth.spline function

```
fit3 <- smooth.spline(age, wage, df = 16)
fit4 <- smooth.spline(age, wage, cv = TRUE)
```



Splines in R

Can also be used in GAMs

- S() function for spline
- Lo() function for local regression

```
gam1 <- gam(wage ~ s(year, 4) + s(age, 5) + education, data = Wage)
gam2 <- gam(wage ~ s(year, df = 4) + lo(age, span=0.7) +
              education, data = Wage)
```

References

Introduction to Statistical learning with Applications in R

https://bookdown.org/tpinto_home/Beyond-Linearity/piecewise-regression-and-splines.html

https://jbhender.github.io/Stats506/F18/GP/Group9.html