

5.6 Activities

Exercise 1

In this exercise we follow the Lab notes from Section 7.8 of James et al. (2013), i.e. the ISLR book. We will skip Section 7.8.1 on polynomial regression and step functions but it is recommended to read through this section (combined with the course notes). Instead we focus on section 7.8.2 to learn how to fit regression splines, natural splines and smoothing splines.

We will focus on the **Wage** dataset available in the **R** package **ISLR**.

1. Consider a spline regression to fit **wage** to **age** considering knots at 25, 40 and 60. Represent the estimated regression line for a sequence of age values of length 100 ranging from the minimum to the maximum observed values. Add a confidence interval using ± 2 standard error.
2. How many basis function were used in the previous question? Now, instead of specifying the locations of the knots, use the **df** and **degree** arguments to specify a quadratic spline regression with 6 degrees of freedom. Fit the model and display it as in 1). Where were the knots located?
3. Use the **ns()** function to fit a natural cubic spline with 6 degree of freedom. Superimpose it with the spline regression from 1). What is the main comment that you can make when comparing both models?
4. Use the **smooth.spline()** function to fit a smoothing spline to the data. In the first place use the argument **df = 16** allowing for a 16 degrees of freedom. In a second step, consider instead the argument **cv = TRUE** to use cross-validation to select the optimal degree of freedom. Display both fitted lines. What do the respective models do?

