

PLS-Regression.R

vijaykalmath

2022-01-04

```
# Principal Components Regression
```

```
# PCR is part of the pls library
```

```
library(pls)
```

```
##
```

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

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

```
##
```

```
##      loadings
```

```
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 dummy variables
```

```
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)
```

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

```
summary(pls.fit)
```

```
## Data:      X dimension: 131 19
```

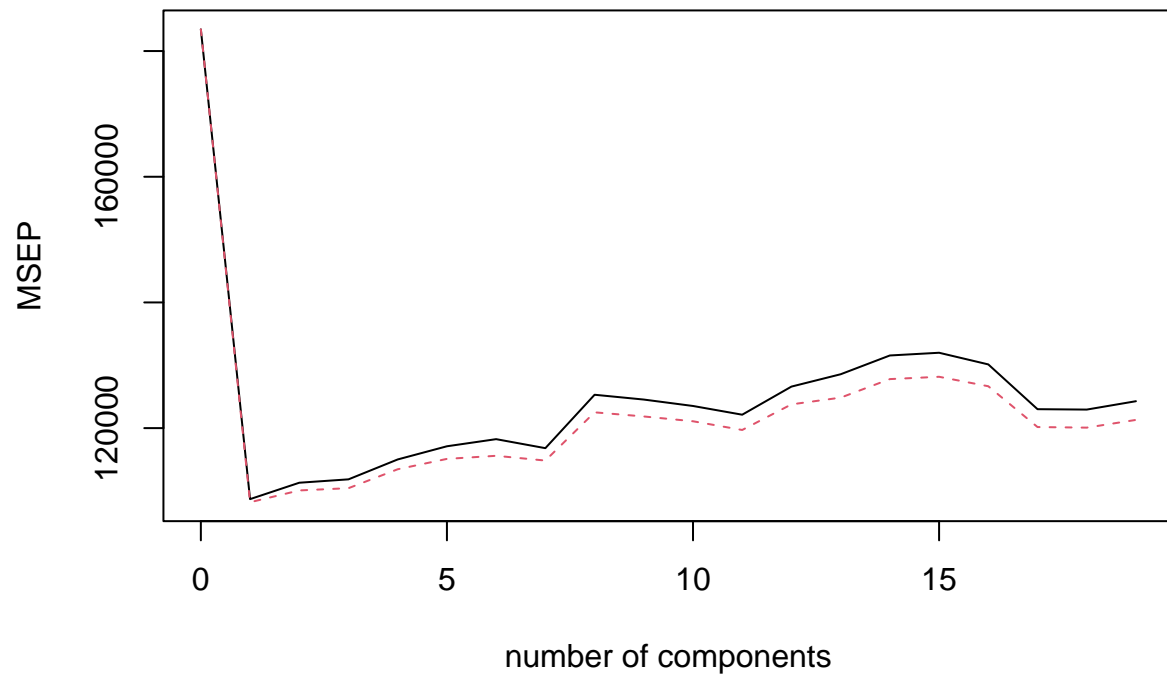
```

## Y dimension: 131 1
## Fit method: kernelpls
## 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           428.3   329.7   333.6   334.4   339.1   342.2   343.9
## adjCV        428.3   328.9   331.8   332.3   336.8   339.3   340.0
##      7 comps  8 comps  9 comps 10 comps 11 comps 12 comps 13 comps
## CV           341.8    354   352.9   351.5   349.5   355.8   358.6
## adjCV        338.9    350   349.1   348.0   346.0   351.8   353.3
##      14 comps 15 comps 16 comps 17 comps 18 comps 19 comps
## CV           362.7   363.3   360.8   350.7   350.6   352.5
## adjCV        357.5   358.0   355.9   346.7   346.5   348.3
##
## TRAINING: % variance explained
##      1 comps  2 comps  3 comps  4 comps  5 comps  6 comps  7 comps  8 comps
## X           39.13   48.80   60.09   75.07   78.58   81.12   88.21   90.71
## Salary      46.36   50.72   52.23   53.03   54.07   54.77   55.05   55.66
##      9 comps 10 comps 11 comps 12 comps 13 comps 14 comps 15 comps
## X           93.17   96.05   97.08   97.61   97.97   98.70   99.12
## Salary      55.95   56.12   56.47   56.68   57.37   57.76   58.08
##      16 comps 17 comps 18 comps 19 comps
## X           99.61   99.70   99.95   100.00
## Salary      58.17   58.49   58.56   58.62

```

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

Salary



```
pls.pred = predict(pls.fit,x[test,],ncomp = 2)
```

```
mean((pls.pred - y[test])^2)
```

```
## [1] 145367.7
```

```
pls.fit = plsrf(Salary~.,data=Hitters,scale=TRUE,ncomp=2)
```

```
summary(pls.fit)
```

```
## Data:      X dimension: 263 19
## Y dimension: 263 1
## Fit method: kernelppls
## Number of components considered: 2
## TRAINING: % variance explained
##           1 comps  2 comps
## X           38.08   51.03
## Salary      43.05   46.40
```