StepAicModel.R

rocka

2023-12-10

```
#Linear Regression using Stepwise search for minimum AIC models
```

```
library(MASS)
library(dplyr)
```

```
##
## Attaching package: 'dplyr'

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

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
df<-read.csv("C:/Users/rocka/OneDrive/Documents/output_file.csv")

na_count <- colSums(is.na(df))
print(na_count)
```

```
##     Cement       Slag     FlyAsh      Water      SuPly CoarseAggr   FineAggr
##          0          0          0          0          0          0          0
##        Age        CCS
##          0          0
```

```
df <- rename(df, response=CCS)

SEED<-1234

# Search for minimum AIC model using all the data
modeltry <- lm(response ~ .^2, data = df)
length(coef(modeltry))
```

```
## [1] 37
```

```r
step.model <- stepAIC(modeltry, trace=0)
length(coef(step.model))
```

```
## [1] 30
```

```r
round(coef(step.model), 3)
```

```
##         (Intercept)              Cement                Slag             FlyAsh
##            -166.032               0.329              -0.091              -0.036
##               Water               SuPly          CoarseAggr            FineAggr
##               1.746              -2.847               0.072              -0.123
##                 Age         Cement:Slag       Cement:FlyAsh        Cement:Water
##              -0.635               0.000               0.000              -0.002
##        Cement:SuPly     Cement:FineAggr          Cement:Age         Slag:FlyAsh
##              -0.003               0.000               0.001               0.000
##          Slag:Water       Slag:FineAggr           Slag:Age        FlyAsh:Water
##              -0.001               0.000               0.001              -0.002
##        FlyAsh:SuPly     FlyAsh:FineAggr          FlyAsh:Age         Water:SuPly
##              -0.006               0.000               0.002               0.010
##    Water:CoarseAggr      Water:FineAggr    SuPly:CoarseAggr           SuPly:Age
##              -0.001               0.000               0.003               0.006
## CoarseAggr:FineAggr        FineAggr:Age
##               0.000               0.001
```

```r
AIC(modeltry, step.model)
```

```
##            df      AIC
## modeltry   38 7342.436
## step.model 31 7333.433
```

```r
# Validation-set approach to estimate MSE
# Create indices for training and test data
set.seed(SEED)
nobs <- nrow(df)
train <- sample(1:nobs, nobs/2)
test <- (-train)

# Computing test-set mean-sum-of-squares for use in computing
# test-set R-squared
mean <- mean(df$response[test])
MSS <- mean((df$response[test]-mean)^2)

#Searching for minimum AIC model using the training data
step.model <- stepAIC(lm(response~.^2, data=df[train,]),
                      trace=0)

#Using the test data to estimate MSE, RMSE, and test-set R-squared
truth <- df[test,]$response
pred <- predict(step.model, newdata=df[test,])
(MSE <- mean((truth-pred)^2))
```

```
## [1] 83.71395
```

```
(RMSE <- sqrt(MSE))
```

```
## [1] 9.149533
```

```
(R_sq <- 1 - MSE/MSS)
```

```
## [1] 0.6928072
```

```
#Doing a final build of the model using all of the data
names(step.model)
```

```
##  [1] "coefficients"  "residuals"     "effects"       "rank"
##  [5] "fitted.values" "assign"        "qr"            "df.residual"
##  [9] "xlevels"       "call"          "terms"         "model"
## [13] "anova"
```

```
model.final <- lm(step.model$model, data=df)

(coef <- round(coef(model.final), 5))
```

```
##       (Intercept)            Cement              Slag            FlyAsh
##         -307.73536           0.47072          -0.25695          -0.03691
##             Water             SuPly        CoarseAggr          FineAggr
##           1.65414           1.96411           0.26005          -0.04170
##               Age       Cement:Slag     Cement:FlyAsh      Cement:Water
##           -0.56812           0.00020           0.00029          -0.00157
##      Cement:SuPly Cement:CoarseAggr        Cement:Age       Slag:FlyAsh
##           -0.00496          -0.00009           0.00057           0.00044
##   Slag:CoarseAggr     Slag:FineAggr          Slag:Age      FlyAsh:Water
##           0.00003           0.00030           0.00080          -0.00153
##      FlyAsh:SuPly   FlyAsh:FineAggr        FlyAsh:Age  Water:CoarseAggr
##           -0.00680           0.00034           0.00180          -0.00125
##         SuPly:Age      FineAggr:Age
##           0.00659           0.00057
```

```
# number of effects (other than the intercept) in the final model
(length(coef) -1)
```

```
## [1] 25
```