STAT 447B 2014–15 Assignment 1

Will Welch

Due: Friday, September 19, 2014

1. (See Lecture 3 slides.) We want to fit a linear-splines regression model relating a response variable y to a single explanatory variable x. The spline model has one knot at κ_1, which is a known number. Consider the following model to explain the way the data are generated:

$$\begin{aligned} Y &= \beta_0 + \beta_1 x + \epsilon \quad \text{(for data with } x \leq \kappa_1\text{)} \\ Y &= \beta_2 + \beta_3 x + \epsilon \quad \text{(for data with } x > \kappa_1\text{)}, \end{aligned}$$

   where ϵ represents random error with expectation 0.

   (a) We want $E(Y \mid x)$ to be continuous at $x = \kappa_1$. What constraint is imposed on the parameters? How many free (linearly independent) parameters does the regression model have?
   (b) Hence rewrite the above model in terms of free parameters only.
   (c) Suppose $\kappa_1 = 10$, and the first 5 values of x in the data set are

$$6, 12, 29, 10, 18.$$

   Write down the first 5 rows of the expanded-design matrix $\mathbf{X}$ that would be passed to `lm` to fit the linear-splines regression model *as we have specified it here*. (There is no need to include a column of 1's in $\mathbf{X}$.)

2. (For this question, please submit your `R` code as well as well as your results.) Install the `R` package `ISLR` and make the `Wage` data available:

```
library(ISLR)
data(Wage, package = 'ISLR')
```

   (Note the capitalization of `Wage` here.) We will fit two regression models relating the response variable `logwage` to the explanatory variable `age`.

   (a) Fit a simple linear regression model and comment on a summary of the fit.
   (b) Fit a linear-splines regression model with one knot at the median value of `age`. (Set up the model as we have done in class, not as in Question 1.) Comment on a summary of the fit. In particular, explain whether the effect of changing `age` by 1 year seems to have a constant effect on the expectation of `logwage` over the range of `age` values in the data.