

SelectionMethodsLinearRegression.R

vijaykalmath

2022-01-04

```
# Selection for Multiple Linear Regresion
```

```
library(MASS)
library(ISLR)
library(tidyverse)
```

```
## -- Attaching packages ----- tidyverse 1.3.1 --
```

```
## v ggplot2 3.3.5    v purrr  0.3.4
## v tibble  3.1.4    v dplyr  1.0.7
## v tidyr   1.1.3    v stringr 1.4.0
## v readr   2.0.1    v forcats 0.5.1
```

```
## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()     masks stats::lag()
## x dplyr::select() masks MASS::select()
```

```
Boston$lstat_pow5 = Boston$lstat^5
Boston$lstat_pow4 = Boston$lstat^4
Boston$lstat_pow3 = Boston$lstat^3
Boston$lstat_pow2 = Boston$lstat^2
```

```
mlr.fit = lm(medv ~ ., data = Boston)
```

```
summary(mlr.fit)
```

```
##
```

```
## Call:
```

```
## lm(formula = medv ~ ., data = Boston)
```

```
##
```

```
## Residuals:
```

```
##      Min       1Q   Median       3Q      Max
## -14.6753  -2.4250  -0.2148   1.7398  25.9904
```

```
##
```

```
## Coefficients:
```

```
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  6.491e+01  5.655e+00  11.479  < 2e-16 ***
## crim        -1.428e-01  2.886e-02  -4.946  1.04e-06 ***
## zn           1.324e-02  1.232e-02   1.075  0.282973
```

```
## indus      -5.338e-03  5.404e-02  -0.099  0.921359
## chas       1.944e+00  7.519e-01   2.586  0.009996 **
## nox       -1.563e+01  3.327e+00  -4.699  3.41e-06 ***
## rm        2.661e+00  3.937e-01   6.758  4.00e-11 ***
## age       2.251e-02  1.183e-02   1.904  0.057546 .
## dis      -1.138e+00  1.752e-01  -6.497  2.03e-10 ***
## rad       2.945e-01  5.761e-02   5.112  4.58e-07 ***
## tax      -1.113e-02  3.267e-03  -3.407  0.000712 ***
## ptratio   -8.409e-01  1.146e-01  -7.334  9.34e-13 ***
## black     8.091e-03  2.338e-03   3.461  0.000586 ***
## lstat     -9.026e+00  1.247e+00  -7.240  1.76e-12 ***
## lstat_pow5 -1.301e-05  3.561e-06  -3.654  0.000286 ***
## lstat_pow4  1.382e-03  3.359e-04   4.114  4.56e-05 ***
## lstat_pow3 -5.416e-02  1.161e-02  -4.663  4.02e-06 ***
## lstat_pow2  9.887e-01  1.809e-01   5.466  7.34e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.118 on 488 degrees of freedom
## Multiple R-squared:  0.8063, Adjusted R-squared:  0.7995
## F-statistic: 119.5 on 17 and 488 DF,  p-value: < 2.2e-16
```

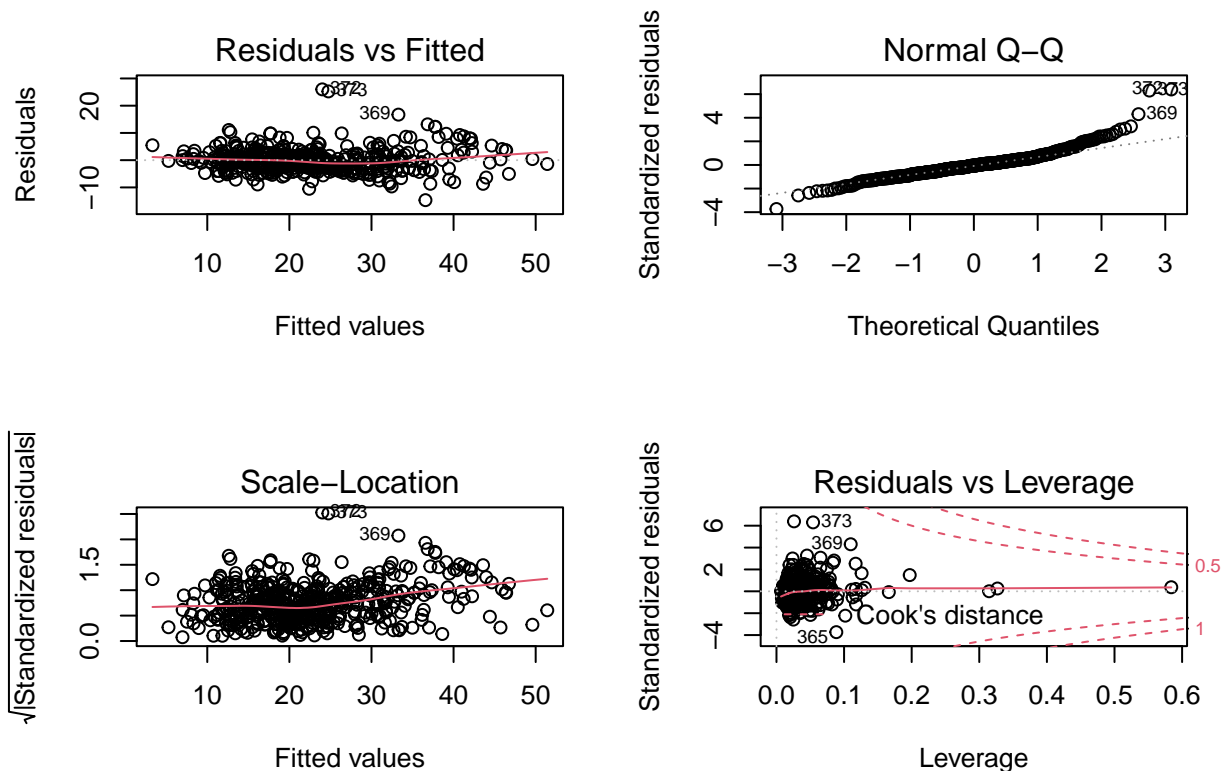
```
par(mfrow=c(2,2));plot(mlr.fit)
```

```
# Forward Stepwise regression model
step.model <- stepAIC(mlr.fit , direction = "forward",
                      trace = FALSE)
summary(step.model)
```

```
##
## Call:
## lm(formula = medv ~ crim + zn + indus + chas + nox + rm + age +
##      dis + rad + tax + ptratio + black + lstat + lstat_pow5 +
##      lstat_pow4 + lstat_pow3 + lstat_pow2, data = Boston)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -14.6753  -2.4250  -0.2148   1.7398  25.9904
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  6.491e+01  5.655e+00  11.479  < 2e-16 ***
## crim        -1.428e-01  2.886e-02  -4.946  1.04e-06 ***
## zn          1.324e-02  1.232e-02   1.075  0.282973
## indus       -5.338e-03  5.404e-02  -0.099  0.921359
## chas        1.944e+00  7.519e-01   2.586  0.009996 **
## nox       -1.563e+01  3.327e+00  -4.699  3.41e-06 ***
## rm         2.661e+00  3.937e-01   6.758  4.00e-11 ***
## age        2.251e-02  1.183e-02   1.904  0.057546 .
## dis      -1.138e+00  1.752e-01  -6.497  2.03e-10 ***
## rad       2.945e-01  5.761e-02   5.112  4.58e-07 ***
## tax      -1.113e-02  3.267e-03  -3.407  0.000712 ***
## ptratio   -8.409e-01  1.146e-01  -7.334  9.34e-13 ***
```

```
## black      8.091e-03  2.338e-03  3.461 0.000586 ***
## lstat      -9.026e+00  1.247e+00 -7.240 1.76e-12 ***
## lstat_pow5 -1.301e-05  3.561e-06 -3.654 0.000286 ***
## lstat_pow4  1.382e-03  3.359e-04  4.114 4.56e-05 ***
## lstat_pow3 -5.416e-02  1.161e-02 -4.663 4.02e-06 ***
## lstat_pow2  9.887e-01  1.809e-01  5.466 7.34e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.118 on 488 degrees of freedom
## Multiple R-squared:  0.8063, Adjusted R-squared:  0.7995
## F-statistic: 119.5 on 17 and 488 DF,  p-value: < 2.2e-16
```

```
par(mfrow=c(2,2));plot(step.model)
```



```
# Backward Stepwise regression model
step.model <- stepAIC(mlr.fit , direction = "backward",
                      trace = FALSE)
summary(step.model)

##
## Call:
## lm(formula = medv ~ crim + chas + nox + rm + age + dis + rad +
##     tax + ptratio + black + lstat + lstat_pow5 + lstat_pow4 +
##     lstat_pow3 + lstat_pow2, data = Boston)
```

```
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -14.7408  -2.4455  -0.2211   1.7028  26.1262
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  6.567e+01  5.599e+00  11.729 < 2e-16 ***
## crim        -1.407e-01  2.877e-02  -4.889 1.38e-06 ***
## chas         1.923e+00  7.483e-01   2.569 0.010481 *
## nox         -1.593e+01  3.204e+00  -4.972 9.16e-07 ***
## rm          2.692e+00  3.902e-01   6.899 1.62e-11 ***
## age         2.141e-02  1.177e-02   1.819 0.069565 .
## dis        -1.049e+00  1.527e-01  -6.873 1.92e-11 ***
## rad         2.913e-01  5.519e-02   5.279 1.96e-07 ***
## tax        -1.065e-02  2.893e-03  -3.681 0.000258 ***
## ptratio     -8.812e-01  1.081e-01  -8.155 2.95e-15 ***
## black        8.128e-03  2.335e-03   3.481 0.000544 ***
## lstat       -9.122e+00  1.232e+00  -7.402 5.87e-13 ***
## lstat_pow5  -1.293e-05  3.547e-06  -3.645 0.000296 ***
## lstat_pow4   1.378e-03  3.344e-04   4.120 4.45e-05 ***
## lstat_pow3  -5.421e-02  1.155e-02  -4.695 3.47e-06 ***
## lstat_pow2   9.947e-01  1.794e-01   5.543 4.87e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.115 on 490 degrees of freedom
## Multiple R-squared:  0.8058, Adjusted R-squared:  0.7998
## F-statistic: 135.5 on 15 and 490 DF, p-value: < 2.2e-16

par(mfrow=c(2,2));plot(step.model)
```

