Practical 6

Statistical Learning, 2011

Niels Richard Hansen
September 21, 2011

Regression and basis expansions

This exercise uses the LA ozone data. You should also keep, as in a previous exercise, 2/3 of the data set as a training data set and 1/3 as a test data set.

1. Consider as in the book, page 155, the regression of ozone concentration on the daggot pressure gradient (dpg) alone. Use the `smooth.spline` function to reproduce a similar regression.

   In the following couple of questions you are asked to investigate if the suggested non-linearity from the one-dimensional regression remains if we do a multiple regression model of ozone concentration.

2. Use `bs` combined with ordinary least squares regression to estimate a full model where the dpg variable is expanded in the B-spline basis and the other variables have a linear effect. Try different choices of degrees of freedom and plot the resulting estimated effects of `bs`.

3. Do a formal ANOVA test where you test the above model against a model with only the linear effect of dpg.

4. Compute the training and test error using the full model with dpg expanded in a B-spline basis for a range of choices of degrees of freedom.