**Alex Riles**

**DSCI 425 – Supervised Learning**

**Ridge, Lasso, and Elastic Net Regression (85 points)**

## Problem 1 – NUmBER OF APPLICATIONS RECEIVED BY COLLEGES

This problem is essentially problem #9 (pg. 263 of the text), although I am making some minor changes and additions to the tasks outlined in the text problem.

**Below is a description of the College data frame in the ISLR library.**

**U.S. News and World Report's College Data**

**Description**Statistics for a large number of US Colleges from the 1995 issue of US News and World Report.

**Usage**

College

**Format**A data frame with 777 observations on the following 18 variables.

Private - A factor with levels No and Yes indicating private or public university

**Apps - Number of applications received**

Percent Accepted = (Apps/Accepts) x 100%

**Accept - Number of applications accepted**

Enroll - Number of new students enrolled

Top10perc - Pct. new students from top 10% of H.S. class

Top25perc - Pct. new students from top 25% of H.S. class

F.Undergrad - Number of fulltime undergraduates

P.Undergrad - Number of parttime undergraduates

Outstate - Out-of-state tuition

Room.Board - Room and board costs

Books - Estimated book costs

Personal - Estimated personal spending  
PhD - Pct. of faculty with Ph.D.'s

Terminal - Pct. of faculty with terminal degree

S.F.Ratio - Student/faculty ratio

perc.alumni - Pct. alumni who donate

Expend - Instructional expenditure per student

Grad.Rate - Graduation rate

**Source**This dataset was taken from the StatLib library which is maintained at Carnegie Mellon University. The dataset was used in the ASA Statistical Graphics Section's 1995 Data Analysis Exposition.

**References**Games, G., Witten, D., Hastie, T., and Tibshirani, R. (2013) *An Introduction to Statistical Learning with applications in R*, [www.StatLearning.com](http://127.0.0.1:29902/help/library/ISLR/html/www.StatLearning.com), Springer-Verlag, New York

**Form a new data frames called College2 and College4**

College2 = data.frame(PctAccept=100\*(College$Accept/College$Apps),College[,-c(2:3)])  
  
The command above forms the response PctAccept and then removes from the original data the Accept and Apps variables. The command below forms a data frame with the log transformations applied to the variables that are grossly skewed to the right.

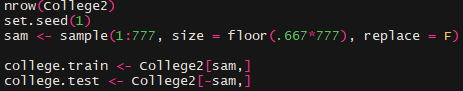
attach(College)  
  
College4 = data.frame(logApps=log(Apps),Private,logAcc=log(Accept),logEnr=log(Enroll),Top10perc,  
Top25perc,logFull=log(F.Undergrad),logPart=log(P.Undergrad),Outstate,Room.Board,Books,Personal,PhD,Terminal,S.F.Ratio,perc.alumni,logExp=log(Expend),Grad.Rate)

detach(College)

You will be using the data frame College2 for parts (a) – (h) of this problem with PctAccept as the response.

1. Split the data into a training set and a test set by forming the indices for the training and test sets. Use *p = .667*, i.e. use two-thirds of the data to train the models. ***Note: That none of commands below show fitting to just a training set!*** (1 pt.)

**Use set.seed(1) before splitting your data!**



1. Fit an OLS model for number of applications using the **training set**, and report the mean RMSEP for the test set. (4 pts.)

X = model.matrix(PctAccept~.,data=College2)[,-1]

y = College2$PctAccept

Xs = scale(X)

College2.temp = data.frame(y,Xs)

sam = sample(1:length(y),floor(.6667\*length(y)),replace=F)

PA.ols = lm(y~Xs,data=College2.temp,subset=sam)

ypred = predict(PA.ols,newdata=College2.temp[-sam,])

RMSEP.ols = sqrt(mean((y[-sam]-ypred)^2))

RMSEP.ols

RMSEP = 16.9856

1. Fit a sequence of ridge and lasso regression models on the **training** data set using the commands below given for the ridge models. The lambda sequence (grid) is formed to create the sequence of models. Create two plots showing the parameter shrinkage, one with the norm constraint on the x-axis and one with log lambda values on the x-axis. Discuss. (6 pts.)

grid = 10^seq(10,-2,length=200)

ridge.mod = glmnet(Xs,y,alpha=0,lambda=grid)

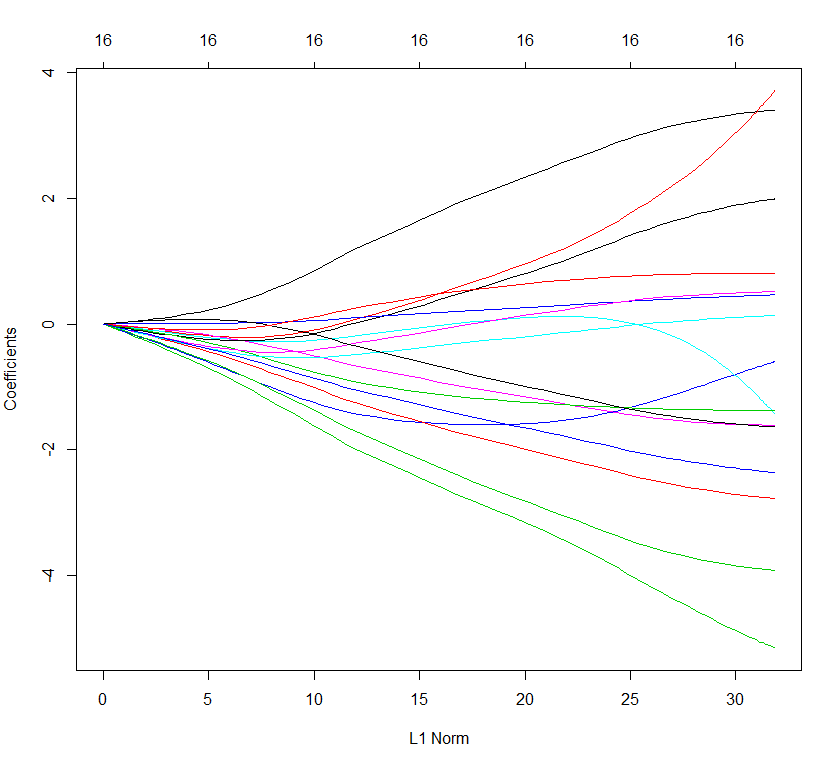
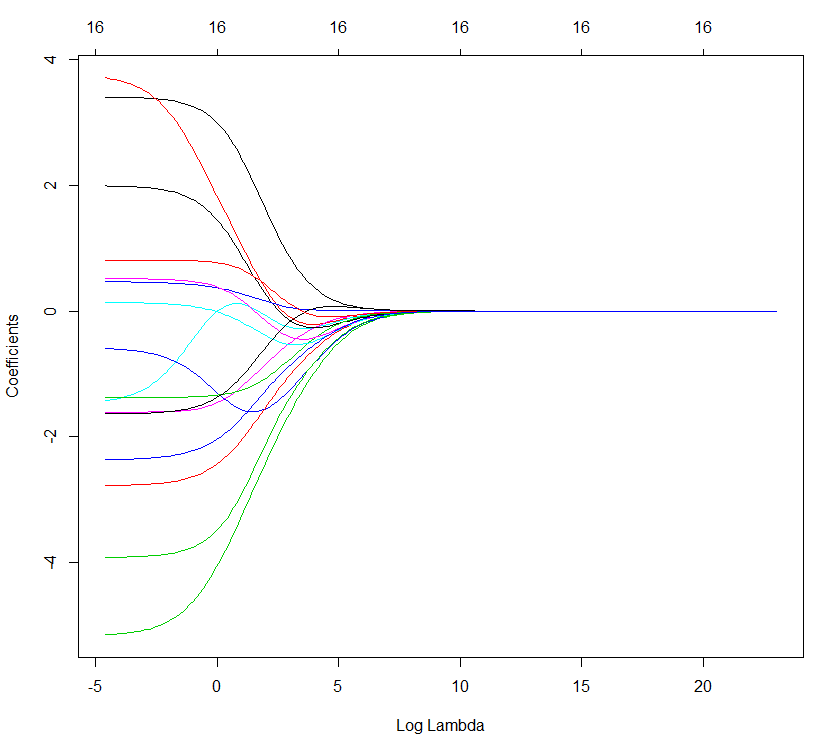
lasso.mod = glmnet(Xs,y,alpha=1,lambda=grid)

plot(ridge.mod) 🡨 trace plots with norm constraint on the x-axis

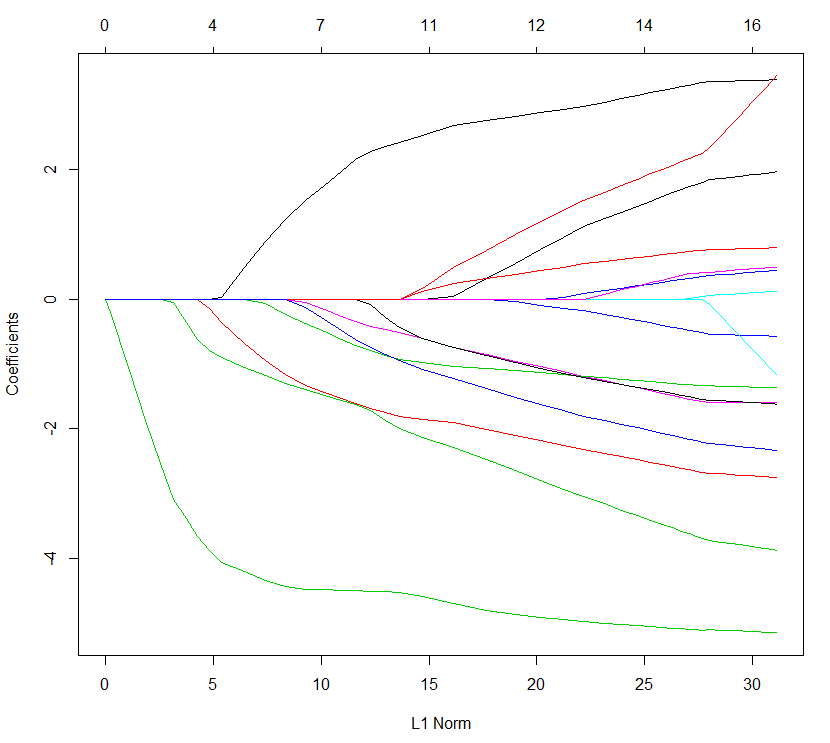
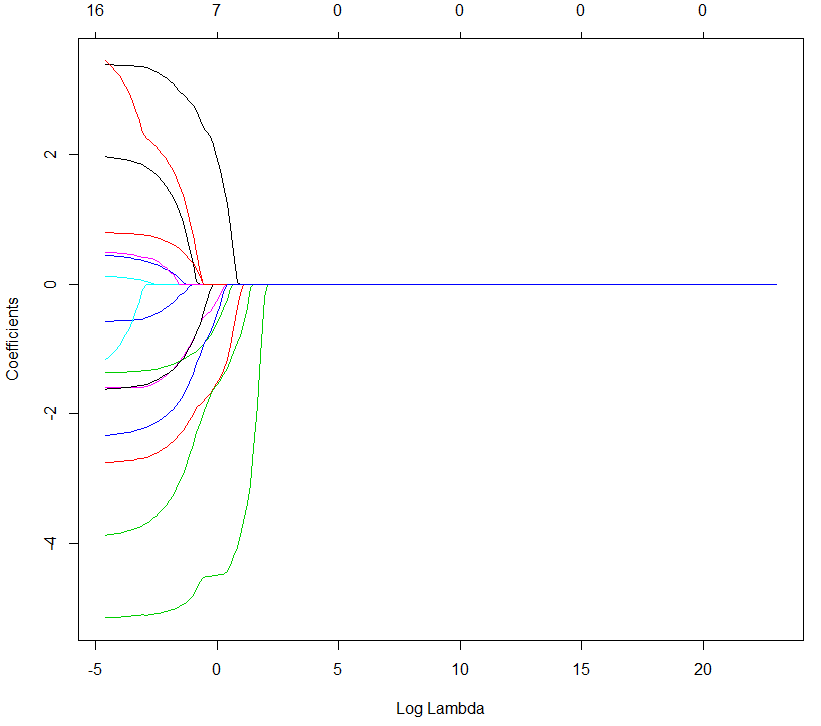
plot(ridge.mod,xvar=”lambda”) 🡨 trace plots with log lambda on the x-axis

Similarly for plotting the coefficient shrinkage for the LASSO model.

Ridge

Lasso

When looking in the graphs you can see in both set of graphs that the lasso models parameters shrink earlier than the Ridge model. That could be due how lasso’s parameters do hit zero and ridge’s parameters comes close to hitting zero.

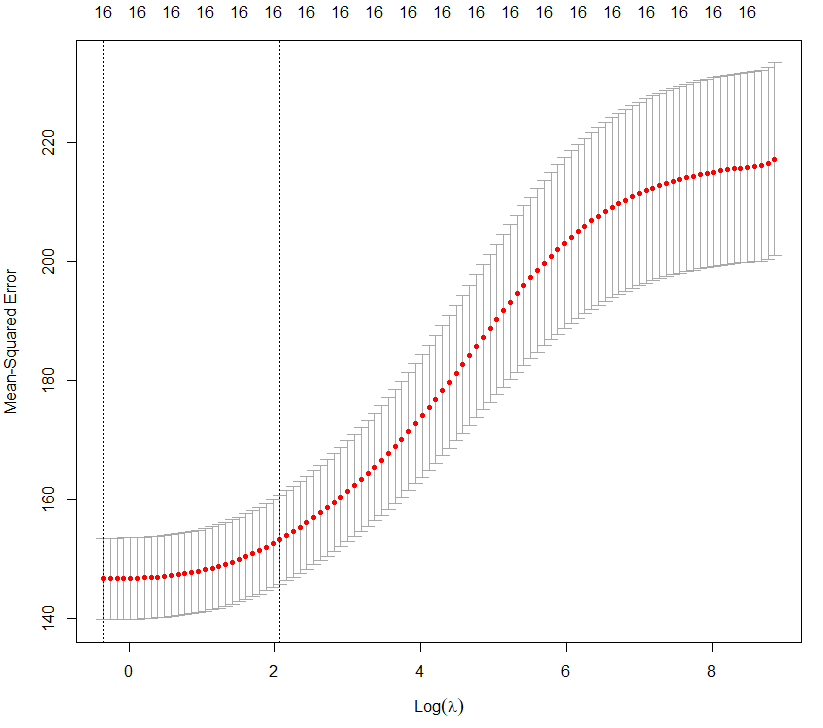
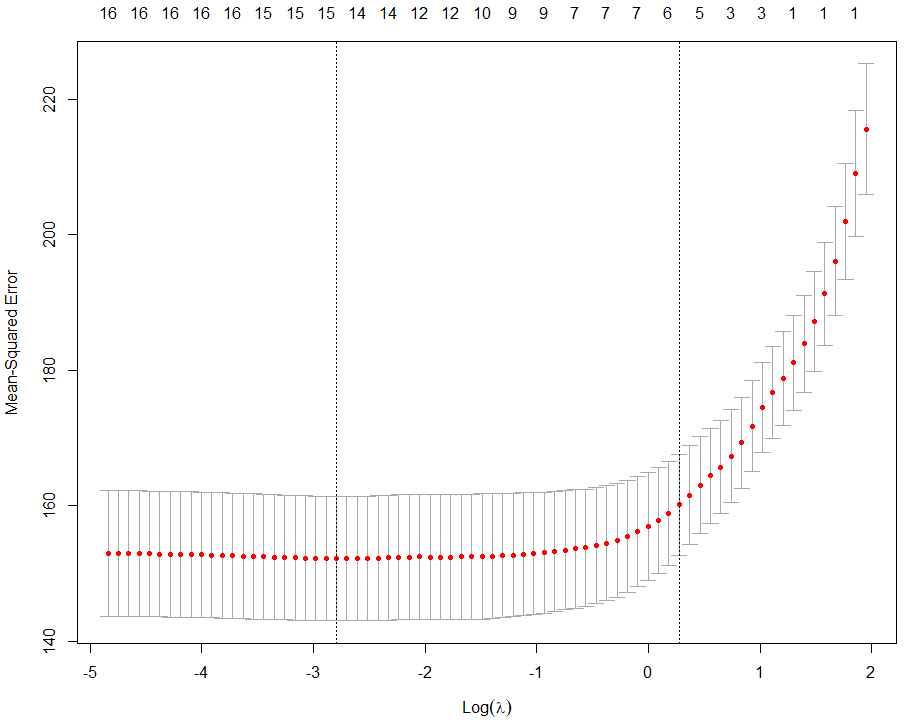
1. Use cross-validation to determine the “optimal” values for the shrinkage parameter for both ridge and lasso and plot the results. The commands for ridge regression are given below. (4 pts.)

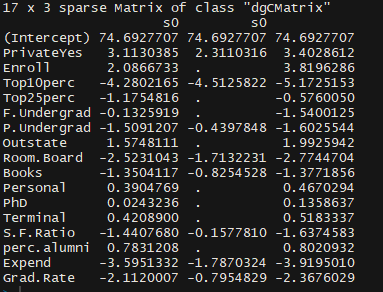
cv.out = cv.glmnet(X,y,alpha=0)

plot(cv.out)

bestlam = cv.out$lambda.min

Ridge Lasso

1. Using the optimal lambda (bestlam) for both ridge and Lasso regression fit both models and compare the estimated coefficients for the OLS, ridge, and Lasso regressions. Discuss. (3 pts.)  
   

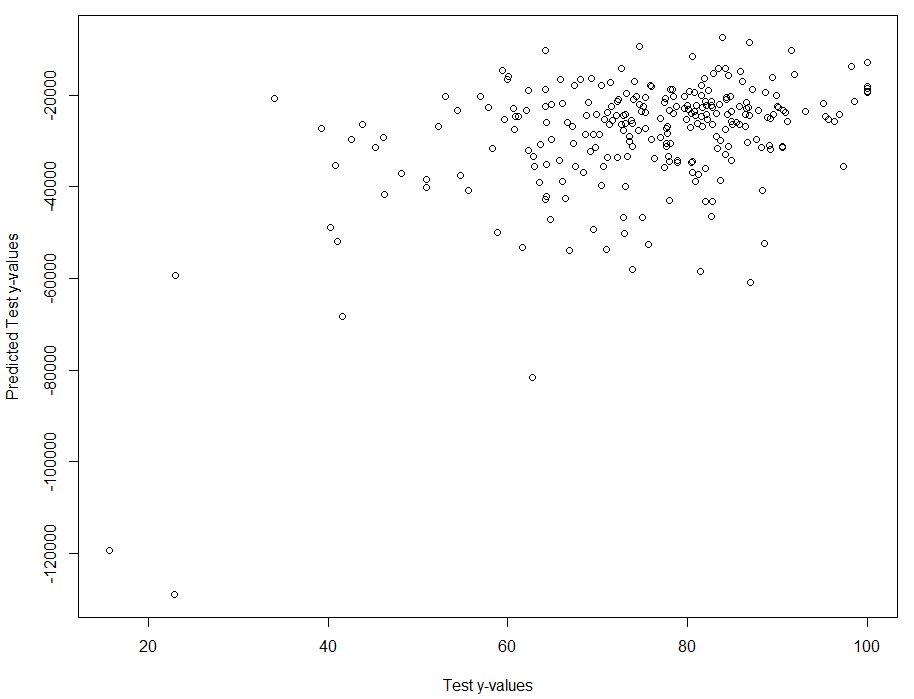
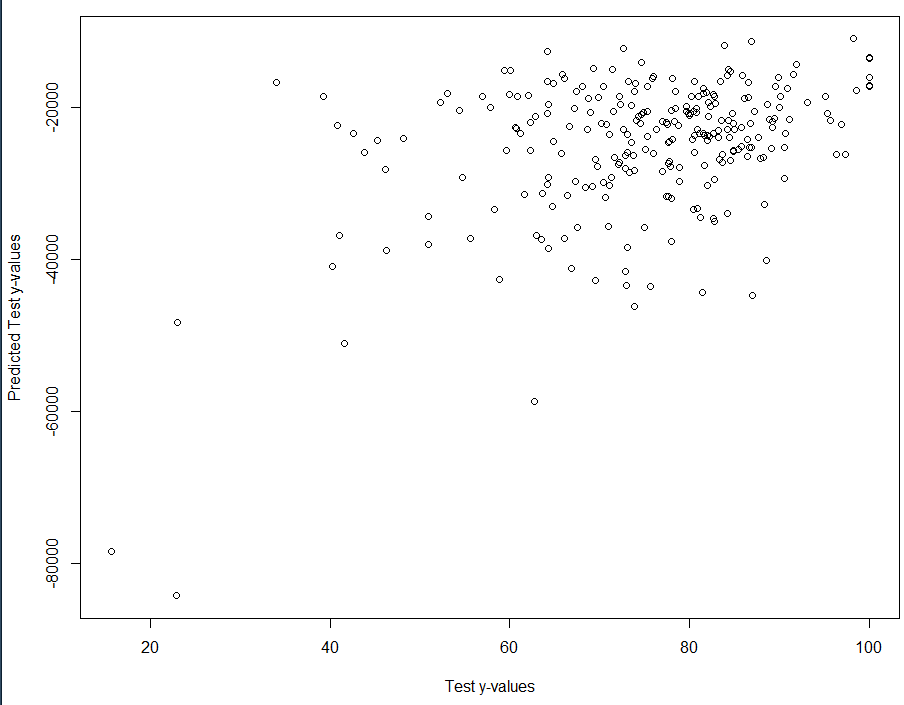
When looking at the coefficients a few interesting things is how there are different variable not used in the Lasso model but look to have a good amount of important in the other two models.

There is one variable that has the most effect in all three models and it’s called Top10perc.

1. Construct a plot of the predicted test y values vs. the actual y test values for both the ridge and Lasso regression models. Discuss. (4 pts.)

plot(y[test],predict(***modelname***,newx=X[test,]),xlab=”Test y-values”,ylab=”Predicted Test y-values”)

Ridge Lasso

Based on the looks on the graphs

1. Using the optimal lambda (bestlam) for both ridge and Lasso regression find the mean RMSEP for the **test** set. How do the mean RMSEP compare for the OLS, ridge, and Lasso regression models? Discuss. (3 pts.)

OLS = 15.83466 Lasso = 11.67995 Ridge = 11.60358

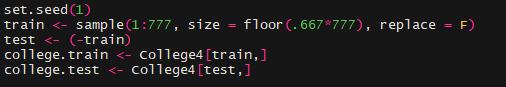
1. Use Monte Carlo Split-Sample Cross-Validation to estimate the mean RMSEP for the OLS, Ridge, and Lasso regressions above. Which model has best predictive performance? (5 pts.)

OLS = 16.31574 Ridge = 12.15552 Lasso = 12.39634

The model that did the best in predictive performance was the Ridge model.

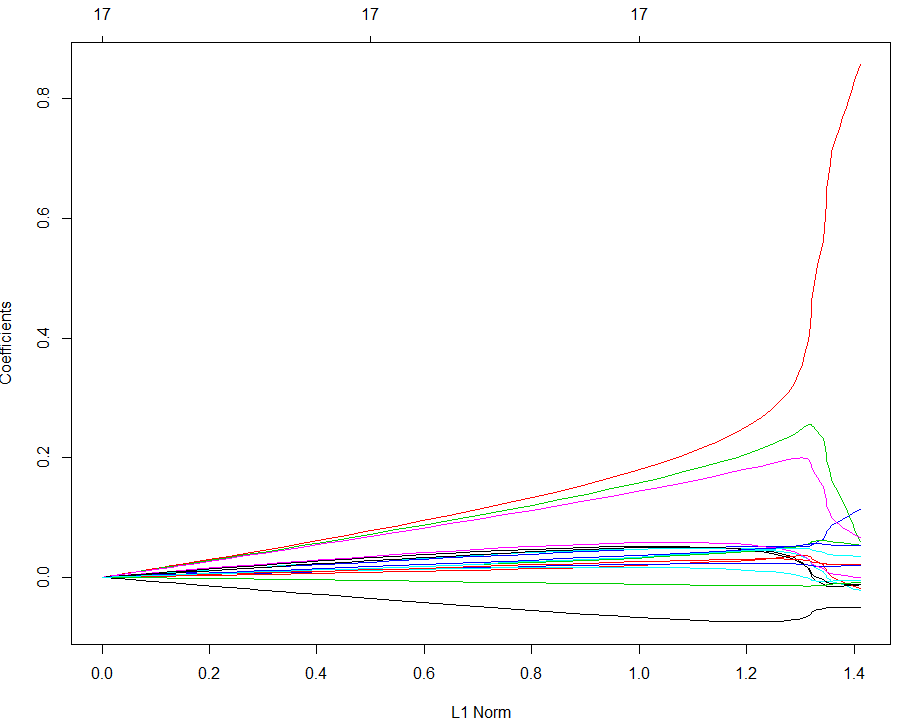
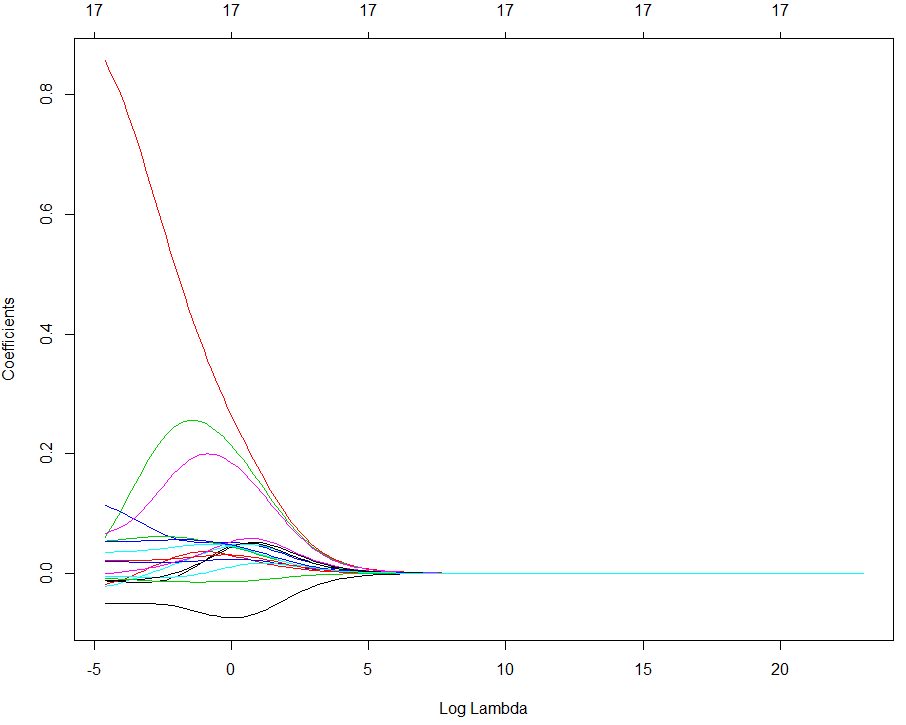
1. Repeat (a) – (h) using the College4 data frame with logApps as the response. (30 pts.)

a)

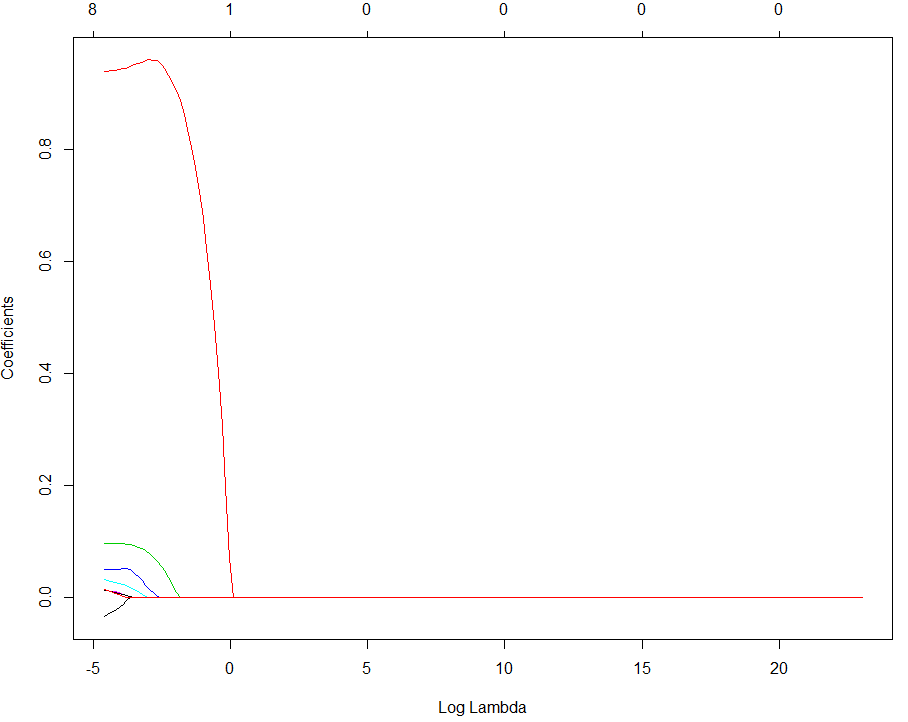
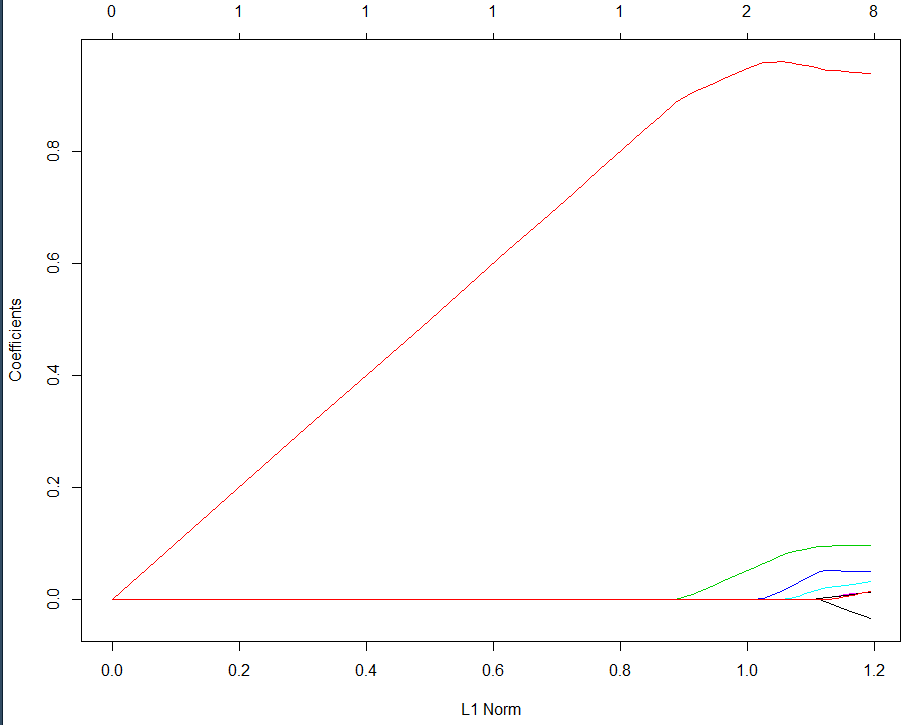


**b)** RMSEP = 1.489735

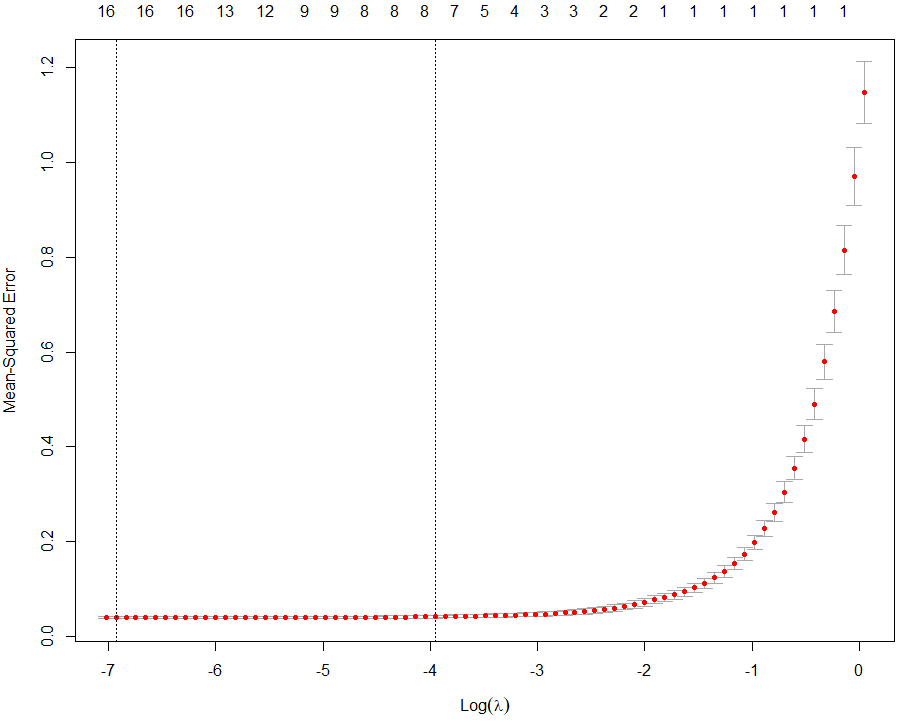
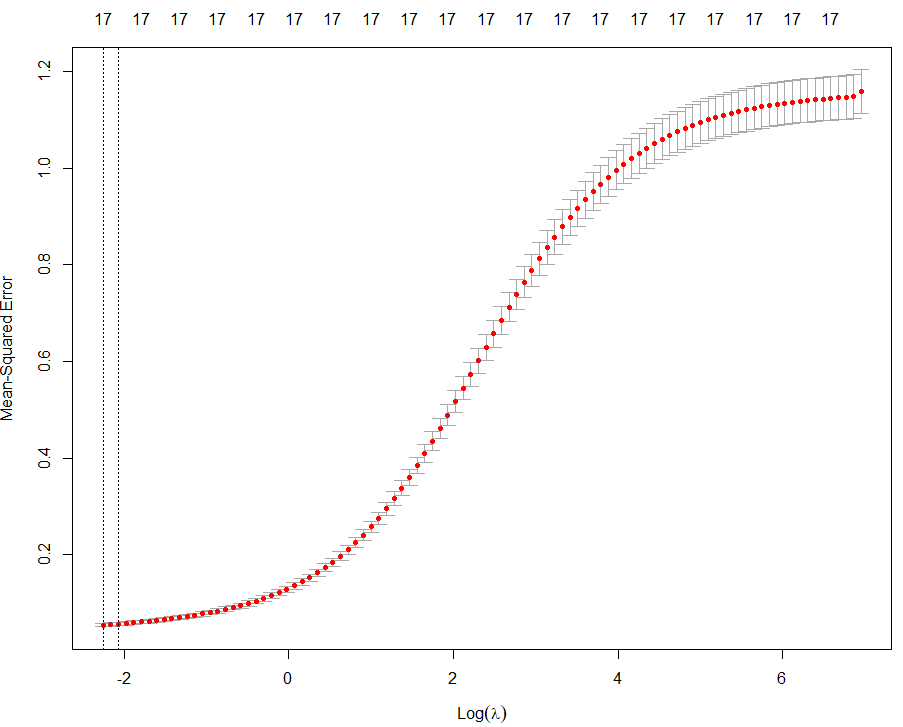
c) Ridge

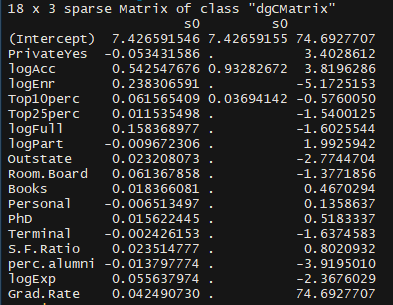
**Lasso**

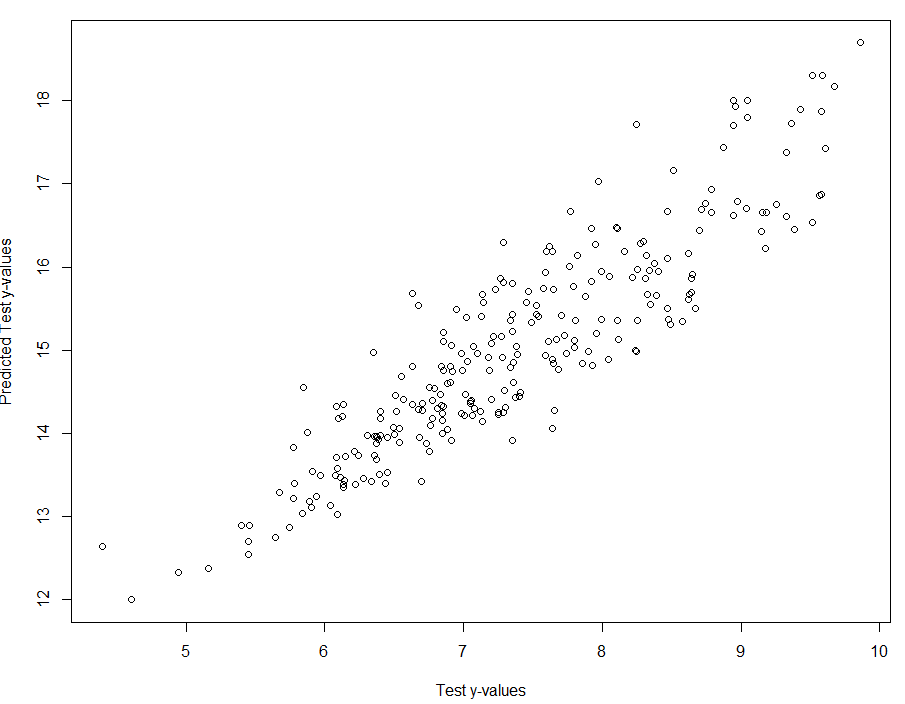
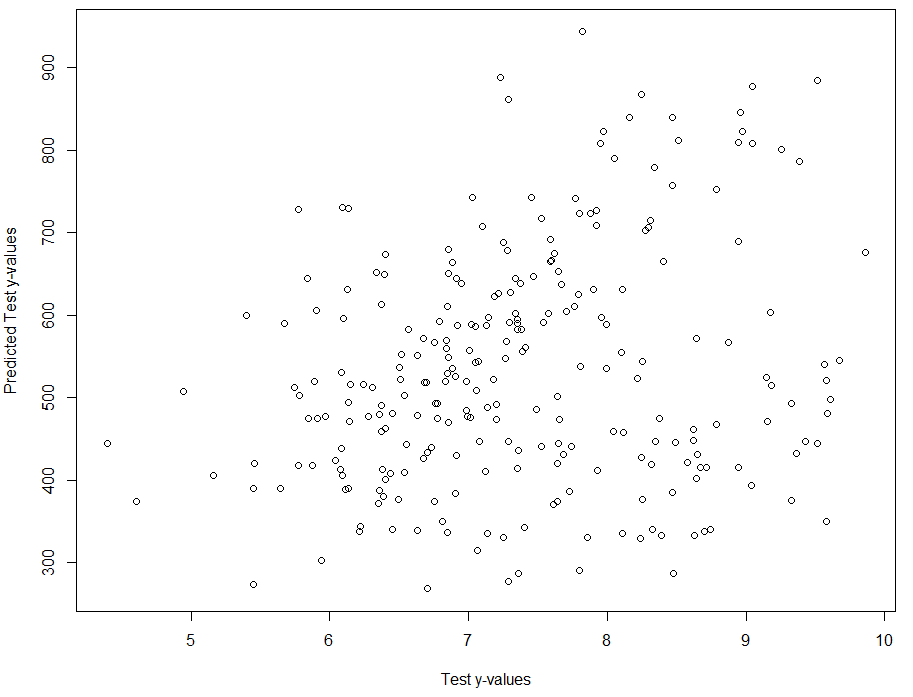


**d) Ridge Lasso**



**e)**



**f) Ridge Lasso** 

**g) OLS = 1.489735 Lasso = 1.31942 Ridge = 1.353942**

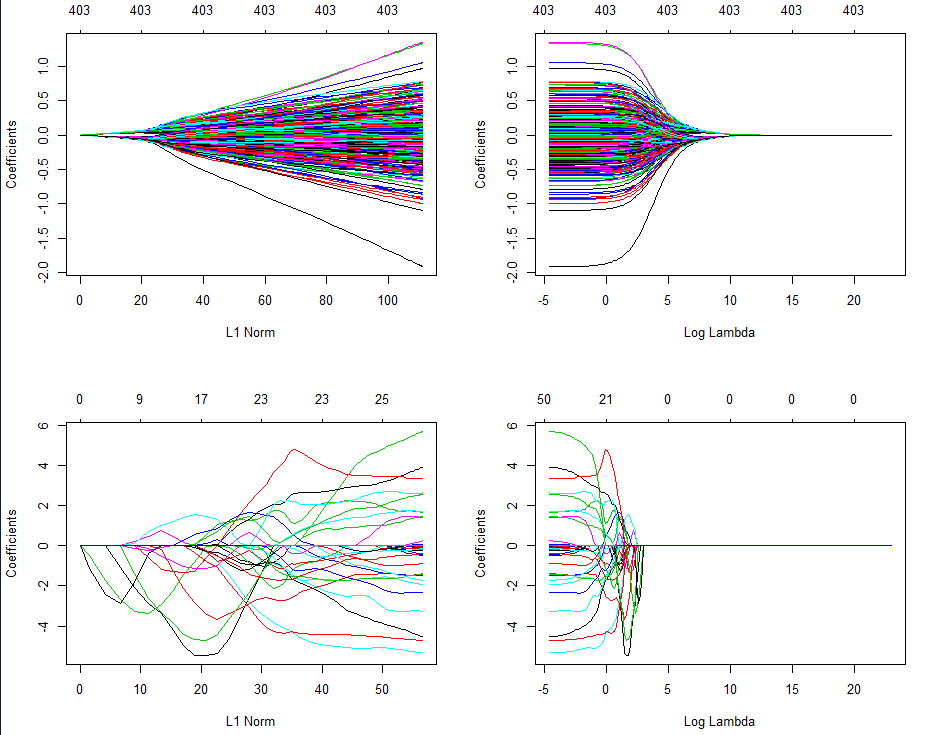
**h) OLS = 1.538292 Lasso = .2456127 Ridge = .233585**

The model that did the best in predictive performance was the Ridge model.

**PROBLEM 2 – PREDICTING AGE USING GENETIC INFORMATION**

Using the data frame Lu2004 contained in Lu2004.RData file I e-mailed you or read in the **Lu2004.csv** file from the website. The response Age is the first column of the data frame, the remaining 403 columns are genetic marker intensity measurements for 403 different genes. Use ridge and Lasso regression to develop optimal models for predicting Age of the subject. Do not worry about a training/test set approach as there are only n = 30 subjects in the full data set. This is an example of a *wide data* problem because *n << p* (because *30 << 403)*! I have also e-mailed the research paper by Lu et al. (2004) that they published based on their analysis of these data.

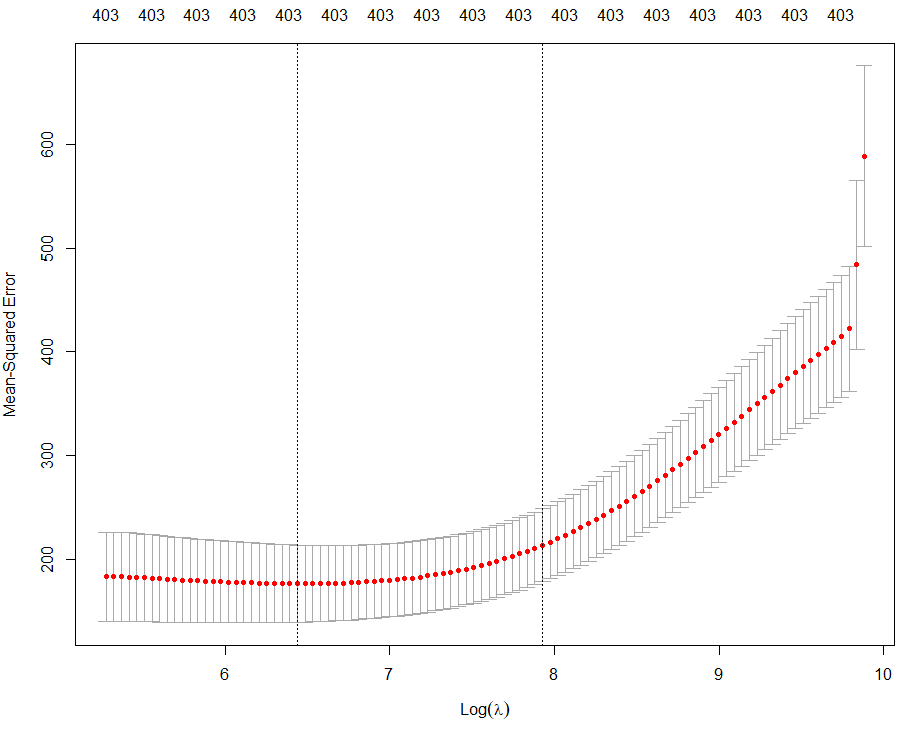
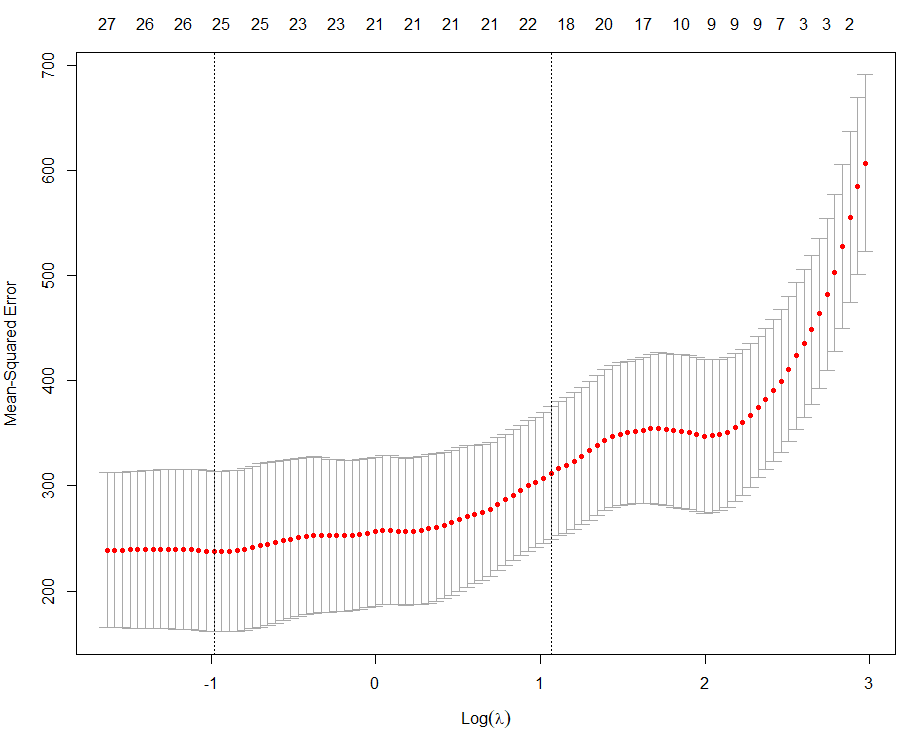
1. Generate a sequence of ridge and Lasso regression models using the same grid values used in Problem 1. Create two plots showing the coefficient shrinkage with different x-axes for both ridge and Lasso regressions as in part (c) for Problem 1. Briefly discuss these plots. (6 pts.)



Lasso

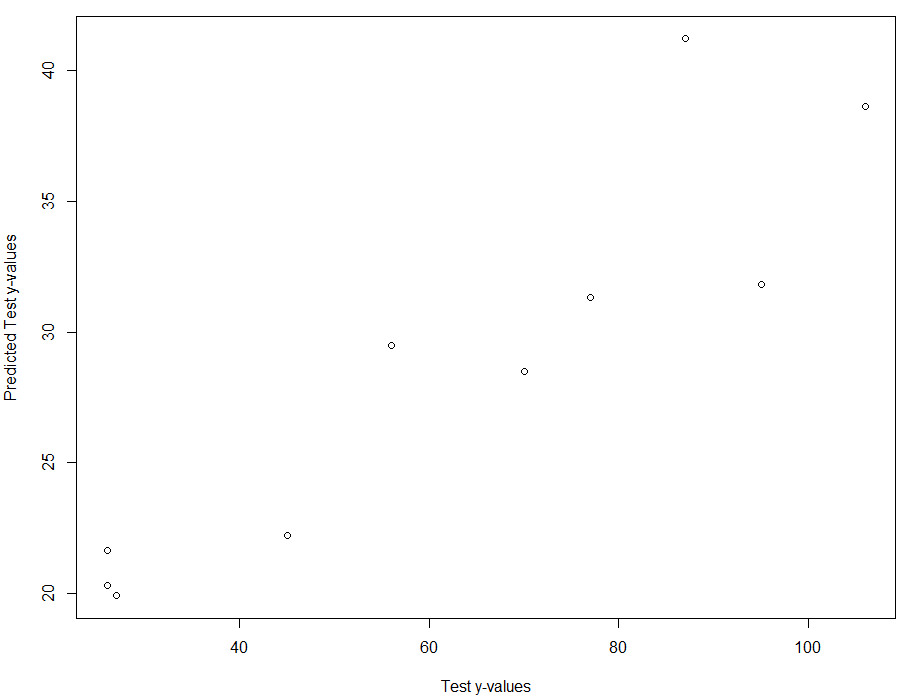
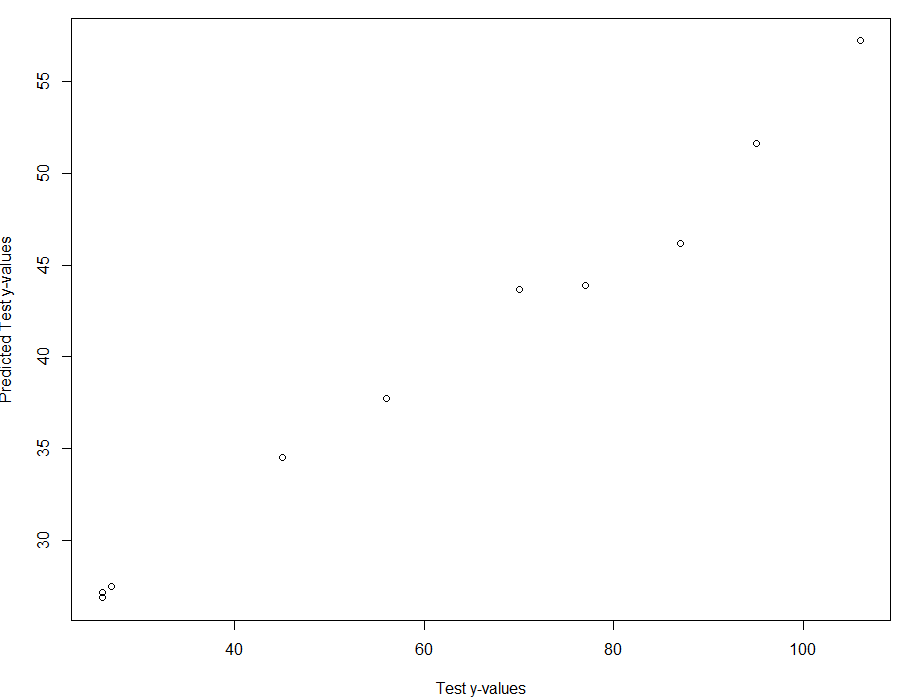
Ridge

1. Find the optimal  for ridge and Lasso regression using the cv.glmnet function. Also show plots of the cross-validation results for both methods. Discuss. (4 pts.)

Ridge Lasso  
 

1. Fit the optimal ridge and Lasso regression models and construct plots of the predicted ages vs. actual age . Also find the correlation between and the correlation squared. Note the correlation between squared is the R-square for the model. Which model predicts subject age better? (5 pts.)

Ridge Lasso

The model that predicted subject age better is Lasso.

1. Using the better of the two models as determined from part (c), examine and interpret the estimated coefficients. If the researchers ask you “*which genes are most related or useful in determining the age of the subject?*”, what would you tell them, i.e. give a list of specific genes to answer this question. (5 pts.)

The most useful genes are x1585\_at, x31771\_at, x32892\_at, x33507\_g\_at, x35569\_at, x37812\_at, x38474\_at, x39287\_at, and x841\_at.

1. Use Monte Carlo cross-validation estimate the prediction accuracies for both ridge and Lasso regression for these data. (Use *p = .75* and *B = 1000*.) (5 pts.)

Ridge = RMSEP = 14.98274 MAEP= 12.56791 MAPEP= 0.2508979

Lasso = RMSEP = 15.2809 MAEP= 12.84116 MAPEP= 0.2396507

1. **BONUS:** Fit an Elastic Net to these data, fine tune it, and compare the predictive performance to ridge and LASSO. (10 pts.)

**Alpha = .1**

RMSEP = 14.18095 MAEP= 11.84796 MAPEP= 0.2228254

The Elastic Net did better than the Lasso and Ridge models by a percent or 2.