Lab 11

Daniel Tshiani

2025-06-08

1

```
library(pls)
```

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

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

```
load("../data/Auto-3.rda")
```

a

```
pcr.fit = pcr(mpg~.-name-origin+as.factor(origin),data=Auto)

lm.fit <- lm(mpg~.-name-origin+as.factor(origin),data=Auto)
```

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

```
## Data:    X dimension: 392 8
##  Y dimension: 392 1
## Fit method: svdpc
## Number of components considered: 8
## TRAINING: % variance explained
##      1 comps  2 comps  3 comps  4 comps  5 comps  6 comps  7 comps  8 comps
## X      99.76    99.96   100.00   100.00   100.00   100.00   100.00   100.00
## mpg    69.35    70.09    70.75    80.79    80.88    80.91    80.93    82.42
```

```
summary(lm.fit)
```

```
##
## Call:
## lm(formula = mpg ~ . - name - origin + as.factor(origin), data = Auto)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -9.0095 -2.0785 -0.0982  1.9856 13.3608
##
## Coefficients:
##                       Estimate Std. Error t value Pr(>|t|)
## (Intercept)        -1.795e+01  4.677e+00  -3.839 0.000145 ***
## cylinders          -4.897e-01  3.212e-01  -1.524 0.128215
```

```
## displacement        2.398e-02  7.653e-03   3.133 0.001863 **
## horsepower         -1.818e-02  1.371e-02  -1.326 0.185488
## weight             -6.710e-03  6.551e-04 -10.243  < 2e-16 ***
## acceleration        7.910e-02  9.822e-02   0.805 0.421101
## year                7.770e-01  5.178e-02  15.005  < 2e-16 ***
## as.factor(origin)2  2.630e+00  5.664e-01   4.643 4.72e-06 ***
## as.factor(origin)3  2.853e+00  5.527e-01   5.162 3.93e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.307 on 383 degrees of freedom
## Multiple R-squared:  0.8242, Adjusted R-squared:  0.8205
## F-statistic: 224.5 on 8 and 383 DF,  p-value: < 2.2e-16
```

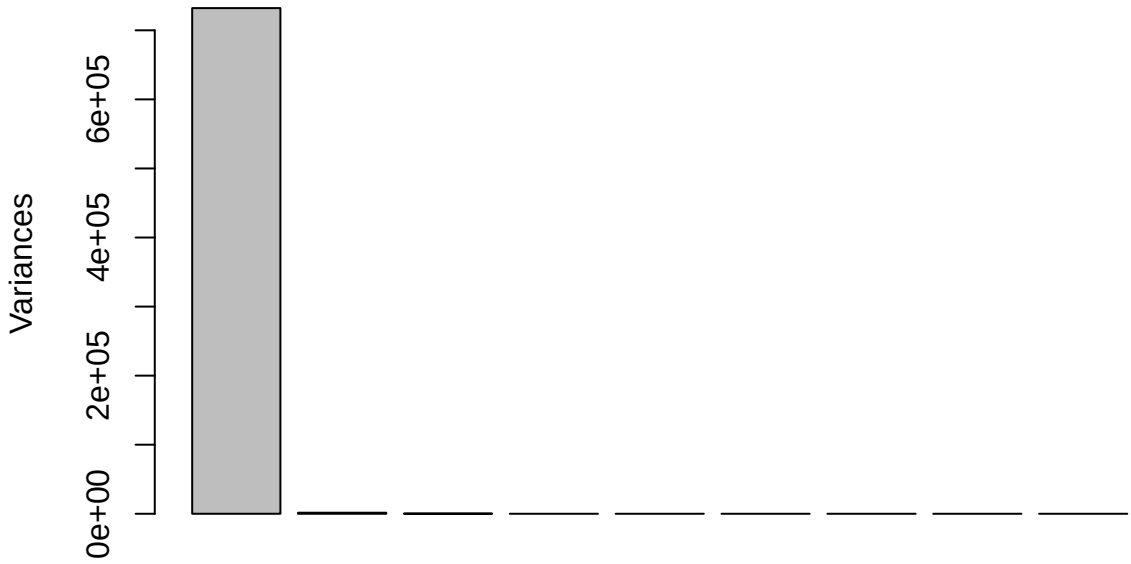
with OLS, about 82% of variation is explained in the model. in PCA, about 82% of variation is explained in the model with 8 predictors.

b

```
X <-model.matrix(lm.fit)
library(stats)
pc <- prcomp(X)
pc
```

```
## Standard deviations (1, .., p=9):
## [1] 855.6585163  38.9097121  16.1620689   3.3135262   1.6966834   0.5249057
## [7]   0.4167494   0.2446327   0.0000000
##
## Rotation (n x k) = (9 x 9):
##                             PC1           PC2          PC3          PC4
## (Intercept)         0.0000000000  0.0000000000  0.000000000  0.000000000
## cylinders          -0.0017926225  0.0133245279  0.007294275 -0.001414710
## displacement       -0.1143412856  0.9457785881  0.303312504  0.009143349
## horsepower         -0.0389670412  0.2982553337 -0.948761071  0.043076559
## weight             -0.9926735354 -0.1207516411  0.002454212 -0.001480458
## acceleration        0.0013528348 -0.0348264293  0.077006895 -0.059516278
## year                0.0013368415 -0.0238516081  0.042819254  0.996935229
## as.factor(origin)2  0.0001308250 -0.0024889942 -0.002857670 -0.022100094
## as.factor(origin)3  0.0002103564 -0.0003765828 -0.004796684  0.012089823
##                             PC5           PC6           PC7           PC8 PC9
## (Intercept)         0.000000e+00  0.0000000000  0.0000000000  0.000000e+00   1
## cylinders           1.719368e-02 -0.9911554803  0.1211162208  4.909265e-02   0
## displacement       -1.059355e-02  0.0146594359 -0.0006512752 -4.394368e-03   0
## horsepower         -8.646402e-02 -0.0038232742  0.0034425206  4.435100e-03   0
## weight              3.152970e-03  0.0002093216 -0.0003053766 -5.729471e-06   0
## acceleration       -9.944974e-01 -0.0168319859  0.0012233398  1.799780e-03   0
## year               -5.549653e-02  0.0001647840  0.0240346554 -7.643176e-03   0
## as.factor(origin)2 -9.052576e-05  0.0483462982  0.6888706846 -7.229226e-01   0
## as.factor(origin)3 -1.150938e-03 -0.1214929883 -0.7142804151 -6.891098e-01   0
```

```
screeplot(pc)
```




```
summary(pc)
```

```
## Importance of components:
##                             PC1      PC2      PC3     PC4   PC5    PC6    PC7
## Standard deviation     855.6585 38.90971 16.16207 3.31353 1.697 0.5249 0.4167
## Proportion of Variance   0.9976  0.00206  0.00036 0.00001 0.000 0.0000 0.0000
## Cumulative Proportion    0.9976  0.99962  0.99998 1.00000 1.000 1.0000 1.0000
##                            PC8 PC9
## Standard deviation     0.2446   0
## Proportion of Variance 0.0000   0
## Cumulative Proportion  1.0000   1
```

c

```
pc_s <- prcomp(X[,-1], scale = TRUE)
pc_s
```

```
## Standard deviations (1, .., p=8):
## [1] 2.1334996 1.1355990 0.9496774 0.8046002 0.6267811 0.3540064 0.2423474
## [8] 0.1800515
##
## Rotation (n x k) = (8 x 8):
##                              PC1         PC2        PC3         PC4         PC5
## cylinders           -0.4418880 -0.03183661  0.1513759  0.08112451 -0.17557152
## displacement        -0.4568839 -0.02468263  0.1334604  0.05932358 -0.09729606
## horsepower          -0.4427613 -0.04007041 -0.1065626 -0.15129036 -0.20758380
## weight              -0.4343646  0.02796626  0.2382679  0.10749689 -0.31123269
## acceleration         0.3038014  0.14422734  0.4574875  0.70453506 -0.30582654
## year                 0.2153593 -0.27979875  0.7122813 -0.58611507 -0.13985800
## as.factor(origin)2   0.1648074  0.71513286 -0.1544922 -0.33648172 -0.55320193
## as.factor(origin)3   0.2131795 -0.62087485 -0.3881227  0.04831876 -0.63295244
##                             PC6          PC7         PC8
## cylinders           -0.714889530  0.219309935  0.42922755
```

```
## displacement       -0.157462297  0.046158763 -0.85611061
## horsepower          0.571133678  0.610447413  0.16180753
## weight              0.318542871 -0.705651981  0.21433158
## acceleration        0.127257355  0.266640825 -0.02042327
## year                0.007589381  0.067111659 -0.01832329
## as.factor(origin)2 -0.114101557 -0.006045755 -0.07244053
## as.factor(origin)3 -0.084877251 -0.059530689 -0.06868871
```

d

```
set.seed(1234)
pcr_reg <- pcr(mpg~.-name-origin+as.factor(origin),data=Auto, scale=TRUE, validation = "CV")
summary(pcr_reg)
```

```
## Data:    X dimension: 392 8
##  Y dimension: 392 1
## Fit method: svdpc
## Number of components considered: 8
##
## VALIDATION: RMSEP
## Cross-validated using 10 random segments.
##        (Intercept)  1 comps  2 comps  3 comps  4 comps  5 comps  6 comps
## CV           7.815    4.163    4.037    4.038    3.620    3.625    3.541
## adjCV        7.815    4.162    4.034    4.036    3.617    3.621    3.537
##        7 comps  8 comps
## CV       3.427    3.359
## adjCV    3.422    3.354
##
## TRAINING: % variance explained
##      1 comps  2 comps  3 comps  4 comps  5 comps  6 comps  7 comps  8 comps
## X       56.9    73.02    84.29    92.38    97.29    98.86    99.59   100.00
## mpg     71.8    73.64    73.96    79.25    79.25    80.22    81.55    82.42
```

2

```
plsr.fit = plsr(mpg~.-name-origin+as.factor(origin),data=Auto)

summary(plsr.fit)
```

```
## Data:    X dimension: 392 8
##  Y dimension: 392 1
## Fit method: kernelpls
## Number of components considered: 8
## TRAINING: % variance explained
##      1 comps  2 comps  3 comps  4 comps  5 comps  6 comps  7 comps  8 comps
## X      99.76    99.96   100.00   100.00   100.00   100.00   100.00   100.00
## mpg    69.35    70.48    72.06    80.85    80.99    82.16    82.41    82.42
```

```
pls_reg <- plsr(mpg~.-name-origin+as.factor(origin),data=Auto, scale = T, validation = "CV")

summary(pls_reg)
```

```
## Data:    X dimension: 392 8
##  Y dimension: 392 1
```

```
## Fit method: kernelpls
## Number of components considered: 8
##
## VALIDATION: RMSEP
## Cross-validated using 10 random segments.
##        (Intercept)  1 comps  2 comps  3 comps  4 comps  5 comps  6 comps
## CV           7.815    3.981    3.614    3.541    3.418    3.402    3.355
## adjCV        7.815    3.980    3.611    3.537    3.413    3.396    3.348
##        7 comps  8 comps
## CV       3.365    3.357
## adjCV    3.360    3.353
##
## TRAINING: % variance explained
##      1 comps  2 comps  3 comps  4 comps  5 comps  6 comps  7 comps  8 comps
## X      56.73    68.84    80.75    84.08    93.48    94.88    99.33   100.00
## mpg    74.32    79.37    80.29    81.71    82.00    82.35    82.38    82.42
```

the model improves up until about 4 or 5 components, then it stays around the 82% of explaination in variation mark.

```
validationplot(pls_reg)
```