Lab 7

Daniel Tshiani

2025-05-26

a

```
library(ISLR)
library(dplyr)
```

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

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

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

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

b

```
Auto <- Auto %>%
  mutate(Economy = case_when(
    mpg <= 17 ~ "Heavy",
    mpg <= 22.75 ~ "OK",
    mpg <= 29 ~ "Eco",
    mpg > 29 ~ "Excellent"
  )) %>%
  mutate(Economy = as.factor(Economy)) %>%
  mutate(origin = as.factor(origin))
```

c

```
library(MASS)
```

```
##
## Attaching package: 'MASS'
```

```
## The following object is masked from 'package:dplyr':
##
##     select
```

```r
table(Auto$Economy)
```

```
##
##       Eco Excellent     Heavy        OK
##       101        95        99        97
```

```r
lda_result <- lda(Economy ~ mpg + cylinders + displacement + weight + acceleration + year + origin, data
lda_result
```

```
## Call:
## lda(Economy ~ mpg + cylinders + displacement + weight + acceleration +
##     year + origin, data = Auto)
##
## Prior probabilities of groups:
##       Eco Excellent     Heavy        OK
## 0.2576531 0.2423469 0.2525510 0.2474490
##
## Group means:
##                mpg cylinders displacement   weight acceleration     year
## Eco       25.88911  4.277228     128.7178 2501.069     16.35248 76.03960
## Excellent 34.30842  4.073684     101.7895 2157.958     16.65053 79.17895
## Heavy     14.26869  7.717172     335.7879 4114.384     13.37071 73.38384
## OK        19.62990  5.793814     209.2371 3116.237     15.82577 75.43299
##              origin2    origin3
## Eco       0.31683168 0.20792079
## Excellent 0.23157895 0.51578947
## Heavy     0.03030303 0.00000000
## OK        0.11340206 0.09278351
##
## Coefficients of linear discriminants:
##                          LD1          LD2          LD3
## mpg          -0.3389905364 -0.264405943 -0.103651807
## cylinders     0.2317355612 -0.422308212  1.163393044
## displacement  0.0018855775 -0.005475355 -0.004392613
## weight        0.0003569182 -0.001252238 -0.001962502
## acceleration  0.0300076370  0.061115872  0.180788046
## year         -0.0378684709  0.085925497  0.194580341
## origin2       0.1702613411 -0.399663658 -0.729557654
## origin3      -0.1481699449 -1.137048668  0.234903830
##
## Proportion of trace:
##    LD1    LD2    LD3
## 0.9406 0.0525 0.0069
```

Heavy economy cars tend to have more cylinders, more displacement, higher weight, and lower mpg. Excellent and Eco economy cars tend to be lighter and more fuel efficient.

d

```r
library(MASS)

lda_cv <- lda(Economy ~ mpg + cylinders + displacement + weight + acceleration + year + origin, data = A

conf_matrix <- table(True = Auto$Economy, Predicted = lda_cv$class)
```

```r
print(conf_matrix)
```

```
##            Predicted
## True        Eco Excellent Heavy OK
##   Eco        95         0     0  6
##   Excellent  13        82     0  0
##   Heavy       0         0    85 14
##   OK         11         0     8 78
```

```r
accuracy <- mean(Auto$Economy == lda_cv$class)
print(paste("Classification accuracy (CV):", round(accuracy * 100, 2), "%"))
```

```
## [1] "Classification accuracy (CV): 86.73 %"
```

e

```r
lda_cv_prior <- lda(Economy ~ mpg + cylinders + displacement + weight + acceleration + year + origin,
                    data = Auto,
                    prior = c(0.25, 0.25, 0.25, 0.25),
                    CV = TRUE)

conf_matrix_prior <- table(True = Auto$Economy, Predicted = lda_cv_prior$class)
print(conf_matrix_prior)
```

```
##            Predicted
## True        Eco Excellent Heavy OK
##   Eco        94         0     0  7
##   Excellent  12        83     0  0
##   Heavy       0         0    85 14
##   OK         11         0     7 79
```

```r
accuracy_prior <- mean(Auto$Economy == lda_cv_prior$class)
print(paste("Classification accuracy with custom priors:", round(accuracy_prior * 100, 2), "%"))
```

```
## [1] "Classification accuracy with custom priors: 86.99 %"
```

the classification accuracy slightly decreased.

f

```r
lda_cv_prior2 <- lda(Economy ~ mpg + cylinders + displacement + weight + acceleration + year + origin,
                    data = Auto,
                    prior = c(0.4,0.3,0.2,0.1),
                    CV = TRUE)

conf_matrix_prior2 <- table(True = Auto$Economy, Predicted = lda_cv_prior2$class)
print(conf_matrix_prior2)
```

```
##            Predicted
## True        Eco Excellent Heavy OK
##   Eco        98         0     0  3
##   Excellent  14        81     0  0
##   Heavy       0         0    88 11
##   OK         17         0     9 71
```

```r
accuracy_prior2 <- mean(Auto$Economy == lda_cv_prior2$class)
print(paste("Classification accuracy with custom priors:", round(accuracy_prior2 * 100, 2), "%"))
```

```
## [1] "Classification accuracy with custom priors: 86.22 %"
```

the classification accuracy slightly decreased again.