PCR-Regression.R

vijaykalmath

2022-01-04

```r
# Principal Components Regression

# PCR is part of the pls library

library(pls)
```

```
##
## Attaching package: 'pls'

## The following object is masked from 'package:stats':
##
##     loadings
```

```r
library(ISLR)
# Ridge Regression and Lasso Regression

# GLMNET package is used for Lasso and Ridge Regression

# set seed for reproducibility
set.seed(1)

# Omitting NA
Hitters = na.omit(Hitters)

# Model Matrix produces matrix with 19 predictors but also transforms any qualitative variables into du
x = model.matrix(Salary~.,data=Hitters)[,-1]

y = Hitters$Salary


# Define Train and test set
train <- sample(1:nrow(x), nrow(x) / 2)
test <- (-train)

set.seed(2)

pcr.fit = pcr(Salary~.,data=Hitters,scale=TRUE,validation = "CV")

summary(pcr.fit)
```
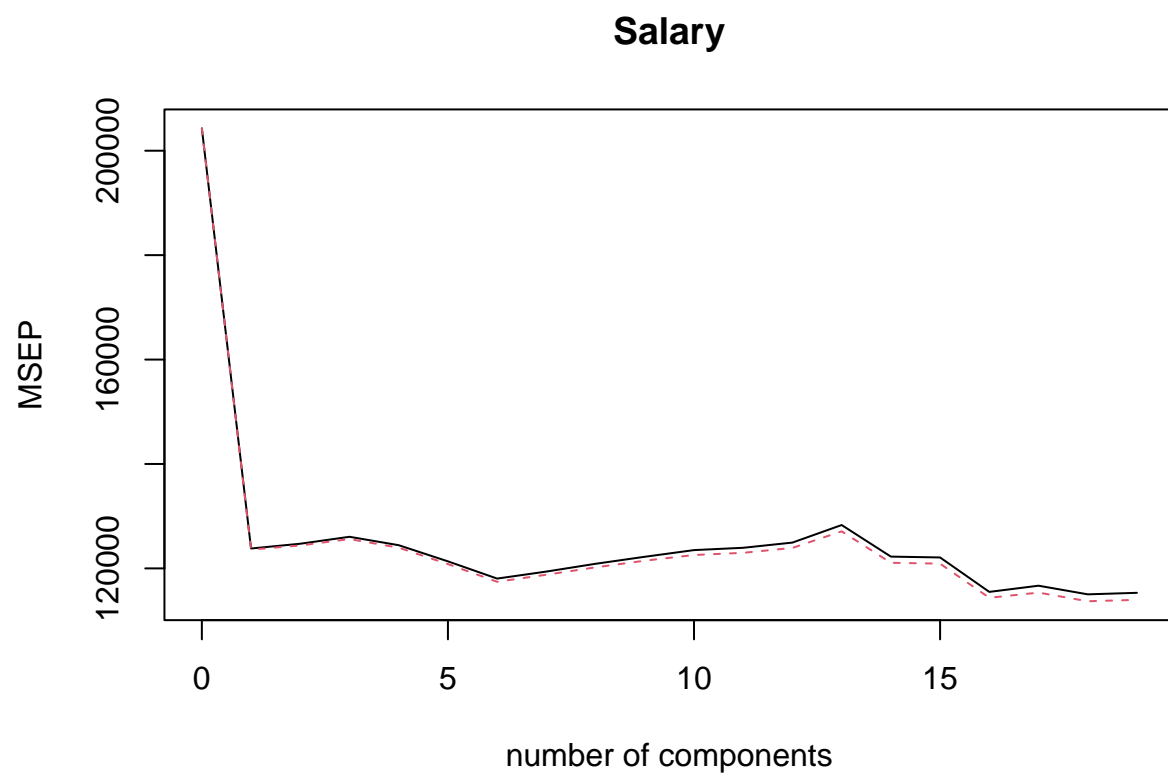
```
## Data:    X dimension: 263 19
```

```
##   Y dimension: 263 1
## Fit method: svdpc
## Number of components considered: 19
##
## VALIDATION: RMSEP
## Cross-validated using 10 random segments.
##        (Intercept)  1 comps  2 comps  3 comps  4 comps  5 comps  6 comps
## CV             452    351.9    353.2    355.0    352.8    348.4    343.6
## adjCV          452    351.6    352.7    354.4    352.1    347.6    342.7
##        7 comps  8 comps  9 comps  10 comps  11 comps  12 comps  13 comps
## CV       345.5    347.7    349.6     351.4     352.1     353.5     358.2
## adjCV    344.7    346.7    348.5     350.1     350.7     352.0     356.5
##        14 comps  15 comps  16 comps  17 comps  18 comps  19 comps
## CV        349.7     349.4     339.9     341.6     339.2     339.6
## adjCV     348.0     347.7     338.2     339.7     337.2     337.6
##
## TRAINING: % variance explained
##         1 comps  2 comps  3 comps  4 comps  5 comps  6 comps  7 comps  8 comps
## X         38.31    60.16    70.84    79.03    84.29    88.63    92.26    94.96
## Salary    40.63    41.58    42.17    43.22    44.90    46.48    46.69    46.75
##         9 comps  10 comps  11 comps  12 comps  13 comps  14 comps  15 comps
## X         96.28     97.26     97.98     98.65     99.15     99.47     99.75
## Salary    46.86     47.76     47.82     47.85     48.10     50.40     50.55
##         16 comps  17 comps  18 comps  19 comps
## X          99.89     99.97     99.99    100.00
## Salary     53.01     53.85     54.61     54.61
```
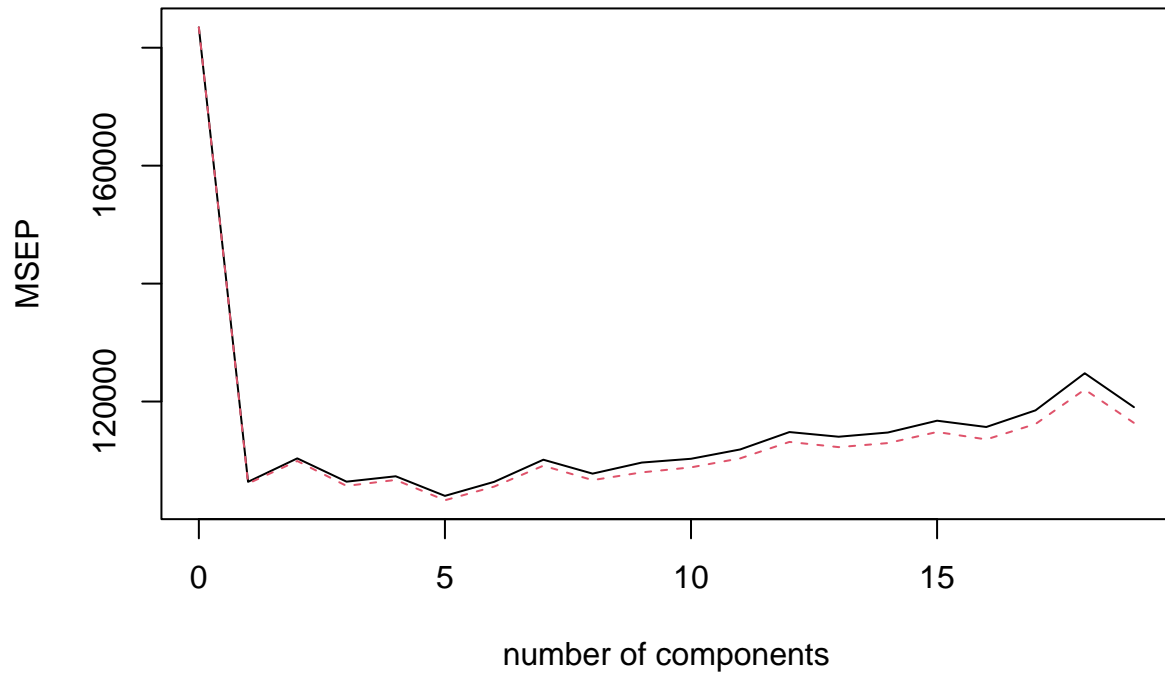
```
validationplot(pcr.fit,val.type="MSEP")
```

Salary


```r
set.seed(1)

pcr.fit = pcr(Salary~.,data=Hitters,subset=train,scale=TRUE,validation="CV")

validationplot(pcr.fit,val.type="MSEP")
```

Salary

MSEP vs number of components

```
pcr.pred = predict(pcr.fit,x[test,],ncomp=7)
mean((pcr.pred - y[test])^2)
```

```
## [1] 140751.3
```

```
pcr.fit= pcr(y~x,scale=TRUE,ncomp=7)
```

```
summary(pcr.fit)
```

```
## Data:    X dimension: 263 19
##  Y dimension: 263 1
## Fit method: svdpc
## Number of components considered: 7
## TRAINING: % variance explained
##    1 comps  2 comps  3 comps  4 comps  5 comps  6 comps  7 comps
## X    38.31    60.16    70.84    79.03    84.29    88.63    92.26
## y    40.63    41.58    42.17    43.22    44.90    46.48    46.69
```