Hw7_Ex_4

Manuel Alejandro Garcia Acosta

10/30/2019

```r
library(BioStatR)
library(olsrr)
```

```
##
## Attaching package: 'olsrr'
## The following object is masked from 'package:datasets':
##
##     rivers
```

Exercise 4 Homework 7

Read the data

```r
setwd('/home/noble_mannu/Documents/PhD/First/STAT_2131_Applied_Statistical_Methods_I/HW7')
steam_data <- read.table('steam_text.txt', header = TRUE)
```

Making the regression model

```r
m1 <- lm(steam ~ fat+glycerine+wind+frezday+temp, data = steam_data)
```

We display the summary of our model.

```r
summary(m1)
```

```
##
## Call:
## lm(formula = steam ~ fat + glycerine + wind + frezday + temp,
##     data = steam_data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.2682 -0.4438  0.1410  0.4043  1.2165
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  9.879818   2.165236   4.563 0.000213 ***
## fat          0.684238   0.542857   1.260 0.222772
## glycerine    0.407284   3.450937   0.118 0.907290
## wind         0.002434   0.102714   0.024 0.981343
## frezday     -0.007772   0.028595  -0.272 0.788712
## temp        -0.083786   0.018113  -4.626 0.000184 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

```
## Residual standard error: 0.6841 on 19 degrees of freedom
## Multiple R-squared:  0.8607, Adjusted R-squared:  0.824
## F-statistic: 23.47 on 5 and 19 DF,  p-value: 1.6e-07
```

Performing forward and backward model selection

Performing forward selection

```r
alpha.1 <- 0.1
forward.1 <- ols_step_forward_p(m1, penter = alpha.1)
```

```
## Forward Selection Method
## ---------------------------
##
## Candidate Terms:
##
## 1. fat
## 2. glycerine
## 3. wind
## 4. frezday
## 5. temp
##
## We are selecting variables based on p value...
##
## Variables Entered:
##
## - temp
## - fat
##
## No more variables to be added.
##
## Final Model Output
## ------------------
##
##                         Model Summary
## ---------------------------------------------------------------
## R                       0.927       RMSE                 0.637
## R-Squared               0.860       Coef. Var            6.761
## Adj. R-Squared          0.847       MSE                  0.406
## Pred R-Squared          0.826       MAE                  0.492
## ---------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                               ANOVA
## ---------------------------------------------------------------------
##               Sum of
##              Squares        DF    Mean Square      F         Sig.
## ---------------------------------------------------------------------
## Regression     54.884         2         27.442    67.597    0.0000
## Residual        8.931        22          0.406
```

```
## Total          63.816        24
## -----------------------------------------------------------------
##
##                                    Parameter Estimates
## ----------------------------------------------------------------------------------------
##       model       Beta    Std. Error    Std. Beta       t        Sig      lower      upper
## ----------------------------------------------------------------------------------------
## (Intercept)      9.474         0.962                  9.850    0.000      7.479     11.469
##        temp     -0.080         0.008       -0.845   -10.588    0.000     -0.095     -0.064
##         fat      0.762         0.159        0.382     4.784    0.000      0.431      1.092
## ----------------------------------------------------------------------------------------
```

```r
forward.1$predictors   # Model only contains temp and fat
```

```
## [1] "temp" "fat"
```

```r
# Our final model only has temp and fat as predictors
```

Performing backward selection

```r
alpha.2 <- 0.2
backward.1 <- ols_step_backward_p(m1, penter = alpha.2)
```

```
## Backward Elimination Method
## ---------------------------
##
## Candidate Terms:
##
## 1 . fat
## 2 . glycerine
## 3 . wind
## 4 . frezday
## 5 . temp
##
## We are eliminating variables based on p value...
##
## Variables Removed:
##
## - wind
## - glycerine
## - frezday
##
## No more variables satisfy the condition of p value = 0.3
##
##
## Final Model Output
## ------------------
##
##                         Model Summary
## --------------------------------------------------------------
## R                       0.927       RMSE                0.637
## R-Squared               0.860       Coef. Var           6.761
## Adj. R-Squared          0.847       MSE                 0.406
## Pred R-Squared          0.826       MAE                 0.492
```

```
## -------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                               ANOVA
## -------------------------------------------------------------------
##                Sum of
##               Squares        DF    Mean Square      F         Sig.
## -------------------------------------------------------------------
## Regression     54.884         2         27.442    67.597    0.0000
## Residual        8.931        22          0.406
## Total          63.816        24
## -------------------------------------------------------------------
##
##                                   Parameter Estimates
## ------------------------------------------------------------------------------------------
##       model      Beta    Std. Error    Std. Beta         t        Sig      lower     upper
## ------------------------------------------------------------------------------------------
## (Intercept)     9.474         0.962                   9.850    0.000      7.479    11.469
##         fat     0.762         0.159        0.382      4.784    0.000      0.431     1.092
##        temp    -0.080         0.008       -0.845    -10.588    0.000     -0.095    -0.064
## ------------------------------------------------------------------------------------------
```

```r
backward.1$removed    # Model only contains temp and fat
```

```
## [1] "wind"      "glycerine" "frezday"
```

```r
# Our final model only has temp and fat as predictors
```

Although we weren't required to do so, I performed forward step-wise selection just to see how it went.

Step-wise forward

```r
step.wise <- ols_step_both_p(m1, pent = alpha.1, prem = alpha.2)
```

```
## Stepwise Selection Method
## ---------------------------
##
## Candidate Terms:
##
## 1. fat
## 2. glycerine
## 3. wind
## 4. frezday
## 5. temp
##
## We are selecting variables based on p value...
##
## Variables Entered/Removed:
##
## - temp added
## - fat added
##
## No more variables to be added/removed.
```

```
##
##
## Final Model Output
## ------------------
##
##                         Model Summary
## -------------------------------------------------------------
## R                        0.927       RMSE              0.637
## R-Squared                0.860       Coef. Var         6.761
## Adj. R-Squared           0.847       MSE               0.406
## Pred R-Squared           0.826       MAE               0.492
## -------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                                ANOVA
## -------------------------------------------------------------------
##               Sum of
##              Squares        DF    Mean Square      F         Sig.
## -------------------------------------------------------------------
## Regression    54.884         2         27.442    67.597    0.0000
## Residual       8.931        22          0.406
## Total         63.816        24
## -------------------------------------------------------------------
##
##                                    Parameter Estimates
## -----------------------------------------------------------------------------------------
##       model      Beta    Std. Error    Std. Beta       t        Sig      lower     upper
## -----------------------------------------------------------------------------------------
## (Intercept)     9.474         0.962                  9.850    0.000      7.479    11.469
##        temp    -0.080         0.008       -0.845   -10.588    0.000     -0.095    -0.064
##         fat     0.762         0.159        0.382     4.784    0.000      0.431     1.092
## -----------------------------------------------------------------------------------------
```

```
step.wise$predictors  # Model only contains temp and fat
```

```
## [1] "temp" "fat"
```

```
# Once again our final model only has temp and fat as predictors
```

Now we'll run best subset selection for this model

Running best subset selection

```
Best.subset.steam <- olsrr::ols_step_best_subset(m1)
```

Choosing the model based on AIC

```
which.min(Best.subset.steam$aic)
```

```
## [1] 2
```

```r
# Returns row 2, this corresponds to the model with only fat and temp as predictors
# Prints the names of the predictors used in the best model with AIC criteria
Best.subset.steam$predictors[which.min(Best.subset.steam$aic)]
```

```
## [1] "fat temp"
```

```r
# This shows that the best model using AIC criteria is the one that has fat and temp as predictors
```

Choosing the model based on BIC

```r
which.min(Best.subset.steam$sbc)
```

```
## [1] 2
```

```r
# Returns row 2, this corresponds to the model with only fat and temp as predictors
# Prints the names of the predictors used in the best model with BIC criteria
Best.subset.steam$predictors[which.min(Best.subset.steam$sbc)]
```

```
## [1] "fat temp"
```

```r
# This shows that the best model using BIC criteria is the one that has fat and temp as predictors
```

We conclude that backward, forward, step-wise forward and AIC and BIC methods all returned back the model with only 'fat' and 'temp' as predictors.

Note: With adjusted R2 criteria we get the same model as above

```r
which.max(Best.subset.steam$adjr)
```

```
## [1] 2
```

```r
# Returns row 2, this corresponds to the model with only fat and temp as predictors
# Prints the names of the predictors used in the best model with adjusted R2 criteria
Best.subset.steam$predictors[which.max(Best.subset.steam$adjr)]
```

```
## [1] "fat temp"
```

Note: With R2 criteria we get the full model

```r
which.max(Best.subset.steam$rsquare)
```

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

```r
# Returns row 5, this corresponds to the full model
# Prints the names of the predictors used in the best model with R2 criteria
Best.subset.steam$predictors[which.max(Best.subset.steam$rsquare)]
```

```
## [1] "fat glycerine wind frezday temp"
```