3. Suppose we estimate the regression coefficients in a linear regression model by minimizing

$$\sum_{i=1}^{n}\left(y_i - \beta_0 - \sum_{j=1}^{p}\beta_j x_{ij}\right)^2 \quad \text{subject to} \quad \sum_{j=1}^{p}|\beta_j| \leq s$$

for a particular value of s. For parts (a) through (e), indicate which of i. through v. is correct. Justify your answer.

(a) As we increase s from 0, the training RSS will:

i. Increase initially, and then eventually start decreasing in an inverted U shape.

ii. Decrease initially, and then eventually start increasing in a U shape.

iii. Steadily increase.

iv. Steadily decrease.

v. Remain constant.

(b) Repeat (a) for test RSS.

(c) Repeat (a) for variance.

(d) Repeat (a) for (squared) bias.

(e) Repeat (a) for the irreducible error.

(a) iv. When s is small, the restriction on the parameters is strong and the model cannot fit the training data well. When s increases, the restriction on the parameters becomes smaller, and the parameters become more and more close to the least square estimates. As a result, the RSS decreases steadily to the least square RSS.

(b) ii. When s is small, the model cannot fit the training data well, so both the train and test RSS are large. As s increases, the model can fit better, so the train and test RSS will decrease. Yet when s is too large, the model will overfit the training data and produce high variance prediction, so the test RSS will increase again.

(c) iii. When s is very small, the model is very simple and underfit, producing high bias and low variance estimates (predict a constant when s=0) . When s is very large, the model has little restriction and can be very complicated and overfit. So, the variance will increase steadily.

(d) iv. When s is very small, the model is very simple and underfit, producing high bias and low variance estimates. When s is very large, the model has little restriction and can be very complicated and overfit. So, the bias will decrease steadily.

(e) v. Irreducible error cannot be affected by the models we use.

4. Suppose we estimate the regression coefficients in a linear regression model by minimizing

$$\sum_{i=1}^{n}\left(y_i - \beta_0 - \sum_{j=1}^{p}\beta_j x_{ij}\right)^2 + \lambda\sum_{j=1}^{p}\beta_j^2$$

for a particular value of λ. For parts (a) through (e), indicate which of i. through v. is correct. Justify your answer.

(a) As we increase λ from 0, the training RSS will:

i. Increase initially, and then eventually start decreasing in an inverted U shape.

ii. Decrease initially, and then eventually start increasing in a U shape.

iii. Steadily increase.

iv. Steadily decrease.

v. Remain constant.

(b) Repeat (a) for test RSS.

(c) Repeat (a) for variance.

(d) Repeat (a) for (squared) bias.

(e) Repeat (a) for the irreducible error.

(a) iii. When λ is 0, the model has no restrictions on the parameters and the training RSS is the least square RSS, which is the smallest RSS for regular linear regression. As lambda increases, parameters are forced to be close to zero, and the training RSS goes up.

(b) ii. When λ is small, the model can overfit the training data, so the test RSS is large. As λ increases, the model has more restriction and cannot fit well on the training data, but the test RSS will decrease because the model is robust. Yet when λ is too large, the model will underfit the training data, so the test RSS will increase again.

(c) iv. When λ is very small, the model tends to overfit, producing low bias and high variance estimates. When λ is very large, the model is forced to be simple (constant) and underfit. So, the variance will decrease steadily

(d) iii. When λ is very small, the model tends to overfit, producing low bias and high variance estimates. When λ is very large, the model is forced to be simple (constant) and underfit. So, the bias will increase steadily.

(e) v. Irreducible error cannot be affected by the models we use.

|

5. It is well-known that ridge regression tends to give similar coefficient values to correlated variables, whereas the lasso may give quite different coefficient values to correlated variables. We will now explore this property in a very simple setting.

   Suppose that $n = 2$, $p = 2$, $x_{11} = x_{12}$, $x_{21} = x_{22}$. Furthermore, suppose that $y_1 + y_2 = 0$ and $x_{11} + x_{21} = 0$ and $x_{12} + x_{22} = 0$, so that the estimate for the intercept in a least squares, ridge regression, or lasso model is zero: $\hat{\beta}_0 = 0$.

   (a) Write out the ridge regression optimization problem in this setting.

   (b) Argue that in this setting, the ridge coefficient estimates satisfy $\hat{\beta}_1 = \hat{\beta}_2$.

   (c) Write out the lasso optimization problem in this setting.

   (d) Argue that in this setting, the lasso coefficients $\hat{\beta}_1$ and $\hat{\beta}_2$ are not unique—in other words, there are many possible solutions to the optimization problem in (c). Describe these solutions.

a) $\min \left(y_1 - (\beta_0 + \beta_1 x_{11} + \beta_2 x_{12})\right)^2 + \left(y_2 - (\beta_0 + \beta_1 x_{21} + \beta_2 x_{22})\right)^2 + \lambda(\beta_1^2 + \beta_2^2)$

$\because \hat{\beta}_0 = 0.$

$\therefore \min (y_1 - \beta_1 x_{11} - \beta_2 x_{12})^2 + (y_2 - \beta_1 x_{21} - \beta_2 x_{22})^2 + \lambda(\beta_1^2 + \beta_2^2)$ (i)

b) $(i) = (y_1 - (\beta_1 + \beta_2)x_{11})^2 + (y_2 + (\beta_1 + \beta_2)x_{11})^2 + \lambda(\beta_1^2 + \beta_2^2)$

$\hat{=} L = 2(\beta_1 + \beta_2)^2 x_{11}^2 + \lambda(\beta_1^2 + \beta_2^2) + 2(\beta_1 + \beta_2)x_{11}(y_2 - y_1) + y_1^2 + y_2^2$

Set $\begin{cases} \dfrac{\partial L}{\partial \beta_1} = 0 \\ \dfrac{\partial L}{\partial \beta_2} = 0 \end{cases} \Rightarrow \begin{cases} \hat{\beta}_1 = \dfrac{x_{11}(y_1 - y_2) - 2x_{11}^2 \hat{\beta}_2}{2x_{11}^2 + \lambda} \\ \hat{\beta}_2 = \dfrac{x_{11}(y_1 - y_2) - 2x_{11}^2 \hat{\beta}_1}{2x_{11}^2 + \lambda} \end{cases}$

$\because$ $\hat{\beta}_1$ & $\hat{\beta}_2$ are symmetry in the solution expression.

$\therefore \hat{\beta}_1 = \hat{\beta}_2$

c) $\min (y_1 - \beta_1 x_{11} - \beta_2 x_{12})^2 + (y_1 - \beta_1 x_{21} - \beta_2 x_{22})^2 + \lambda(|\beta_1| + |\beta_2|)$

d). $\min (y_1 - \beta_1 x_{11} - \beta_2 x_{12})^2 + (y_2 - \beta_1 x_{21} - \beta_2 x_{22})^2$ (ii) with $|\beta_1| + |\beta_2| \le s$

$\because x_{11} = x_{12},\ x_{21} = x_{22},\ x_{11} + x_{21} = 0\ \ x_{12} + x_{22} = 0\ \ y_1 + y_2 = 0.$

$\therefore$ (ii) is equal to $\min 2(y_1 - (\beta_1 + \beta_2)x_{11})^2$ with $|\beta_1| + |\beta_2| \le s$

$\therefore$ given y_1 and x_{11}, $(y_1 - (\beta_1 + \beta_2)x_{11})^2$ is minimized to 0 when $\beta_1 + \beta_2 = \dfrac{y_1}{x_{11}}$ $(x_{11} \neq 0)$

$\because \beta_1 + \beta_2 = \dfrac{y_1}{x_1}$ and $\beta_1 + \beta_2 = s$ are parallel

$\therefore$ if $0 < \dfrac{y_1}{x_1} < s$, then all β_1, β_2 s.t. $\beta_1 + \beta_2 = \dfrac{y_1}{x_1}$, $\beta_1 > 0$ $\beta_2 > 0$ are the optimal solutions.

6. We will now explore (6.12) and (6.13) further.

(a) Consider (6.12) with $p = 1$. For some choice of y_1 and $\lambda > 0$, plot (6.12) as a function of β_1. Your plot should confirm that (6.12) is solved by (6.14).

(b) Consider (6.13) with $p = 1$. For some choice of y_1 and $\lambda > 0$, plot (6.13) as a function of β_1. Your plot should confirm that (6.13) is solved by (6.15).

$$\sum_{j=1}^{p}(y_j - \beta_j)^2 + \lambda \sum_{j=1}^{p} \beta_j^2 \tag{6.12}$$

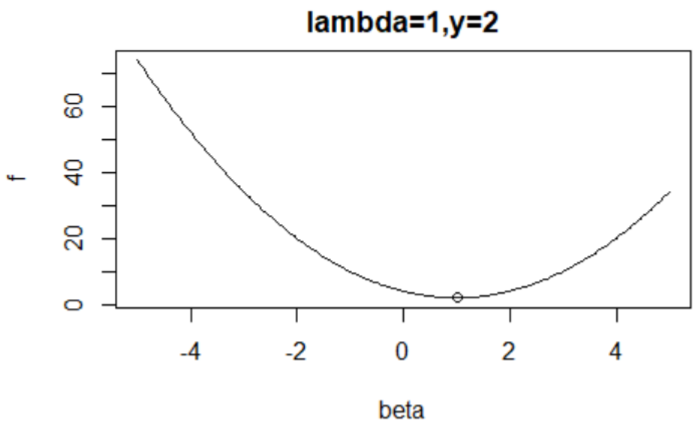
$$\sum_{j=1}^{p}(y_j - \beta_j)^2 + \lambda \sum_{j=1}^{p} |\beta_j| \tag{6.13}$$

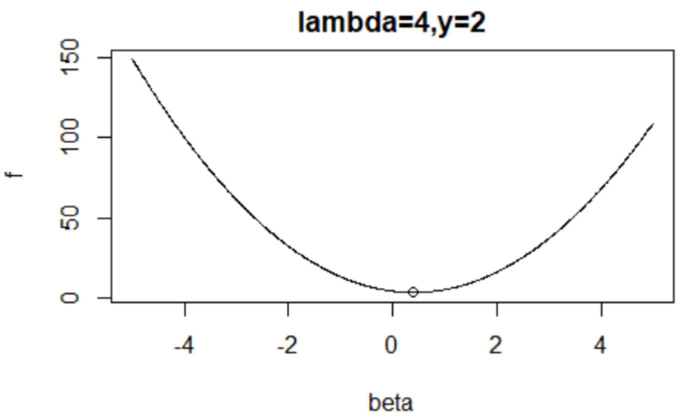
$$\hat{\beta}_j^R = y_j/(1+\lambda), \tag{6.14}$$

$$\hat{\beta}_j^L = \begin{cases} y_j - \lambda/2 & \text{if } y_j > \lambda/2; \\ y_j + \lambda/2 & \text{if } y_j < -\lambda/2; \\ 0 & \text{if } |y_j| \le \lambda/2. \end{cases} \tag{6.15}$$

a)

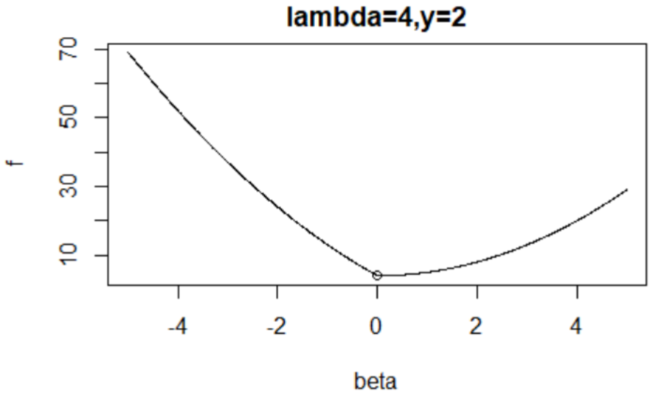
```
lambda = 1
y = 2
beta = seq(-5,5,0.05)
f = (y-beta)**2+lambda*beta**2
plot(beta,f,type='l',main=c('lambda=1,y=2'))
minbeta = y/(lambda+1)
minf = (y-minbeta)**2+lambda*minbeta**2
points(minbeta,minf)
```

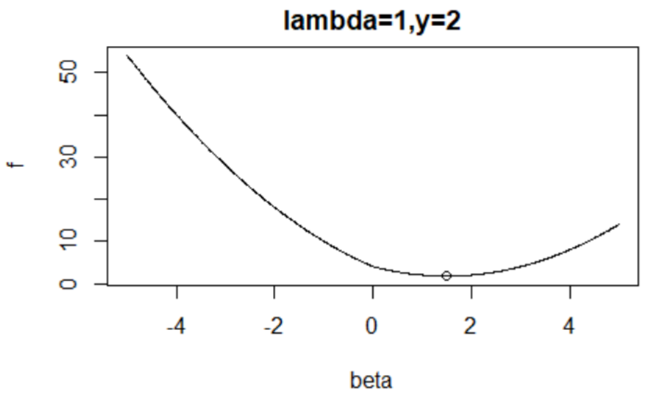


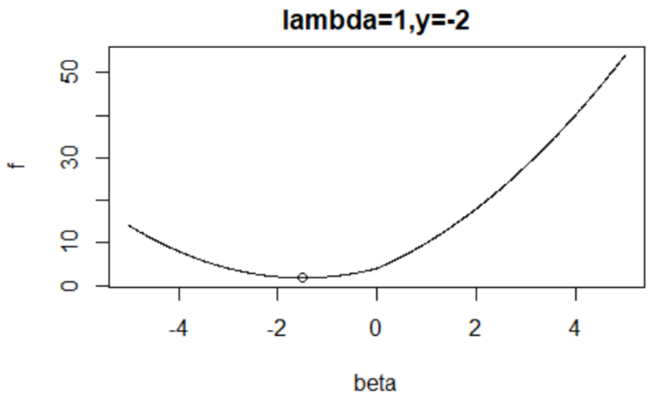


b)

```
lambda = 4
y = 2
beta = seq(-5,5,0.01)
f = (y-beta)**2+lambda*abs(beta)
minbeta=0
if(y>lambda/2){minbeta = y-lambda/2}
if(y< -lambda/2){minbeta = y+lambda/2}
minf = (y-minbeta)**2+lambda*abs(minbeta)
plot(beta,f,type='l',main=c('lambda=4,y=2'))
points(minbeta,minf)
```







HW2

Xinyue Lu

2020/1/23

Consider Prostate cancer study, page 49 the book attached

```
library('lasso2')
```

```
## R Package to solve regression problems while imposing
##   an L1 constraint on the parameters. Based on S-plus Release 2.1
## Copyright (C) 1998, 1999
## Justin Lokhorst   <jlokhors@stats.adelaide.edu.au>
## Berwin A. Turlach <bturlach@stats.adelaide.edu.au>
## Bill Venables     <wvenable@stats.adelaide.edu.au>
##
## Copyright (C) 2002
## Martin Maechler <maechler@stat.math.ethz.ch>
```

```
data(Prostate)
```

1. Discuss the correlation between predictors, list the pairs with strong correlations

The pairs with strong correlations were: (lcavol, lcp), (lcavol, lpsa), (svi, lcp), (lcp, pgg45), (gleason, pgg45)

```
corr.P = cor(Prostate)
col = colnames(Prostate)
cat('The pairs with strong correlations (r>0.6) were:\n')
```

```
## The pairs with strong correlations (r>0.6) were:
```

```
for (i in 1:9){
  for (j in 1:9){
    if (corr.P[i,j] > 0.6 & i<j){cat('(',col[i],',',col[j],')\n')}}}
```

```
## ( lcavol , lcp )
## ( lcavol , lpsa )
## ( svi , lcp )
## ( lcp , pgg45 )
## ( gleason , pgg45 )
```

```
cat('The correlation matrix was:\n\n')
```

```
## The correlation matrix was:
```

```
corr.P
```

```
##            lcavol      lweight       age         lbph         svi
## lcavol  1.0000000  0.194128286 0.2249999  0.027349703  0.53884500
## lweight 0.1941283  1.000000000 0.3075286  0.434934636  0.10877851
## age     0.2249999  0.307528614 1.0000000  0.350185896  0.11765804
## lbph    0.0273497  0.434934636 0.3501859  1.000000000 -0.08584324
## svi     0.5388450  0.108778505 0.1176580 -0.085843238  1.00000000
## lcp     0.6753105  0.100237795 0.1276678 -0.006999431  0.67311118
## gleason 0.4324171 -0.001275658 0.2688916  0.077820447  0.32041222
## pgg45   0.4336522  0.050846821 0.2761124  0.078460018  0.45764762
## lpsa    0.7344603  0.354120390 0.1695928  0.179809410  0.56621822
##                 lcp      gleason      pgg45      lpsa
## lcavol   0.675310484  0.432417056 0.43365225 0.7344603
## lweight  0.100237795 -0.001275658 0.05084682 0.3541204
## age      0.127667752  0.268891599 0.27611245 0.1695928
## lbph    -0.006999431  0.077820447 0.07846002 0.1798094
## svi      0.673111185  0.320412221 0.45764762 0.5662182
## lcp      1.000000000  0.514830063 0.63152825 0.5488132
## gleason  0.514830063  1.000000000 0.75190451 0.3689868
## pgg45    0.631528245  0.751904512 1.00000000 0.4223159
## lpsa     0.548813169  0.368986803 0.42231586 1.0000000
```

2. Fit the two linear models to the lpsa using the original values of the predictors and the standardized (unit variance) ones. Compare the significance of the resulting coefficients. In what follows consider standartized predictors.

When the original predictors were used, lcavol,lweight and svi were significant predictors. When the standardized predictors were used, the significance (p-value) of their coefficients were the same. Yet the intercept changed from insignificant to significant.

```r
m1<- lm(lpsa~ lcavol+lweight+age+ lbph  + svi +lcp +gleason + pgg45,
data=Prostate)
summary(m1)
```

```
##
## Call:
## lm(formula = lpsa ~ lcavol + lweight + age + lbph + svi + lcp +
##     gleason + pgg45, data = Prostate)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.73316 -0.37133 -0.01702  0.41414  1.63811
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.669399   1.296381   0.516  0.60690
## lcavol       0.587023   0.087920   6.677 2.11e-09 ***
## lweight      0.454461   0.170012   2.673  0.00896 **
## age         -0.019637   0.011173  -1.758  0.08229 .
## lbph         0.107054   0.058449   1.832  0.07040 .
## svi          0.766156   0.244309   3.136  0.00233 **
## lcp         -0.105474   0.091013  -1.159  0.24964
## gleason      0.045136   0.157464   0.287  0.77506
## pgg45        0.004525   0.004421   1.024  0.30885
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.7084 on 88 degrees of freedom
## Multiple R-squared:  0.6548, Adjusted R-squared:  0.6234
## F-statistic: 20.86 on 8 and 88 DF,  p-value: < 2.2e-16
```

```r
Prostate.s = Prostate
for (i in 1:8){Prostate.s[i] = scale(Prostate[i])}
m2<- lm(lpsa~ lcavol+lweight+age+ lbph  + svi +lcp +gleason + pgg45,
data=Prostate.s)
summary(m2)
```

```
##
## Call:
## lm(formula = lpsa ~ lcavol + lweight + age + lbph + svi + lcp +
##     gleason + pgg45, data = Prostate.s)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.73316 -0.37133 -0.01702  0.41414  1.63811
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  2.47839    0.07193  34.456  < 2e-16 ***
## lcavol       0.69188    0.10363   6.677 2.11e-09 ***
## lweight      0.22570    0.08443   2.673  0.00896 **
## age         -0.14620    0.08318  -1.758  0.08229 .
## lbph         0.15532    0.08480   1.832  0.07040 .
## svi          0.31718    0.10114   3.136  0.00233 **
## lcp         -0.14748    0.12726  -1.159  0.24964
## gleason      0.03259    0.11371   0.287  0.77506
## pgg45        0.12763    0.12470   1.024  0.30885
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.7084 on 88 degrees of freedom
## Multiple R-squared:  0.6548, Adjusted R-squared:  0.6234
## F-statistic: 20.86 on 8 and 88 DF,  p-value: < 2.2e-16
```

3. Read Section 3.3.1 and produce a plot similar to Figure 3.5 (page 58). List the predictors from the best model with 4 and 5 predictors.

Use all the data as training set, so the sum of residuals were larger than thoes in the textbook.

The predictors of the best model with 4 predictors were: lcavo, lweight, lbph, svi. The predictors of the best model with 5 predictors were: lcavo, lweight, lbph, svi, age.

```r
library(leaps)
```

```
## Warning: package 'leaps' was built under R version 3.6.2
```

```r
prostate.leaps <- regsubsets( lpsa ~ . ,method="exhaustive", data=Prostate.s, nbest=70,really.big=TRUE )
prostate.leaps.sum <- summary( prostate.leaps )
prostate.models <- prostate.leaps.sum$which
prostate.models.size <- as.numeric(attr(prostate.models, "dimnames")[[1]])
prostate.models.rss <- prostate.leaps.sum$rss
prostate.models.best.rss <- tapply( prostate.models.rss, prostate.models.size, min )
prostate.models.best.rss
```

```
##        1        2        3        4        5        6        7        8
## 58.91478 52.96636 47.78496 46.48490 45.52565 44.86669 44.20436 44.16313
```

```r
prostate.dummy <- lm( lpsa ~ 1, data=Prostate.s ) # only intercept model
prostate.models.best.rss <- c(sum(resid(prostate.dummy)^2),prostate.models.best.rss)

cat('The best model with 4 predictors:\n\n')
```

```
## The best model with 4 predictors:
```

```r
index.best4 = which( prostate.models.rss == prostate.models.best.rss[5])
prostate.models[index.best4,]
```

```
## (Intercept)      lcavol     lweight         age        lbph         svi
##        TRUE        TRUE        TRUE       FALSE        TRUE        TRUE
##         lcp     gleason       pgg45
##       FALSE       FALSE       FALSE
```

```r
cat('The best model with 5 predictors:\n\n')
```

```
## The best model with 5 predictors:
```

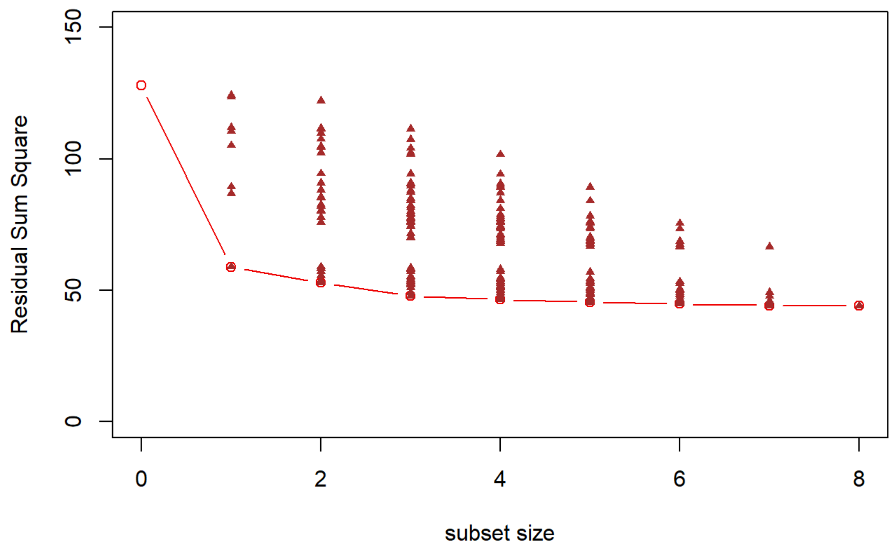
```r
index.best5 = which( prostate.models.rss == prostate.models.best.rss[6])
prostate.models[index.best5,]
```

```
## (Intercept)      lcavol     lweight         age        lbph         svi
##        TRUE        TRUE        TRUE        TRUE        TRUE        TRUE
##         lcp     gleason       pgg45
##       FALSE       FALSE       FALSE
```

```r
cat('Plot Figure 3.5 with all the data:\n\n')
```

```
## Plot Figure 3.5 with all the data:
```

```r
plot( 0:8, prostate.models.best.rss, ylim=c(0, 150),
type="b",xlab="subset size",ylab="Residual Sum Square",col="red2" )
points( prostate.models.size, prostate.models.rss,
pch=17, col="brown",cex=0.7 )
```



4. Run Forward, Backward and Stepwise Selection, write down the resulting best models, are they the same? Compare the models with the one resulted from the best subset selection with the same number of predictors.

All the 3 methods, Forward, Backward and Stepwise selection methods, selected the same 5 predictors in the final model, which were the same 5 predictors as the best subset method selected with 5 predictors in the model.

```
library(MASS)
fit1 <- lm(lpsa ~ ., Prostate.s)
fit2 <- lm(lpsa ~ 1, Prostate.s)
step.forward <- stepAIC(fit2, direction="forward",scope=list(upper=fit1,lower=fit2))
```

```
## Start:  AIC=28.84
## lpsa ~ 1
##
##           Df Sum of Sq     RSS     AIC
## + lcavol   1    69.003  58.915 -44.366
## + svi      1    41.011  86.907  -6.658
## + lcp      1    38.528  89.389  -3.926
## + pgg45    1    22.814 105.103  11.783
## + gleason  1    17.416 110.502  16.641
## + lweight  1    16.041 111.877  17.841
## + lbph     1     4.136 123.782  27.650
## + age      1     3.679 124.239  28.007
## <none>                 127.918  28.838
##
## Step:  AIC=-44.37
## lpsa ~ lcavol
##
##           Df Sum of Sq    RSS     AIC
## + lweight  1    5.9484 52.966 -52.690
## + svi      1    5.2375 53.677 -51.397
## + lbph     1    3.2658 55.649 -47.898
## + pgg45    1    1.6980 57.217 -45.203
## <none>                 58.915 -44.366
## + lcp      1    0.6562 58.259 -43.452
## + gleason  1    0.4156 58.499 -43.053
## + age      1    0.0025 58.912 -42.370
##
## Step:  AIC=-52.69
## lpsa ~ lcavol + lweight
##
##           Df Sum of Sq    RSS     AIC
## + svi      1    5.1814 47.785 -60.676
## + pgg45    1    1.9489 51.017 -54.327
## <none>                 52.966 -52.690
## + lcp      1    0.8371 52.129 -52.235
## + gleason  1    0.7810 52.185 -52.131
## + lbph     1    0.6751 52.291 -51.935
## + age      1    0.4200 52.546 -51.462
##
## Step:  AIC=-60.68
## lpsa ~ lcavol + lweight + svi
##
##           Df Sum of Sq    RSS     AIC
## + lbph     1   1.30006 46.485 -61.352
## <none>                 47.785 -60.676
## + pgg45    1   0.57347 47.211 -59.847
## + age      1   0.40252 47.382 -59.497
## + gleason  1   0.38898 47.396 -59.469
## + lcp      1   0.06411 47.721 -58.806
##
## Step:  AIC=-61.35
## lpsa ~ lcavol + lweight + svi + lbph
##
##           Df Sum of Sq    RSS     AIC
## + age      1   0.95925 45.526 -61.374
## <none>                 46.485 -61.352
## + pgg45    1   0.35332 46.132 -60.092
## + gleason  1   0.21254 46.272 -59.796
## + lcp      1   0.10230 46.383 -59.565
##
## Step:  AIC=-61.37
## lpsa ~ lcavol + lweight + svi + lbph + age
##
##           Df Sum of Sq    RSS     AIC
## <none>                 45.526 -61.374
## + pgg45    1   0.65896 44.867 -60.788
## + gleason  1   0.45598 45.070 -60.351
## + lcp      1   0.12927 45.396 -59.650
```

```
step.forward$anova
```

```
## Stepwise Model Path 
## Analysis of Deviance Table
## 
## Initial Model:
## lpsa ~ 1
## 
## Final Model:
## lpsa ~ lcavol + lweight + svi + lbph + age
## 
## 
##         Step Df   Deviance Resid. Df Resid. Dev       AIC
## 1                                 96  127.91766  28.83755
## 2  + lcavol  1 69.0028744        95   58.91478 -44.36603
## 3 + lweight  1  5.9484273        94   52.96636 -52.69024
## 4     + svi  1  5.1813959        93   47.78496 -60.67600
## 5    + lbph  1  1.3000579        92   46.48490 -61.35159
## 6     + age  1  0.9592528        91   45.52565 -61.37420
```

```
step.backward <- stepAIC(m2, direction="backward")
```

```
## Start:  AIC=-58.32
## lpsa ~ lcavol + lweight + age + lbph + svi + lcp + gleason + 
##     pgg45
## 
##           Df Sum of Sq    RSS     AIC
## - gleason  1    0.0412 44.204 -60.231
## - pgg45    1    0.5258 44.689 -59.174
## - lcp      1    0.6740 44.837 -58.852
## <none>                 44.163 -58.322
## - age      1    1.5503 45.713 -56.975
## - lbph     1    1.6836 45.847 -56.693
## - lweight  1    3.5860 47.749 -52.749
## - svi      1    4.9355 49.099 -50.045
## - lcavol   1   22.3722 66.535 -20.567
## 
## Step:  AIC=-60.23
## lpsa ~ lcavol + lweight + age + lbph + svi + lcp + pgg45
## 
##           Df Sum of Sq    RSS     AIC
## - lcp      1    0.6623 44.867 -60.788
## <none>                 44.204 -60.231
## - pgg45    1    1.1920 45.396 -59.650
## - age      1    1.5166 45.721 -58.959
## - lbph     1    1.7053 45.910 -58.559
## - lweight  1    3.5461 47.751 -54.746
## - svi      1    4.8984 49.103 -52.037
## - lcavol   1   23.5039 67.708 -20.872
## 
## Step:  AIC=-60.79
## lpsa ~ lcavol + lweight + age + lbph + svi + pgg45
## 
##           Df Sum of Sq    RSS     AIC
## - pgg45    1    0.6590 45.526 -61.374
## <none>                 44.867 -60.788
## - age      1    1.2649 46.132 -60.092
## - lbph     1    1.6465 46.513 -59.293
## - lweight  1    3.5646 48.431 -55.373
## - svi      1    4.2503 49.117 -54.009
## - lcavol   1   25.4190 70.286 -19.248
## 
## Step:  AIC=-61.37
## lpsa ~ lcavol + lweight + age + lbph + svi
## 
##           Df Sum of Sq    RSS     AIC
## <none>                 45.526 -61.374
## - age      1    0.9593 46.485 -61.352
## - lbph     1    1.8568 47.382 -59.497
## - lweight  1    3.2250 48.751 -56.735
## - svi      1    5.9517 51.477 -51.456
## - lcavol   1   28.7666 74.292 -15.870
```

```
step.backward$anova
```

```
## Stepwise Model Path 
## Analysis of Deviance Table
## 
## Initial Model:
## lpsa ~ lcavol + lweight + age + lbph + svi + lcp + gleason + 
##     pgg45
## 
## Final Model:
## lpsa ~ lcavol + lweight + age + lbph + svi
## 
## 
##        Step Df   Deviance Resid. Df Resid. Dev       AIC
## 1                                88   44.16313 -58.32161
## 2 - gleason  1 0.04123419        89   44.20436 -60.23109
## 3     - lcp  1 0.66232990        90   44.86669 -60.78848
## 4   - pgg45  1 0.65895836        91   45.52565 -61.37420
```

```
step.both <- stepAIC(m2, direction="both")
```

```
## Start:  AIC=-58.32
## lpsa ~ lcavol + lweight + age + lbph + svi + lcp + gleason + 
##     pgg45
## 
##           Df Sum of Sq    RSS     AIC
## - gleason  1    0.0412 44.204 -60.231
## - pgg45    1    0.5258 44.689 -59.174
## - lcp      1    0.6740 44.837 -58.852
## <none>                 44.163 -58.322
## - age      1    1.5503 45.713 -56.975
## - lbph     1    1.6836 45.847 -56.693
## - lweight  1    3.5860 47.749 -52.749
## - svi      1    4.9355 49.099 -50.045
## - lcavol   1   22.3722 66.535 -20.567
## 
## Step:  AIC=-60.23
## lpsa ~ lcavol + lweight + age + lbph + svi + lcp + pgg45
## 
##           Df Sum of Sq    RSS     AIC
## - lcp      1    0.6623 44.867 -60.788
## <none>                 44.204 -60.231
## - pgg45    1    1.1920 45.396 -59.650
## - age      1    1.5166 45.721 -58.959
## - lbph     1    1.7053 45.910 -58.559
## + gleason  1    0.0412 44.163 -58.322
## - lweight  1    3.5461 47.751 -54.746
## - svi      1    4.8984 49.103 -52.037
## - lcavol   1   23.5039 67.708 -20.872
## 
## Step:  AIC=-60.79
## lpsa ~ lcavol + lweight + age + lbph + svi + pgg45
## 
##           Df Sum of Sq    RSS     AIC
## - pgg45    1    0.6590 45.526 -61.374
## <none>                 44.867 -60.788
## + lcp      1    0.6623 44.204 -60.231
## - age      1    1.2649 46.132 -60.092
## - lbph     1    1.6465 46.513 -59.293
## + gleason  1    0.0296 44.837 -58.852
## - lweight  1    3.5646 48.431 -55.373
## - svi      1    4.2503 49.117 -54.009
## - lcavol   1   25.4190 70.286 -19.248
## 
## Step:  AIC=-61.37
## lpsa ~ lcavol + lweight + age + lbph + svi
## 
##           Df Sum of Sq    RSS     AIC
## <none>                 45.526 -61.374
## - age      1    0.9593 46.485 -61.352
## + pgg45    1    0.6590 44.867 -60.788
## + gleason  1    0.4560 45.070 -60.351
## + lcp      1    0.1293 45.396 -59.650
## - lbph     1    1.8568 47.382 -59.497
## - lweight  1    3.2250 48.751 -56.735
## - svi      1    5.9517 51.477 -51.456
## - lcavol   1   28.7666 74.292 -15.870
```

```
step.both$anova
```

```
## Stepwise Model Path 
## Analysis of Deviance Table
##
## Initial Model:
## lpsa ~ lcavol + lweight + age + lbph + svi + lcp + gleason + 
##     pgg45
##
## Final Model:
## lpsa ~ lcavol + lweight + age + lbph + svi
##
##
##        Step Df   Deviance Resid. Df Resid. Dev       AIC
## 1                                88   44.16313 -58.32161
## 2 - gleason  1 0.04123419        89   44.20436 -60.23109
## 3     - lcp  1 0.66232990        90   44.86669 -60.78848
## 4   - pgg45  1 0.65895836        91   45.52565 -61.37420
```