Homework 2

Due on 03/20/2018

In this exercise, we build nonlinear models using the concrete compressive strength data set. Concrete is the most important material in civil engineering. The concrete compressive strength is a nonlinear function of age and ingredients. These ingredients include cement, blast furnace slag, fly ash, water, superplasticizer, coarse aggregate, and fine aggregate. The data are in "concrete.csv". Concrete compressive strength is the response.

(a) Create scatter plots of response vs. predictors using the function `featurePlot()`.

(b) Perform polynomial regression to predict compressive strength using `water` as the predictor. For $1 \leq d \leq 4$, use cross-validation to select the optimal degree d for the polynomial. What degree was chosen, and how does this compare to the results of hypothesis testing using ANOVA? Make a plot of different polynomial fits to the data.

(c) Fit a **smoothing spline** using `water` as the predictor for a range of degrees of freedom, as well as the degree of freedom obtained by generalized cross-validation, and plot the resulting fits. Describe the results obtained.

(d) Fit a GAM using all the predictors. Plot the results and explain your findings.