

This question relates to the **College** data set.

- (a) Split the data into a training set and a test set. 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.
- (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. Plot the results, and explain your findings.
- (c) Evaluate the model obtained 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?