Homework 8

Due Tuesday, November 19

Show all your work. Data files available from `alr4` or `MASS` package, or on class website.

1. Consider the `mantel` data with the response `Y` and the predictors `X1`, `X2` and `X3`.
   (a) Find AIC and BIC for all possible models (8 in total, including a null model).
   (b) Perform the variable selections in R (try BE and FS, as well as the default and others). What happens? Select a model and defend your choice.
   (c) Find AIC, BIC and C_p for (1) null model, (2) full model, (3) the model that you selected in part (b), **using matrices**.

2. Consider the dataset `bodyfat.txt`. Note that the header is missing from this dataset. The description of the data is as follows

```
The variables listed below, from left to right, are:

 Percent body fat
 Density (weight per volume) determined from underwater weighing
 Age (years)
 Weight (lbs)
 Height (inches)
 Neck circumference (cm)
 Chest circumference (cm)
 Abdomen 2 circumference (cm)
 Hip circumference (cm)
 Thigh circumference (cm)
 Knee circumference (cm)
 Ankle circumference (cm)
 Biceps (extended) circumference (cm)
 Forearm circumference (cm)
 Wrist circumference (cm)
```

   The purpose of this study is to see what variables (if any) contribute significantly to Percent body fat, in a multiple regression model. Perform the analysis (including selection of variables, transformations and diagnostic analysis), and present and defend your final model.

3. For the dataset `BGSboys`, fit a model with `BMI18` as a response and variables for ages 9 or earlier as predictors (variable with 2 or 9 at the end). Perform a complete analysis, including selection of variables, transformations and diagnostic analysis, and summarize your results.

4. Please find the dataset `Boston`. Using `medv` as the response, please perform a complete analysis. Defend and interpret your model, and comment.

5. Find the (leave-one-out) cross validation errors for Problem 1 (8 in total, including a null model).