7.5 Smoothing Splines

Model Fitting

- In general, we want to find some function that fits the data well.
- That is, we want

$$\text{RSS} = \sum_{i=1}^{n} (y_i - g(x_i))^2$$

  to be small.
- However, if this is our only criteria, we can always choose g to *interpolate* all of y.
  - That is, g will be such that RSS ≈ 0, a massive overfit!
- What we want is g such that RSS is small *and* g is *smooth*.

Smoothing Splines

- How can we ensure g is smooth?
- One approach: find G to minimize

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

  where λ is some nonnegative tuning parameter.
  - The g that minimizes this is a *smoothing spline*

Smoothing Splines

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

- We've seen similar "loss + penalty" formulations before (ridge, lasso)
  - The loss function is about fit to data.
  - The penalty is about reducing the variability/flexibility.
- That second derivative is, roughly, a measure of the wiggliness of the function.
  - It's high if g is very wiggly near t and close to 0 if it's very smooth.

Smoothing Splines

- When $\lambda = 0$, g will exactly interpolate the training data.
- As $\lambda \to \infty$, g will converge to the least squares regression line (a line is perfectly smooth).
- So λ controls the bias-variance tradeoff.

Smoothing Splines: Properties

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

- is a piecewise cubic polynomial with knots at all unique values of $x_1, \ldots, x_n$
- has continuous first and second derivatives at each knot
- is linear in the region outside the extreme knots
  - so this is a natural cubic spline with specific knots
  - In fact, it's a *shrunken* version of the natural cubic spline from Section 7.4

Choosing λ

- Knots at each unique x seems like too many degrees of freedom.
- However, the value of λ is going to control the degrees of freedom by controlling the roughness.
- As λ increases from 0 to ∞, effective degrees of freedom df_λ decreases from n to 2.

Degrees of Freedom

- Usually, degrees of freedom refer to the number of free parameters in the model, such as the number of coefficients estimated.
- A smoothing spline has n parameters and so n nominal degrees of freedom, but these are heavily constrained.
- *Effective degrees of freedom* is a measure of the flexibility of the smoothing spline

Effective Degrees of Freedom

- We can write

$$\hat{g}_\lambda = S_\lambda y$$

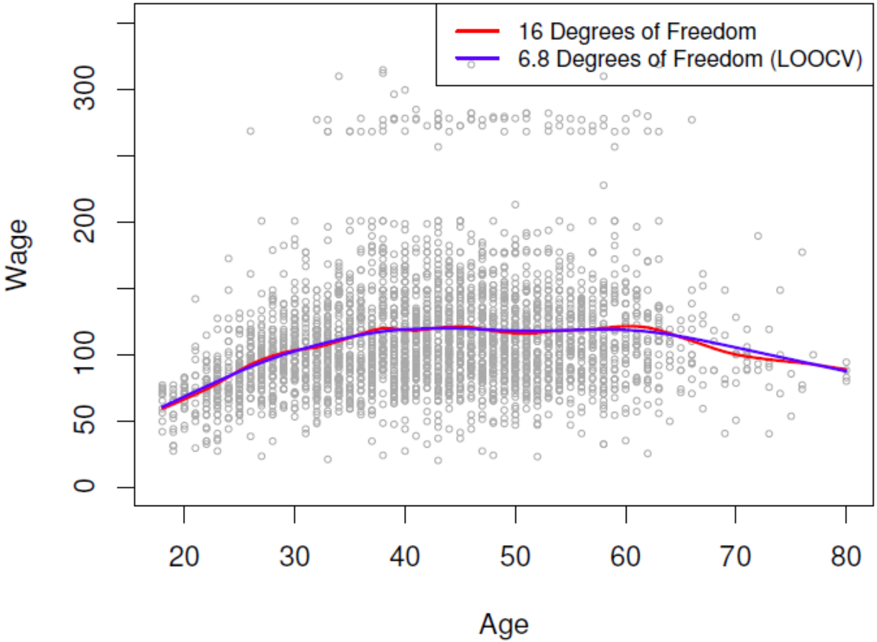
  where $\hat{g}_\lambda$ is the solution to the prior optimization problem.
  - This is a n-vector with the fitted values of the smoothing spline for the training data.
- Then effective degrees of freedom is defined as

$$df_\lambda = \sum_{i=1}^{n} \{S_\lambda\}_{ii}$$

  the sum of the diagonal elements of S_λ.

Selecting λ

- Now, instead of choosing the number of knots, we need to select the value of λ.
- Like the other constrained optimization problems, we use cross validation.
- It turns out the LOOCV error can be computed for smoothing splines using a direct formula
$$\text{RSS}_{\text{CV}}(\lambda) = \sum_{i=1}^{n} \left(y_i - \hat{g}_\lambda^{(-i)(x_i)} \right)^2 = \sum_{i=1}^{n} \left[\frac{y_i - \hat{g}_\lambda(x_i)}{1 - \{S_\lambda\}_{ii}} \right]^2$$
  - The notation $\hat{g}_\lambda(x_i)$ indicates the fitted value at x_i where the fit uses all the training observations *except* the ith.


16 Degrees of Freedom
6.8 Degrees of Freedom (LOOCV)
Wage
Age
0
50
100
200
300
20
30
40
50
60
70
80