

HUDM5126 Linear Models and Regression Analysis Homework 8

Yifei Dong

10/22/2020

0. Data Preparation

```
getwd()

## [1] "/Users/yifei/Documents/Teachers College/Linear Models and Regression/Week 8/hw8/hw8_R"

library(dplyr)

##
## Attaching package: 'dplyr'
##
## The following objects are masked from 'package:stats':
##
##   filter, lag
##
## The following objects are masked from 'package:base':
##
##   intersect, setdiff, setequal, union

library(clusterGeneration)

## Loading required package: MASS

##
## Attaching package: 'MASS'
##
## The following object is masked from 'package:dplyr':
##
##   select

library(bestglm)

## Loading required package: leaps

library(ggplot2)
library(leaps)
library(MASS)
```

1. Model Selection with Kidney Dataset

In this exercise, we will use kidney function data from Exercise 9.15 on p.378.

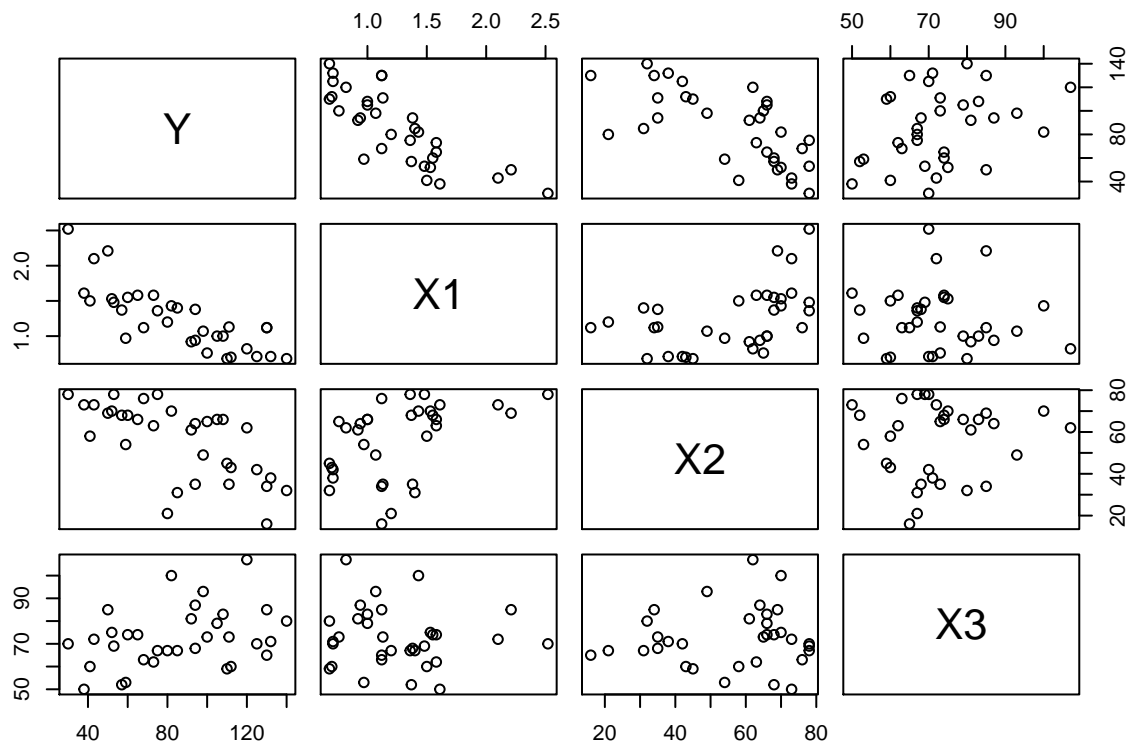
```
# Load Dataset
kidney <- read.table(paste("http://users.stat.ufl.edu/",
                           "~rrandles/sta4210/Rclassnotes/data/textdatasets/KutnerData/",
                           "Chapter%20%209%20Data%20Sets/CH09PR15.txt", sep = ""))
```

```
# change column names
kidney <- kidney %>%
  dplyr::select("Y" = V1, "X1" = V2, "X2" = V3, "X3" = V4)
head(kidney, 10)
```

```
##      Y   X1 X2 X3
## 1  132 0.71 38 71
## 2   53 1.48 78 69
## 3   50 2.21 69 85
## 4   82 1.43 70 100
## 5  110 0.68 45 59
## 6  100 0.76 65 73
## 7   68 1.12 76 63
## 8   92 0.92 61 81
## 9   60 1.55 68 74
## 10  94 0.94 64 87
```

a) Obtain the scatterplot matrix. What does it suggest?

```
pairs(kidney)
```



```
cor(kidney)
```

```
##      Y      X1      X2      X3
## Y  1.000000 -0.80181086 -0.66787239  0.34591487
## X1 -0.8018109  1.00000000  0.46773179 -0.08898262
## X2 -0.6678724  0.46773179  1.00000000  0.06848147
## X3  0.3459149 -0.08898262  0.06848147  1.00000000
```

We can conclude from the scatterplot, as well as the correlation table. The response variable Y has a

strong negative relationship with X_1 and X_2 and a positive relationship with X_3 . Also, I observe some multicollinearity problem because there exists a positive correlation between X_1 and X_2 .

- b) Fit the multiple regression model containing all three predictors as first-order terms. Are all predictors significant?

```
reg1 <- lm(Y~X1+X2+X3, data = kidney)
summary(reg1)
```

```
##
## Call:
## lm(formula = Y ~ X1 + X2 + X3, data = kidney)
##
## Residuals:
```

	Min	1Q	Median	3Q	Max
	-28.668	-7.002	1.518	9.905	16.006

```
##
## Coefficients:
```

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	120.0473	14.7737	8.126	5.84e-09 ***
X1	-39.9393	5.6000	-7.132	7.55e-08 ***
X2	-0.7368	0.1414	-5.211	1.41e-05 ***
X3	0.7764	0.1719	4.517	9.69e-05 ***

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 12.46 on 29 degrees of freedom
## Multiple R-squared:  0.8548, Adjusted R-squared:  0.8398
## F-statistic: 56.92 on 3 and 29 DF,  p-value: 2.885e-12
```

The model shows that all three predictors as first-order terms are significant because of the low p-value. Therefore, all of the three predictors should be retained.

2. Best Subset Selection

- c) Perform best subset selection in order to choose the best model from the pool of possible predictors that includes X_1, X_2, X_3 plus all quadratic terms and all possible interactions. (That is, you should have 9 predictor variables to choose from.) What are the *two* best models according to BIC? Include a plot BIC as evidence which one is the lowest/highest value. Report the coefficients of the best models obtained.

```
# Generate variables
kidney <- kidney %>%
  mutate(X1sq = X1^2, X2sq = X2^2, X3sq = X3^2,
         X1X2 = X1*X2, X1X3 = X1*X3, X2X3 = X2*X3)
head(kidney, 10)

##      Y   X1 X2  X3   X1sq X2sq  X3sq   X1X2   X1X3 X2X3
## 1  132 0.71 38  71 0.5041 1444  5041  26.98  50.41 2698
## 2   53 1.48 78  69 2.1904 6084  4761 115.44 102.12 5382
## 3   50 2.21 69  85 4.8841 4761  7225 152.49 187.85 5865
## 4   82 1.43 70 100 2.0449 4900 10000 100.10 143.00 7000
## 5  110 0.68 45  59 0.4624 2025  3481  30.60  40.12 2655
## 6  100 0.76 65  73 0.5776 4225  5329  49.40  55.48 4745
## 7   68 1.12 76  63 1.2544 5776  3969  85.12  70.56 4788
## 8   92 0.92 61  81 0.8464 3721  6561  56.12  74.52 4941
## 9   60 1.55 68  74 2.4025 4624  5476 105.40 114.70 5032
## 10  94 0.94 64  87 0.8836 4096  7569  60.16  81.78 5568

# Regression with all predictors
summary(lm(Y ~ ., data = kidney))

##
## Call:
## lm(formula = Y ~ ., data = kidney)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -27.663  -6.258   1.832   5.925  20.164
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  68.408996   98.118427   0.697   0.4927
## X1          -93.900207   48.001432  -1.956   0.0627 .
## X2           -1.220775    1.754197  -0.696   0.4935
## X3            3.310486    1.795766   1.843   0.0782 .
## X1sq          8.645235   14.996001   0.577   0.5699
## X2sq          0.008801    0.009514   0.925   0.3646
## X3sq         -0.011568    0.010288  -1.124   0.2725
## X1X2          0.422545    0.732978   0.576   0.5699
## X1X3         -0.030914    0.612987  -0.050   0.9602
## X2X3         -0.011938    0.019095  -0.625   0.5380
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 11.95 on 23 degrees of freedom
## Multiple R-squared:  0.894, Adjusted R-squared:  0.8526
## F-statistic: 21.56 on 9 and 23 DF, p-value: 3.534e-09
```

```

# Reorder the variables
kidney <- kidney[c(2:10, 1)]

bs1 <- bestglm(Xy = kidney,
               family = gaussian,
               IC = "BIC")
bs1$Subsets

##      (Intercept)      X1      X2      X3 X1sq X2sq X3sq X1X2 X1X3 X2X3
## 0      TRUE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## 1      TRUE FALSE FALSE FALSE FALSE FALSE FALSE  TRUE FALSE FALSE
## 2      TRUE FALSE FALSE  TRUE FALSE FALSE FALSE  TRUE FALSE FALSE
## 3      TRUE  TRUE FALSE  TRUE FALSE FALSE FALSE FALSE FALSE  TRUE
## 4*      TRUE  TRUE  TRUE  TRUE FALSE FALSE FALSE  TRUE FALSE FALSE
## 5      TRUE  TRUE  TRUE  TRUE FALSE FALSE  TRUE  TRUE FALSE FALSE
## 6      TRUE  TRUE FALSE  TRUE  TRUE  TRUE  TRUE FALSE FALSE  TRUE
## 7      TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE FALSE FALSE  TRUE
## 8      TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE FALSE  TRUE
## 9      TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE
##      logLikelihood      BIC
## 0      -112.94399 225.8880
## 1      -92.65821 188.8129
## 2      -85.63381 178.2606
## 3      -80.70050 171.8905
## 4*      -78.12089 170.2278
## 5      -76.87208 171.2267
## 6      -76.32314 173.6253
## 7      -76.15234 176.7802
## 8      -75.90910 179.7903
## 9      -75.90727 183.2831

# Select two best models using BIC
bs1$BestModels %>%
  top_n(-2)

```

Selecting by Criterion

```

##      X1      X2      X3 X1sq X2sq X3sq X1X2 X1X3 X2X3 Criterion
## 1 TRUE  TRUE  TRUE FALSE FALSE FALSE  TRUE FALSE FALSE  170.2278
## 2 TRUE FALSE  TRUE  TRUE FALSE FALSE FALSE FALSE  TRUE  170.6920

```

The best model includes the following 4 predictors: X_1 , X_2 , X_3 , and X_1X_2 .

The second best model includes the following 4 predictors: X_1 , X_3 , X_1^2 , and X_2X_3 .

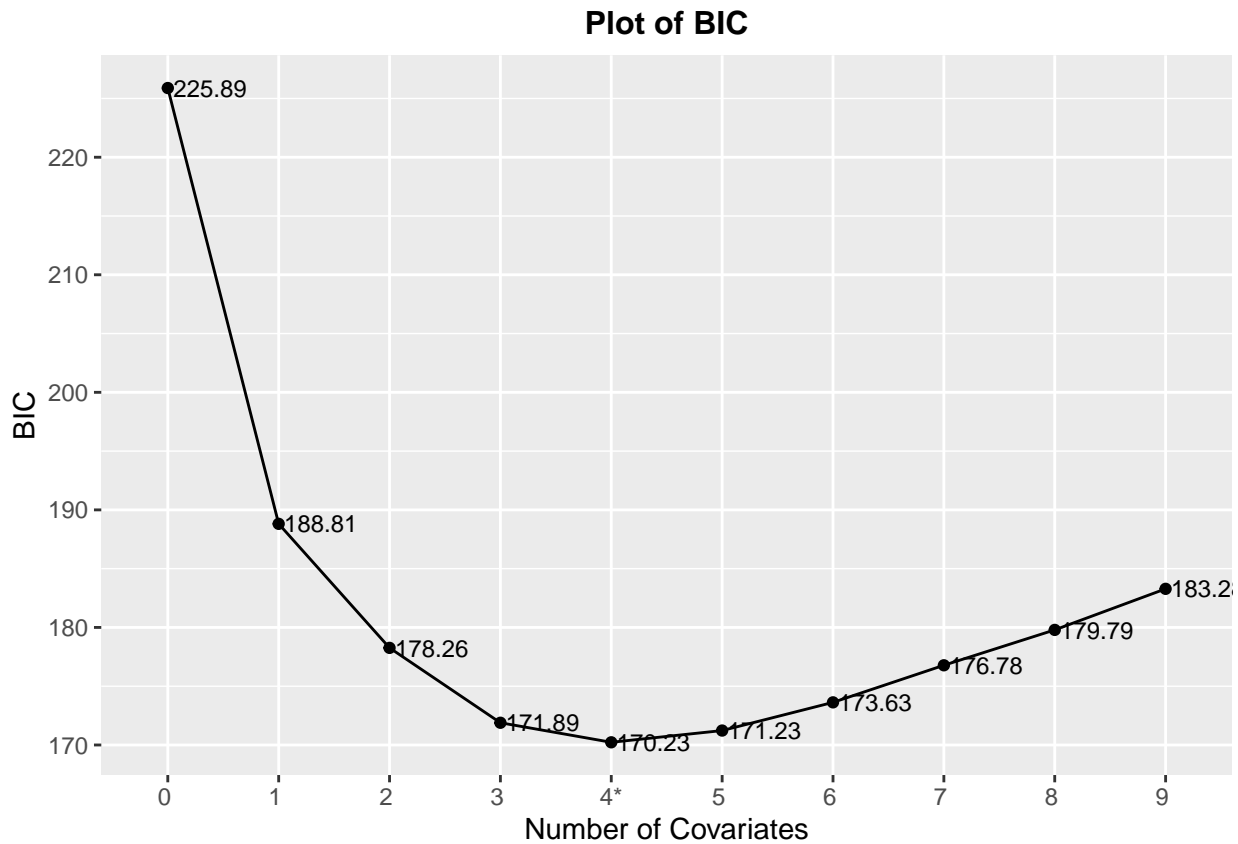
Both two models' BIC values are very close.

```

# Plot of BIC use ggplot2 package
g1 <- ggplot(bs1$Subsets, aes(x = row.names(bs1$Subsets),
                              y = BIC, label = round(BIC, digits = 2)))+
  geom_point()+
  geom_text(hjust = 0, nudge_x = 0.05, size = 3)+
  scale_x_discrete("Number of Covariates")+
  scale_y_continuous("BIC")+
  geom_line(group = 1)+
  ggtitle("Plot of BIC")+

```

```
theme(plot.title = element_text(size = 12, face = "bold", color = "black", hjust = 0.5))
g1
```



```
# The best model
```

```
bs1$BestModel
```

```
##
```

```
## Call:
```

```
## lm(formula = y ~ ., data = data.frame(Xy[, c(bestset[-1], FALSE),
```

```
## drop = FALSE], y = y))
```

```
##
```

```
## Coefficients:
```

```
## (Intercept)          X1          X2          X3          X1X2
```

```
## 181.6978    -95.5593    -1.7536     0.7951     0.8620
```

Based on the BIC plot, the lowest value is the model has 4 covariates and the highest value is the model with 0 covariates, which only includes the intercept.

So the estimated best regression model is:

$$\hat{Y}_i = 181.6978 - 95.5593X_{i1} - 1.7536X_{i2} + 0.7951X_{i3} + 0.8620X_{i1}X_{i2}$$

d) Repeat part c) but this time using AIC. Are the results identical?

```
bs2 <- bestglm(Xy = kidney,
               family = gaussian,
```

```
IC = "AIC")
bs2$Subsets
```

```
##      (Intercept)      X1      X2      X3 X1sq X2sq X3sq X1X2 X1X3 X2X3
## 0      TRUE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## 1      TRUE FALSE FALSE FALSE FALSE FALSE FALSE  TRUE FALSE FALSE
## 2      TRUE FALSE FALSE  TRUE FALSE FALSE FALSE  TRUE FALSE FALSE
## 3      TRUE  TRUE FALSE  TRUE FALSE FALSE FALSE FALSE FALSE  TRUE
## 4      TRUE  TRUE  TRUE  TRUE FALSE FALSE FALSE  TRUE FALSE FALSE
## 5*     TRUE  TRUE  TRUE  TRUE FALSE FALSE  TRUE  TRUE FALSE FALSE
## 6      TRUE  TRUE FALSE  TRUE  TRUE  TRUE  TRUE FALSE FALSE  TRUE
## 7      TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE FALSE FALSE  TRUE
## 8      TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE FALSE  TRUE
## 9      TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE
##      logLikelihood      AIC
## 0      -112.94399 225.8880
## 1      -92.65821 187.3164
## 2      -85.63381 175.2676
## 3      -80.70050 167.4010
## 4      -78.12089 164.2418
## 5*     -76.87208 163.7442
## 6      -76.32314 164.6463
## 7      -76.15234 166.3047
## 8      -75.90910 167.8182
## 9      -75.90727 169.8145
```

```
# Select two best models according to AIC
```

```
bs2$BestModels %>%
  top_n(-2)
```

```
## Selecting by Criterion
```

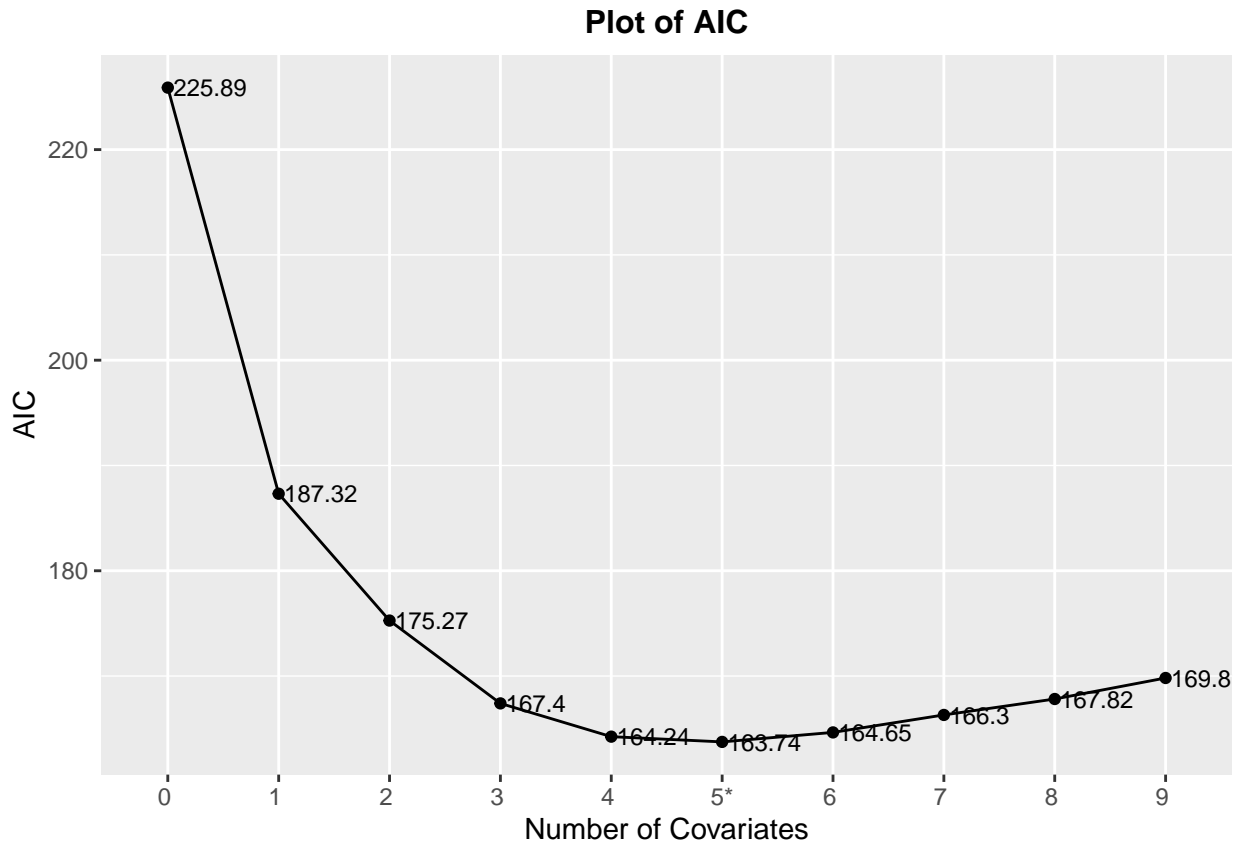
```
##      X1      X2      X3 X1sq X2sq X3sq X1X2 X1X3 X2X3 Criterion
## 1 TRUE TRUE TRUE FALSE FALSE  TRUE TRUE FALSE FALSE 163.7442
## 2 TRUE TRUE TRUE FALSE FALSE FALSE TRUE FALSE FALSE 164.2418
```

Using AIC, the best model includes the following 5 predictors: X_1, X_2, X_3, X_3^2 and X_1X_2 .

The second best model also includes 4 predictors, which are: X_1, X_2, X_3 and X_1X_2 .

```
# Plot of AIC use ggplot2 package
```

```
g2 <- ggplot(bs2$Subsets, aes(x = row.names(bs2$Subsets),
                              y = AIC, label = round(AIC, digits = 2)))+
  geom_point()+
  geom_text(hjust = 0, nudge_x = 0.05, size = 3)+
  scale_x_discrete("Number of Covariates")+
  scale_y_continuous("AIC")+
  geom_line(group = 1)+
  ggtitle("Plot of AIC")+
  theme(plot.title = element_text(size = 12, face = "bold", color = "black", hjust = 0.5))
g2
```



```
# The best model using AIC
bs2$BestModel
```

```
##
## Call:
## lm(formula = y ~ ., data = data.frame(Xy[, c(bestset[-1], FALSE),
##     drop = FALSE], y = y))
##
## Coefficients:
## (Intercept)          X1          X2          X3          X3sq          X1X2
##  103.97280   -93.35747   -1.64122    2.77202   -0.01299    0.80433
```

Based on the AIC plot, the lowest AIC value is the model that has 5 covariates and the highest value is the model with 0 covariates, which only includes the intercept.

So the estimated best regression model is:

$$\hat{Y}_i = 103.9728 - 93.3575X_{i1} - 1.6412X_{i2} + 2.7720X_{i3} - 0.0130X_{i3}^2 + 0.8043X_{i1}X_{i2}$$

The results are not identical. BIC suggests the best model should have 4 covariates, while AIC suggests the best model should have 5 covariates. This is consistent with the theory because BIC imposes additional penalty for more complexity in the model, which means BIC tends to produce a more easily interpretable model.

e) Repeat part c) using adjusted R^2 ? (You might want to use the `leaps` function from the `leaps` package)

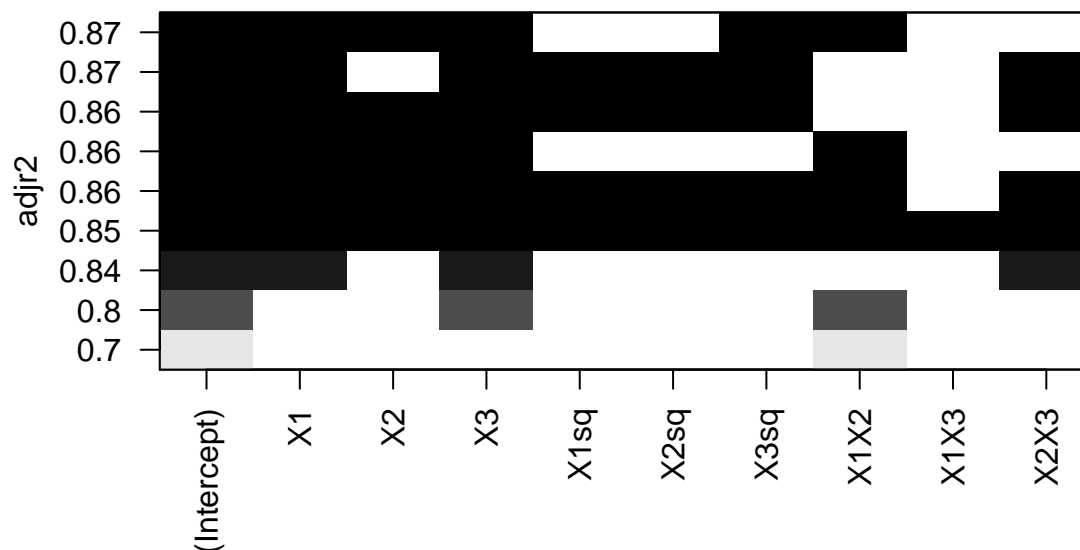
```
rsOut <- regsubsets(Y ~ ., data = kidney, nvmax = 9)
summary(rsOut)
```



```
## Subset selection object
## Call: regsubsets.formula(Y ~ ., data = kidney, nvmax = 9)
## 9 Variables (and intercept)
##      Forced in Forced out
## X1      FALSE      FALSE
## X2      FALSE      FALSE
## X3      FALSE      FALSE
## X1sq     FALSE      FALSE
## X2sq     FALSE      FALSE
## X3sq     FALSE      FALSE
## X1X2     FALSE      FALSE
## X1X3     FALSE      FALSE
## X2X3     FALSE      FALSE
## 1 subsets of each size up to 9
## Selection Algorithm: exhaustive
##      X1 X2 X3 X1sq X2sq X3sq X1X2 X1X3 X2X3
## 1 ( 1 ) " " " " " " " " " " " " " " " "
## 2 ( 1 ) " " " " " " " " " " " " " " "
## 3 ( 1 ) "*" " " " " " " " " " " " " "
## 4 ( 1 ) "*" "*" " " " " " " " " " " "
## 5 ( 1 ) "*" "*" " " " " " " " " " " "
## 6 ( 1 ) "*" " " " " "*" " " " " " " "
## 7 ( 1 ) "*" "*" " " "*" " " " " " " "
## 8 ( 1 ) "*" "*" " " "*" " " " " " " "
## 9 ( 1 ) "*" "*" " " "*" " " " " " " "
```

```
# Adjusted R2
plot(rsOut, scale = "adjr2", main = "Adjusted R squared")
```

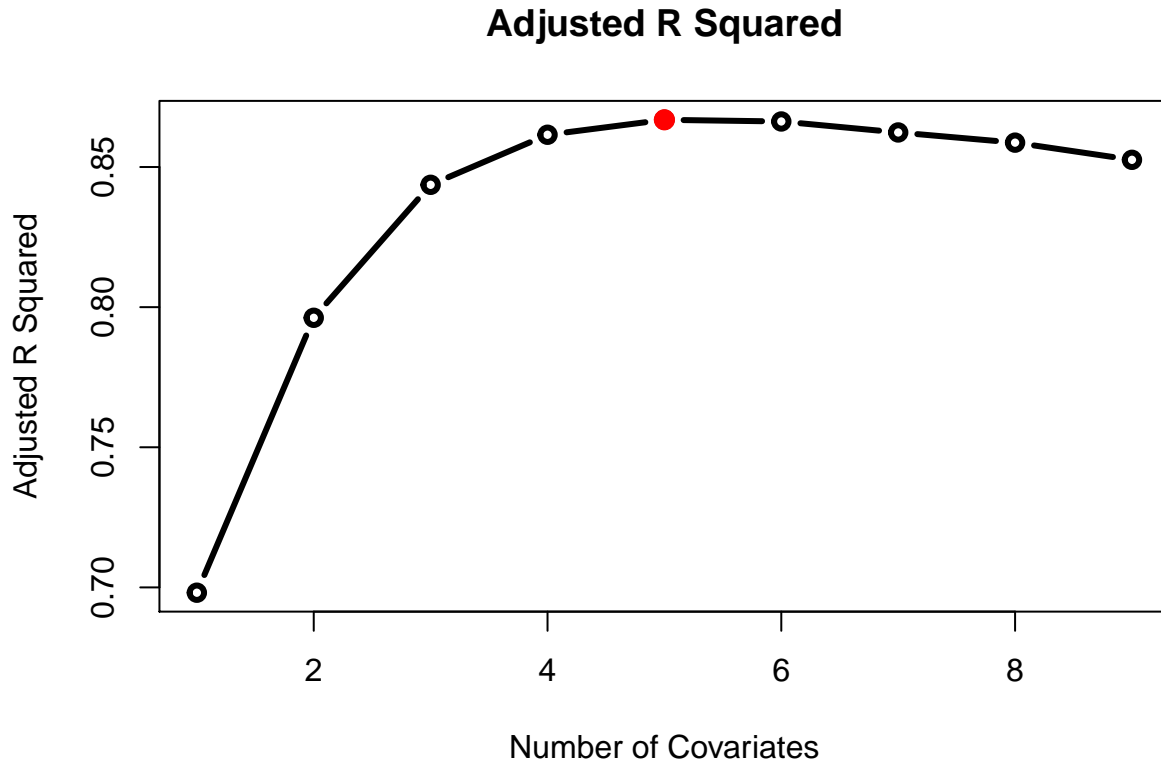
Adjusted R squared



```
reg.summary <- summary(rsOut)
plot(1:9, reg.summary$adjr2, xlab="Number of Covariates",
     ylab="Adjusted R Squared",
     main="Adjusted R Squared", type="b",
```

```
lwd = 3)
which.max(reg.summary$adjr2) # Model with 5 covariates has the largest R-Squared

## [1] 5
points(5,reg.summary$adjr2[5], col="red",cex=2,pch=20) # add a red point for the best model
```



```
# Let's find those five covariates
bs3 <- leaps(x=kidney[, 1:9], y = kidney[, 10], nbest = 9, method = "adjr2")
colnames(bs3$which) <- colnames(kidney[, 0:9]) # change column names to make it clear
```

```
# Sort the two best models by adjusted R2
sort(bs3$adjr2, decreasing = TRUE)
```

```
## [1] 0.86684974 0.86625309 0.86523624 0.86442943 0.86429424 0.86382503
## [7] 0.86357578 0.86325172 0.86278452 0.86233566 0.86226141 0.86181642
## [13] 0.86151029 0.86132854 0.86129136 0.86123924 0.86108542 0.86104547
## [19] 0.86067757 0.86057782 0.85976727 0.85976199 0.85954863 0.85899428
## [25] 0.85887848 0.85869815 0.85866310 0.85856170 0.85847391 0.85816939
## [31] 0.85786324 0.85667234 0.85667216 0.85631283 0.85573878 0.85446881
## [37] 0.85345773 0.85257090 0.85201885 0.85191420 0.85136197 0.85094778
## [43] 0.84883436 0.84801042 0.84365820 0.83979979 0.83783736 0.83520681
## [49] 0.83216621 0.83181519 0.83172029 0.82955652 0.82854662 0.82709626
## [55] 0.82522981 0.79620153 0.78550843 0.73618198 0.73298252 0.72967518
## [61] 0.71650290 0.70330969 0.70081943 0.70014903 0.69810991 0.63138133
## [67] 0.55003092 0.46553310 0.42818429 0.32942820 0.11238896 0.09125894
## [73] 0.08039406
```

```
# Let's find the two best models with largest R-Squared
head(bs3$which[order(bs3$adjr2, decreasing = TRUE), ], 2)
```

```
##      X1      X2      X3  X1sq  X2sq  X3sq  X1X2  X1X3  X2X3
## 5 TRUE  TRUE TRUE FALSE FALSE TRUE  TRUE FALSE FALSE
## 6 TRUE FALSE TRUE  TRUE  TRUE TRUE FALSE FALSE  TRUE
```

Using adjusted R^2 , the best model chooses the following 5 covariates, which are X_1, X_2, X_3, X_3^2 , and X_1X_2 . This matches our graph.

The second best model chooses 6 covariates, which are $X_1, X_3, X_1^2, X_2^2, X_3^2$ and X_2X_3 . This also matches our graph.

```
# Best model using adjusted R^2
coefficients(rsOut, 5) #R^2 = 0.8668
```

```
## (Intercept)          X1          X2          X3          X3sq          X1X2
## 103.97279806 -93.35746856 -1.64121863  2.77202222 -0.01299075  0.80432836
```

So the estimated regression model with largest adjusted R^2 is:

$$\hat{Y}_i = 103.9728 - 93.3575X_{i1} - 1.6412X_{i2} + 2.7720X_{i3} - 0.0130X_{i3}^2 + 0.8043X_{i1}X_{i2}$$

This result matches the result we get using AIC. AIC's best model has the largest adjusted R^2 .

3. Forward Stepwise Selection

- f) Repeat parts c & d, but this time using forward selection, as shown in class with the `stepAIC` function. How does your answer compare to the results in parts c to e?

```
# Forward Selection with AIC
min.model <- lm(Y ~ 1, data = kidney) # start with only intercept
max.model <- lm(Y ~ ., data = kidney) # stop with model with everything
scp <- list(lower = min.model, upper = max.model)
fwd <- stepAIC(min.model, direction = 'forward', scope = scp)

## Start:  AIC=227.89
## Y ~ 1
##
##           Df Sum of Sq    RSS    AIC
## + X1X2    1   21930.7  9064.8 189.32
## + X1      1   19927.0 11068.5 195.91
## + X1sq    1   17484.3 13511.2 202.49
## + X2sq    1   14947.1 16048.4 208.17
## + X2      1   13825.7 17169.9 210.40
## + X1X3    1   10860.3 20135.2 215.65
## + X2X3    1    4343.3 26652.2 224.91
## + X3      1    3708.8 27286.7 225.68
## + X3sq    1    3382.6 27612.9 226.07
## <none>                30995.5 227.89
##
## Step:  AIC=189.32
## Y ~ X1X2
##
##           Df Sum of Sq    RSS    AIC
## + X3      1    3142.79 5922.0 177.27
## + X3sq    1    2832.07 6232.8 178.96
## + X2X3    1    1305.75 7759.1 186.18
## + X1X3    1     826.88 8237.9 188.16
## <none>                9064.8 189.32
## + X1      1     371.15 8693.7 189.94
## + X1sq    1        2.15 9062.7 191.31
## + X2sq    1        0.60 9064.2 191.31
## + X2      1        0.16 9064.7 191.32
##
## Step:  AIC=177.27
## Y ~ X1X2 + X3
##
##           Df Sum of Sq    RSS    AIC
## + X3sq    1     439.64 5482.4 176.72
## <none>                5922.0 177.27
## + X1      1    141.03 5781.0 178.47
## + X2X3    1    117.85 5804.2 178.60
## + X1X3    1     95.05 5827.0 178.73
## + X2      1     71.76 5850.3 178.87
## + X2sq    1     38.61 5883.4 179.05
## + X1sq    1     38.56 5883.5 179.05
##
## Step:  AIC=176.72
```

```
## Y ~ X1X2 + X3 + X3sq
##
##           Df Sum of Sq    RSS    AIC
## <none>                5482.4 176.72
## + X1      1    278.642 5203.8 177.00
## + X1X3    1    270.423 5212.0 177.05
## + X2X3    1     29.961 5452.4 178.54
## + X2      1      9.384 5473.0 178.67
## + X2sq    1      1.933 5480.5 178.71
## + X1sq    1      0.497 5481.9 178.72

fwd$coefficients

## (Intercept)          X1X2           X3          X3sq
## -16.20594947 -0.59715466  3.18911906 -0.01585003

# Do the results match up with best subset selection?
(d1 = names(fwd$coefficients)[-1]) # Names of predictor variables without the intercept

## [1] "X1X2" "X3"   "X3sq"

(minAIC = which.min(bs2$Subsets$AIC)) # Which model is the best for AIC

## [1] 6

# Names of predictor variables without intercept: Best Subset Selection
(d2 = names(bs2$Subsets[minAIC, bs2$Subsets[minAIC,] == TRUE])[-1])

## [1] "X1"   "X2"   "X3"   "X3sq" "X1X2"

# Check if the two names sets are equivalent:
d1 %in% d2

## [1] TRUE TRUE TRUE

d2 %in% d1

## [1] FALSE FALSE  TRUE  TRUE  TRUE
```

Using the Forward Stepwise selection approach, AIC identified 3 covariates, which are X_1X_2 , X_3 , X_3^2 . The AIC via best subset identified 5 covariates, which are X_1 , X_2 , X_3 , X_3^2 , X_1X_2 . The best subset selection suggests 2 more covariates compare to the Forward Stepwise selection method. The results are not matching.

```
# Forward selection with BIC (k = log(N)), StepAIC can do the same job for BIC
fwd2 <- stepAIC(min.model,
               direction = 'forward',
               scope = scp,
               k = log(nrow(kidney)))
```

```
## Start:  AIC=229.38
## Y ~ 1
##
##           Df Sum of Sq    RSS    AIC
## + X1X2    1    21930.7  9064.8 192.31
## + X1      1    19927.0 11068.5 198.90
## + X1sq    1    17484.3 13511.2 205.48
## + X2sq    1    14947.1 16048.4 211.16
## + X2      1    13825.7 17169.9 213.39
```

```
## + X1X3 1 10860.3 20135.2 218.65
## + X2X3 1 4343.3 26652.2 227.90
## + X3 1 3708.8 27286.7 228.68
## + X3sq 1 3382.6 27612.9 229.07
## <none> 30995.5 229.38
```

```
##
```

```
## Step: AIC=192.31
```

```
## Y ~ X1X2
```

```
##
```

```
##      Df Sum of Sq  RSS   AIC
## + X3  1  3142.79 5922.0 181.76
## + X3sq 1  2832.07 6232.8 183.44
## + X2X3 1  1305.75 7759.1 190.67
## <none>      9064.8 192.31
## + X1X3 1   826.88 8237.9 192.65
## + X1  1   371.15 8693.7 194.43
## + X1sq 1    2.15 9062.7 195.80
## + X2sq 1    0.60 9064.2 195.80
## + X2  1    0.16 9064.7 195.81
```

```
##
```

```
## Step: AIC=181.76
```

```
## Y ~ X1X2 + X3
```

```
##
```

```
##      Df Sum of Sq  RSS   AIC
## <none>      5922.0 181.76
## + X3sq 1   439.64 5482.4 182.71
## + X1  1   141.03 5781.0 184.46
## + X2X3 1   117.85 5804.2 184.59
## + X1X3 1    95.05 5827.0 184.72
## + X2  1    71.76 5850.3 184.85
## + X2sq 1    38.61 5883.4 185.04
## + X1sq 1    38.56 5883.5 185.04
```

```
fwd2$coefficients
```

```
## (Intercept)      X1X2      X3
## 73.6149703 -0.5964808 0.7648475
```

```
# Do the results match up with best subset selection?
```

```
(d3 <- names(fwd2$coefficients)[-1])
```

```
## [1] "X1X2" "X3"
```

```
(minBIC = which.min(bs1$Subsets$BIC))
```

```
## [1] 5
```

```
(d4 <- names(bs1$Subsets[minBIC, bs1$Subsets[minBIC,] == TRUE])[-1])
```

```
## [1] "X1" "X2" "X3" "X1X2"
```

```
# Check if the two names sets are equivalent:
```

```
d3 %in% d4
```

```
## [1] TRUE TRUE
```

```
d4 %in% d3
```

```
## [1] FALSE FALSE TRUE TRUE
```

Using the Forward Stepwise selection approach, BIC identified 2 covariates, which are X_1X_2, X_3 . The BIC via best subset identified 4 covariates, which are X_1, X_2, X_3, X_1X_2 . The best subset selection suggests 2 more covariates compare to the Forward Stepwise selection method. The results are also not matching.