April 11, 2018

Assignment 4

Sanchit Singhal

INF 397 – Statistical Analysis and Learning w/ Prof. Varun Rai

Spring 2018

The University of Texas at Austin

Question 1

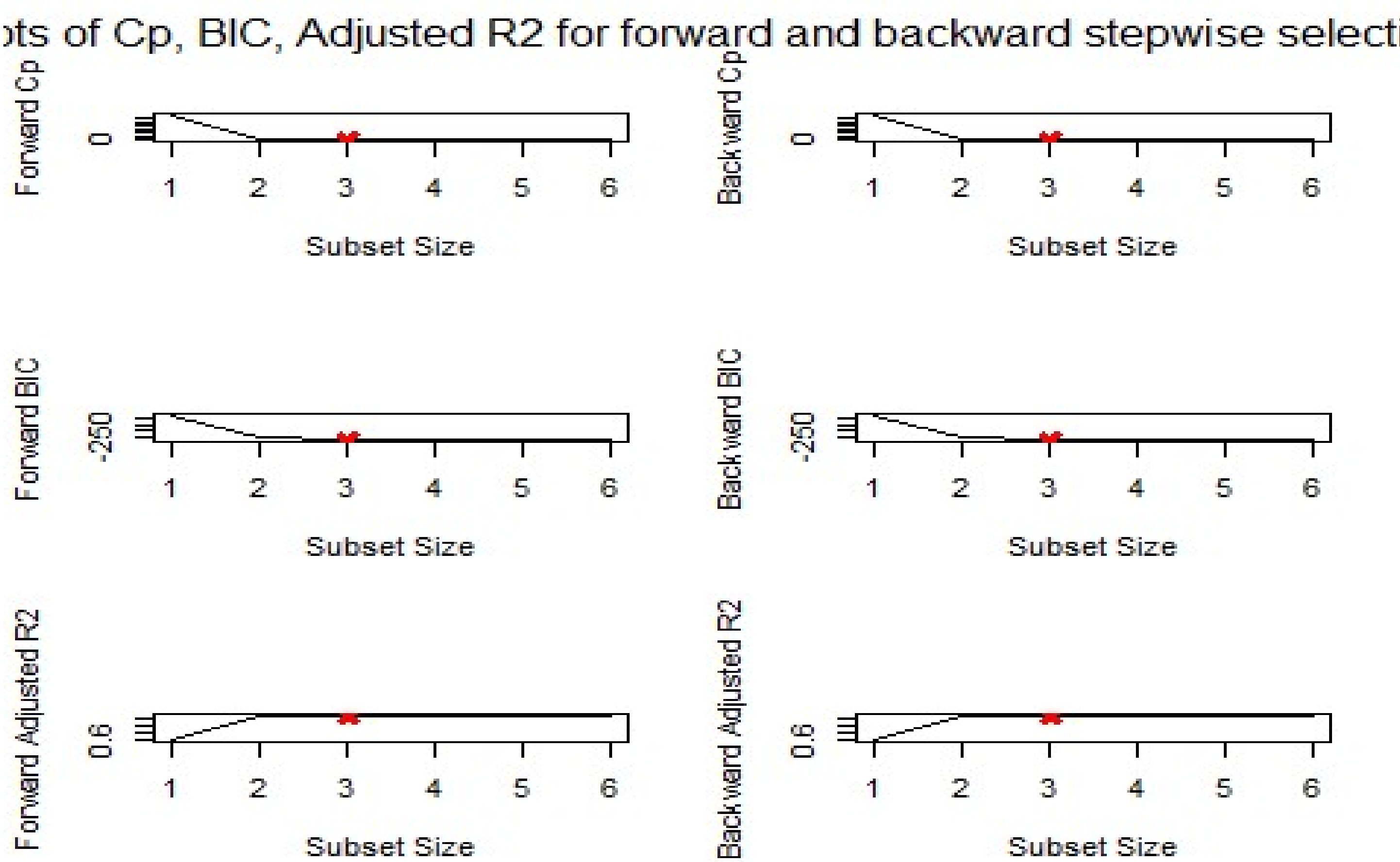
<u>Summary of subsection "Another Formulation for Ridge Regression and the Lasso"</u>

Ridge Regression and the Lasso can be interpreted as more practical alternatives to the best subset selection method. Since in the best subset selection, all models containing s predictors are built, the problem can be rephrased to say: find the set of coefficients that minimize RSS subject to the constrain that no greater than s predictors are non-zero. This wording of the best subset selection is similar to both shrinkage methods in which we are solving a similar problem but with different constraints. In the lasso, we try to find the set of coefficients that minimize RSS given that there is an upper limit to the absolute value of the coefficients, s. Depending on the value of s, the coefficients are selected (large s values will yield the least squares model, whereas small s values lead to smaller coefficients). Similarly, ridge regression works the same way, under a budget, but uses the square of the coefficients rather than the absolute values. In both cases, the shrinkage methods solve the best subset selection problem but under constraints more feasibility – and therefore can be thought to a more practical alternative to the best subset selection method. It should be noted that the lasso is more closely related to the subset selection method however because it allows for variable selection and yields sparse results.

```
### Question 2

# a)

set.seed(1)

# Generate predictor(X), noise(eps)

X = rnorm(100)
eps = rnorm(100)

# Generate coefficients(beta), response(Y)

beta0 = 3
beta1 = 2
beta2 = -3
beta3 = 0.3

Y = beta0 + beta1 * X + beta2 * X^2 + beta3 * X^3 + eps

## b)

# Build forward and backward selection models

library(leaps)

## Warning: package 'leaps' was built under R version 3.4.4

data <- data.frame(y = Y, x = X)

fwd_selection.model <- regsubsets(y ~ x + I(x^2) + I(x^3) + I(x^4) + I(x^5) + I(x^6), dat
a = data, nvmax = 10, method = "forward")
bwd_selection.model <- regsubsets(y ~ x + I(x^2) + I(x^3) + I(x^4) + I(x^5) + I(x^6), dat
a = data, nvmax = 10, method = "backward")
```

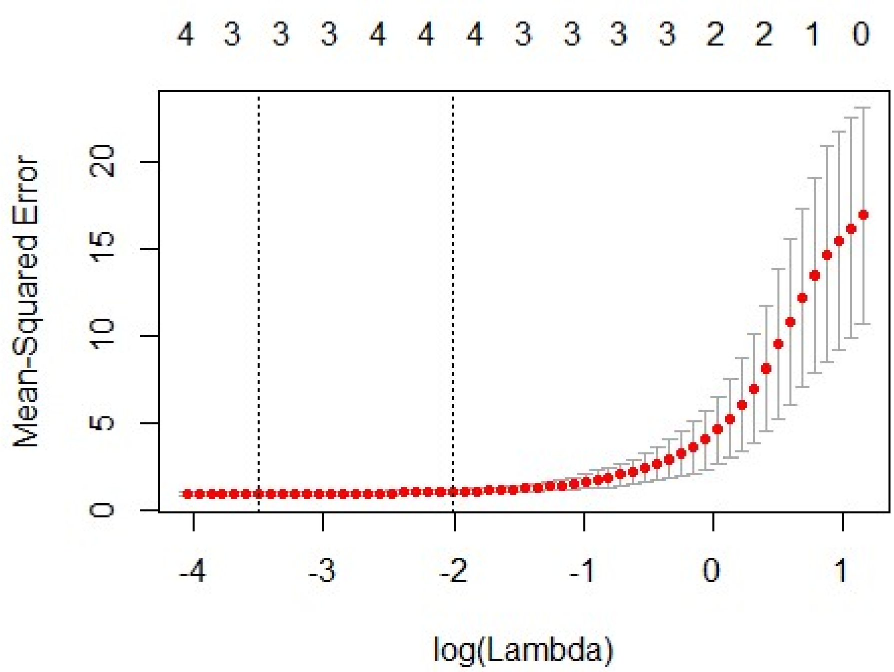
```r
# Analysis of results

fwd_selection.summary = summary(fwd_selection.model)
bwd_selection.summary = summary(bwd_selection.model)

par(mfrow = c(3, 2))

plot(fwd_selection.summary$cp, xlab = "Subset Size", ylab = "Forward Cp", pch = 20, type
= "l")
points(which.min(fwd_selection.summary$cp), fwd_selection.summary$cp[which.min(fwd_select
ion.summary$cp)], pch = 4, col = "red", lwd = 3)

plot(bwd_selection.summary$cp, xlab = "Subset Size", ylab = "Backward Cp", pch = 20, type
= "l")
points(which.min(bwd_selection.summary$cp), bwd_selection.summary$cp[which.min(bwd_select
ion.summary$cp)], pch = 4, col = "red", lwd = 3)

plot(fwd_selection.summary$bic, xlab = "Subset Size", ylab = "Forward BIC", pch = 20, typ
e = "l")
points(which.min(fwd_selection.summary$bic), fwd_selection.summary$bic[which.min(fwd_sele
ction.summary$bic)], pch = 4, col = "red", lwd = 3)

plot(bwd_selection.summary$bic, xlab = "Subset Size", ylab = "Backward BIC", pch = 20, ty
pe = "l")
points(which.min(bwd_selection.summary$bic), bwd_selection.summary$bic[which.min(bwd_sele
ction.summary$bic)], pch = 4, col = "red", lwd = 3)

plot(fwd_selection.summary$adjr2, xlab = "Subset Size", ylab = "Forward Adjusted R2", pch
= 20, type = "l")
points(which.max(fwd_selection.summary$adjr2), fwd_selection.summary$adjr2[which.max(fwd_
selection.summary$adjr2)], pch = 4, col = "red", lwd = 3)

plot(bwd_selection.summary$adjr2, xlab = "Subset Size", ylab = "Backward Adjusted R2", pc
h = 20, type = "l")
points(which.max(bwd_selection.summary$adjr2), bwd_selection.summary$adjr2[which.max(bwd_
selection.summary$adjr2)], pch = 4, col = "red", lwd = 3)

mtext("Plots of Cp, BIC, Adjusted R2 for forward and backward stepwise selection", side =
3, line = -2, outer = TRUE)

# Determine coefficients of models

fwd_selection.coef <- coefficients(fwd_selection.model, id=3)
bwd_selection.coef <- coefficients(bwd_selection.model, id=3)

print(fwd_selection.coef)

## (Intercept)           x      I(x^2)      I(x^5)
##  3.07154970  2.24612357 -3.14807309  0.05744276

print(bwd_selection.coef)

## (Intercept)           x      I(x^2)      I(x^5)
##  3.07154970  2.24612357 -3.14807309  0.05744276
```

```
# Comment on Results

# For both forward and backward stepwise selection, all three estimates of test
# error (Cp, BIC, Adjusted R2) selected a 3 predictor model. Both approaches
# also selected the same predictors : X, X^2, & X^5.

# c)

library(glmnet)

## Warning: package 'glmnet' was built under R version 3.4.4

## Loading required package: Matrix

## Warning: package 'Matrix' was built under R version 3.4.3

## Loading required package: foreach

## Warning: package 'foreach' was built under R version 3.4.4

## Loaded glmnet 2.0-16
```




```
# Build Lasso model

xmatrix <- model.matrix(y ~ x + I(x^2) + I(x^3) + I(x^4) + I(x^5) + I(x^6), data = data)[
, -1]
lasso <- cv.glmnet(xmatrix, Y, alpha = 1)
```

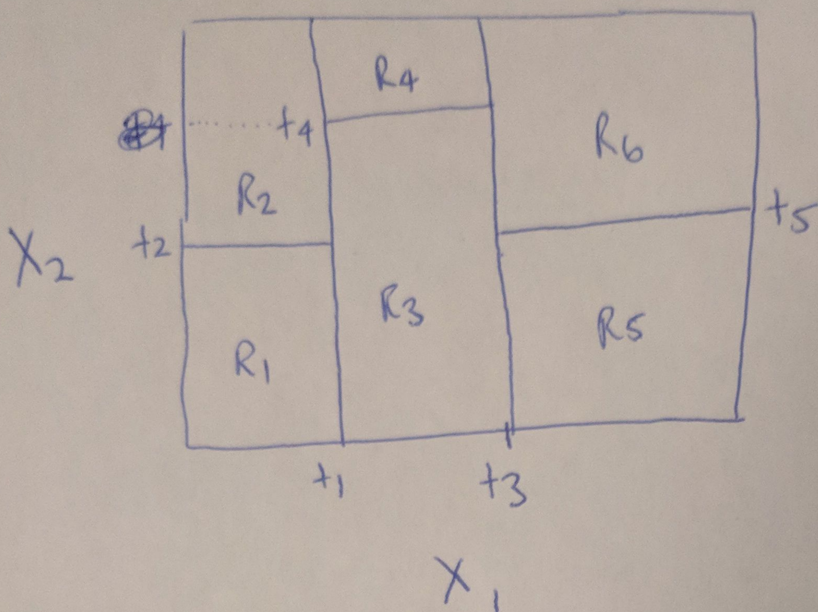
```r
par(mfrow = c(1, 1))
plot(lasso)
```




```r
# Determine Optimal value of Lambda

bestlambda <- lasso$lambda.min
print(bestlambda)

## [1] 0.0301936

# Fit Lasso model and get coefficients

lasso.fit <- glmnet(xmatrix, Y, alpha = 1)
predict(lasso.fit, s = bestlambda, type = "coefficients")[1:7, ]

## (Intercept)           x      I(x^2)      I(x^3)      I(x^4)      I(x^5)
##  3.04683073  2.22998261 -3.11276660  0.00000000  0.00000000  0.05482882
##      I(x^6)
##  0.00000000

# Comment on Results

# Lasso selected a 3 predictor model as well - using the same predictors : X,
# X^2, X^5. The rest of the coefficients are zero. Although the values for the
# coefficients obtained from the lasso are similar in magnitude to the values
# selected through the forward and backward stepwise selection, they are not
# exactly the same.
```

Question 3

2D Feature Space



Decision Tree