STATS 500 - Homework 8

Due Wednesday, November 25

Based on Problem 4., page 180 in Faraway

Take the `fat` data, and use the percentage of body fat as the response and the other variables as potential predictors. Remove every tenth observation from the data for use as a test sample. Use the remaining data as a training sample building the following models:

1. Linear regression with all predictors

2. Linear regression with variables selected using AIC

3. Principal component regression

4. Partial least squares

5. Ridge regression

Use the models you find to predict the response in the test sample. Make a report on the performance of the models.

Hints:

Background: The data on the percentage of body fat, age, weight, height, and ten body circumference measurements (e.g., abdomen) are recorded for 252 men. Body fat is estimated through an underwater weighing technique, but this is inconvenient to use widely.

The columns in the dataset are:

1. Percent body fat using Brozek's equation, "457/Density − 414.2"

2. Percent body fat using Siri's equation, "495/Density − 450"

3. Density (gm/cm^3)

4. Age (yrs)

5. Weight (lbs)

6. Height (inches)

7. Adiposity index = Weight/Height^2(kg/m^2)

8. Fat Free Weight = (1 − fraction of body fat) × Weight, using Brozek's formula (lbs)

9. Neck circumference (cm)

10. Chest circumference (cm)

11. Abdomen circumference (cm) "at the umbilicus and level with the iliac crest"

12. Hip circumference (cm)

13. Thigh circumference (cm)

14. Knee circumference (cm)

15. Ankle circumference (cm)

16. Extended biceps circumference (cm)

17. Forearm circumference (cm)

18. Wrist circumference (cm) "distal to the styloid processes"

Your models should predict body fat according to `siri`. Do not use Brozek's body fat (`brozek`), density (`density`) or Fat Free Weight (`free`) as predictors. You also need to remove every tenth observation from the data for use as a test sample. So you may start with something like:

```
> library(faraway)
> data(fat)
> index <- seq(10, 250, by=10)
## Extract data and remove ``brozek'', ``density'' and ``free''
> train <- fat[-index, -c(1, 3, 8)]
> test <- fat[index, -c(1, 3, 8)]
```