STA 5207 Assignment 9

Due Friday November 12

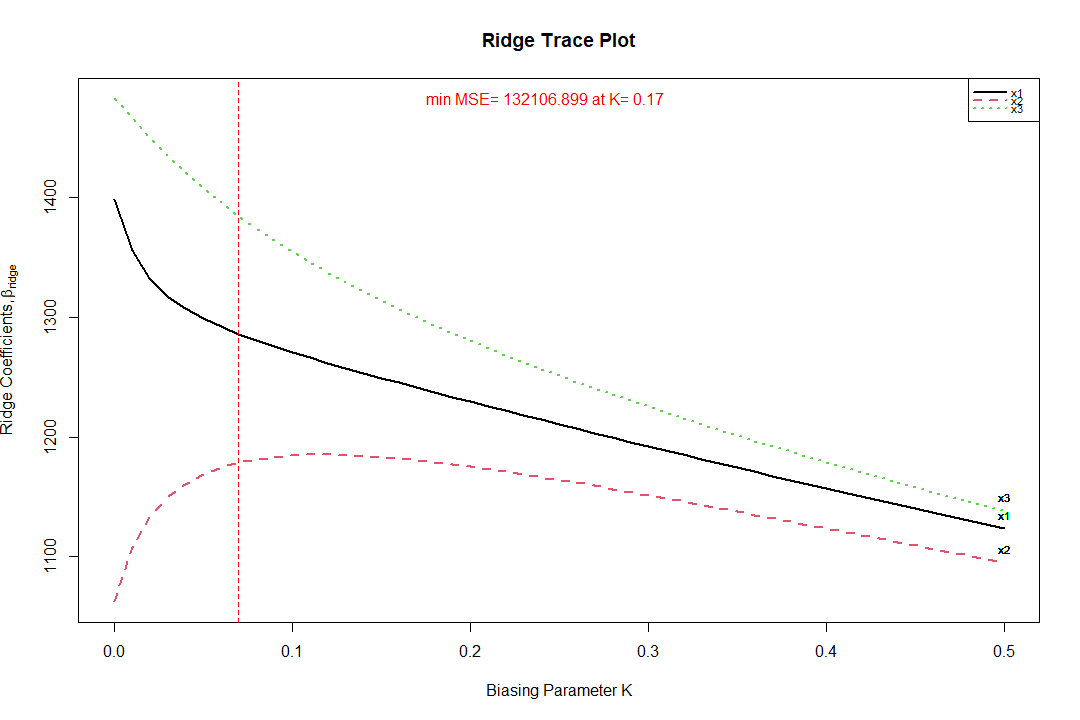
1. (50 points total) Consider the data set Assignment.txt. There is a response variable and 3 predictors:

*y*: number of worker hours over a month

*x*1: number of cases handled

*x*2: eligible population for the hospital

*x*3: number of operating rooms

1. (40 points) Perform ridge regression and give the fitted model for a suitably chosen value of λ. There is not a single correct answer, but you must explain the reasons for your choice. No output required.
   1. After performing ridge regression and checking the VIF’s, the first, and lowest, value of λ where all VIF’s are less than 10 **and** the changes in the estimates are small as λ increases is λ = 0.07.
2. (10 points) Submit the plot of the standardized coefficient estimates.
3. (50 points total) The data set has prostate in the faraway package has 97 observations and 8 predictors. A study on 97 men with prostate cancer who were due to receive a radical prostatectomy. The data set has the following variables:

lcavol: log(cancer volume)

lweight: log(prostate weight)

age: age

lbph: log(benign prostatic hyperplasia amount)

svi: seminal vesicle invasion

lcp: log(capsular penetration)

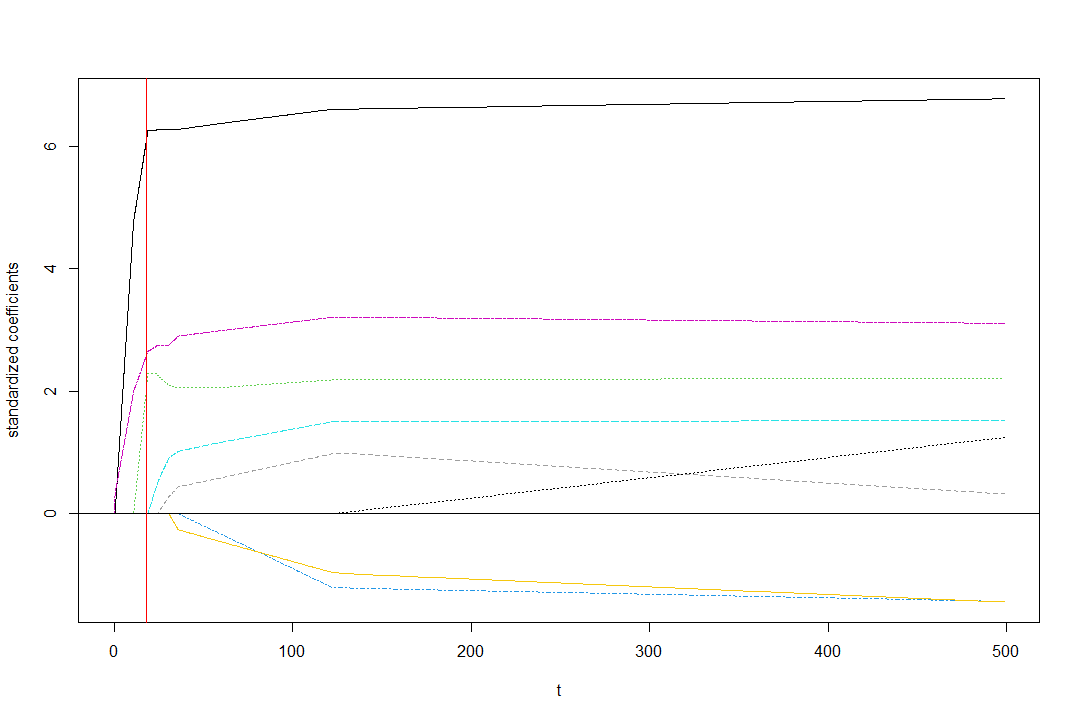
gleason: Gleason score

pgg45: percentage Gleason scores 4 or 5

lpsa: log(prostate specific antigen)

This data set is prostate.txt in SAS Studio and includes observation number as the first column.

1. (40 points) Use lpsa as the response and the other variables as predictors. Perform Adaptive Lasso using BIC as the criteria and give the fitted model. No output required. If you are using SAS, make sure that you make a note of this because PROC GLMSELECT can give a slightly different solution than the R function.



1. (10 points) Submit the plot of the solution path of the standardized coefficient estimates.