Stat 897 Fall 2017 Data Analysis Assignment 8

Penn State

Due October 22, 2017

In this assignment we again use the College data found in the ISLR library, with the 600 observation training set (using the rest as the test data).

(a) Using out-of-state tuition as the response and the other variables as the predictors, perform forward stepwise selection on the training set in order to identify a satisfactory model that uses just a subset of the predictors. Please set the seed at 801 once at the beginning before choosing the training set.

(b) Fit a GAM on the training data, using out-of-state tuition as the response and the features selected in the previous step as the predictors. Use appropriate nonlinear components (e.g. natural splines, step functions) for the variables that need it as in the salary example from the book. Plot the results, and explain your findings. What nonlinear components did you use?

(c) Evaluate both models obtained from part (a) and (b) on the test set, and explain the results obtained.

(d) For which variables, if any, is there evidence of a non-linear relationship with the response?