

HW10, Question 1

The dataset `fat` (type `library(faraway)` followed by `data(fat)`) consists of 18 measurements in 252 samples. The variables `brozek`, `siri` and `density` are different ways of measuring total body fat, while the remaining 15 variables are different variables that may be helpful in predicting body fat. For this exercise, take `siri` as the variable of interest, omit `brozek` and `density`, and use the other 15 variables as potential covariates. Also, split the data into a test dataset consisting of every tenth sample (i.e. rows 10, 20, ..., 250 of the data) and the training dataset which is everything else.

Fit a linear regression to the training data by each of the following methods:

- (a) Linear regression fitted to all 15 covariates;
- (b) Linear regression fitted to a subset of covariates using variable selection;
- (c) Principal components regression;
- (d) Partial least squares regression;
- (e) Lasso regression.

For each method, you should use your own judgment about any decisions you take along the way (e.g. how many covariates to include in a variable selection) but say explicitly what choices you are making.

Then, compare each of the methods by using them to predict the test data, using the mean squared error between the predictions of `siri` and the test values as a comparison criterion. What are your conclusions about the relative performances of the different methods?

HW10, Question 2

See the file “Logistic Regression Exercises” in the Resources>Homework Assignments folder on sakai; do questions 8.15 and 8.17. Answer all parts, but in particular:

- (a) For 8.15, verify the parameter estimates for both models and carry out an anova test between them
- (b) For 8.17, calculate the probability that a male possum with a skull 63 mm wide, a tail 37 cm long, and a total length 83 cm is from Victoria. Also quote a 95% confidence interval for that probability.

The data file possum.csv is in the Resources>Data folder on sakai

The slides for the logistic regression lectures are in the file LogisticRegressionSlides.pdf in the Homework Assignments folder on sakai.