ch7-2

Harry Woo

2020 6 16

rm(list = ls())  
library(MASS)  
library(klaR)  
library(dplyr)  
library(gt)

city <- read.csv("ex7-2.csv")  
head(city) %>% gt()

type

population

pollution

ind

11

48

ind

8

20

ind

12

25

ind

13

32

ind

6

42

ind

19

25

summary(city)

## type population pollution   
## Length:30 Min. : 4.00 Min. :18.00   
## Class :character 1st Qu.:13.25 1st Qu.:25.00   
## Mode :character Median :24.00 Median :27.50   
## Mean :21.87 Mean :30.37   
## 3rd Qu.:28.75 3rd Qu.:34.75   
## Max. :45.00 Max. :48.00

str(city)

## 'data.frame': 30 obs. of 3 variables:  
## $ type : chr "ind" "ind" "ind" "ind" ...  
## $ population: int 11 8 12 13 6 19 21 30 18 25 ...  
## $ pollution : int 48 20 25 32 42 25 43 24 35 27 ...

city\_lda <- lda(type ~ ., data = city)  
city\_lda #그룹이 2개이므로 1개 선형판별함수 출력

## Call:  
## lda(type ~ ., data = city)  
##   
## Prior probabilities of groups:  
## ind mer   
## 0.5 0.5   
##   
## Group means:  
## population pollution  
## ind 19.60000 31.86667  
## mer 24.13333 28.86667  
##   
## Coefficients of linear discriminants:  
## LD1  
## population 0.07528885  
## pollution -0.05751264

pred\_lda <- predict(city\_lda, newdata = city)  
names(pred\_lda)

## [1] "class" "posterior" "x"

head(pred\_lda$class)

## [1] ind ind ind ind ind mer  
## Levels: ind mer

head(pred\_lda$posterior) %>% gt()

ind

mer

0.7194049

0.2805951

0.5572716

0.4427284

0.5555477

0.4444523

0.5965942

0.4034058

0.7226470

0.2773530

0.4880780

0.5119220

head(pred\_lda$x) %>% gt()

LD1

-1.83227833

-0.44779106

-0.43419884

-0.76149845

-1.86364676

0.09282311

confu\_mat <- table(city$type, pred\_lda$class)  
confu\_mat

##   
## ind mer  
## ind 9 6  
## mer 6 9

error <- 1 - sum(diag(confu\_mat))/sum(confu\_mat)  
error

## [1] 0.4

# 변수선택 가능? 안되는데? 안되는게 정상이라고.

# city\_forward <- greedy.wilks(type ~ ., data = city)  
# ?greedy.wilks

# 분류함수  
source("classfunc.r")  
X <- city[, -1]  
classfunc\_result <- classfunc.lda(X, city$type)  
classfunc\_result$class.func %>% gt()

ind

mer

-12.7537645

-12.702302

0.3272460

0.365933

0.5556657

0.526113

classfunc\_result

## Call:  
## lda(groups ~ ., data = as.data.frame(x))  
##   
## Prior probabilities of groups:  
## ind mer   
## 0.5 0.5   
##   
## Group means:  
## population pollution  
## ind 19.60000 31.86667  
## mer 24.13333 28.86667  
##   
## Coefficients of linear discriminants:  
## LD1  
## population 0.07528885  
## pollution -0.05751264

X2 <- X %>%   
 mutate(ind = -12.7537645 + 0.3272460 \* population + 0.5556657 \* pollution,  
 mer = -12.702302 + 0.365933 \* population + 0.526113 \* pollution) %>%   
 mutate(type = city$type, type2 = ifelse(ind > mer, "ind", "mer"))  
  
sum(as.data.frame(pred\_lda$class) == X2["type2"])/nrow(X2)

## [1] 1

head(X2) %>% gt()

population

pollution

ind

mer

type

type2

11

48

17.5178951

16.576385

ind

ind

8

20

0.9775175

0.747422

ind

ind

12

25

5.0648300

4.841719

ind

ind

13

32

9.2817359

8.890443

ind

ind

6

42

12.5476709

11.590042

ind

ind

19

25

7.3555520

7.403250

ind

mer

X2

## population pollution ind mer type type2  
## 1 11 48 17.5178951 16.576385 ind ind  
## 2 8 20 0.9775175 0.747422 ind ind  
## 3 12 25 5.0648300 4.841719 ind ind  
## 4 13 32 9.2817359 8.890443 ind ind  
## 5 6 42 12.5476709 11.590042 ind ind  
## 6 19 25 7.3555520 7.403250 ind mer  
## 7 21 43 18.0120266 17.605150 ind ind  
## 8 30 24 10.3995923 10.902400 ind mer  
## 9 18 35 12.5849630 12.298447 ind ind  
## 10 25 27 10.4303594 10.651074 ind mer  
## 11 20 43 17.6847806 17.239217 ind ind  
## 12 24 26 9.5474477 9.759028 ind mer  
## 13 29 27 11.7393434 12.114806 ind mer  
## 14 28 18 6.4111061 7.013856 ind mer  
## 15 30 43 20.9572406 20.898547 ind ind  
## 16 9 30 6.8614205 6.374485 mer ind  
## 17 14 24 5.1636563 5.047472 mer ind  
## 18 17 28 8.3680571 8.249723 mer ind  
## 19 10 45 15.5236520 14.632113 mer ind  
## 20 4 27 3.5581934 2.966481 mer ind  
## 21 24 40 17.3267675 17.124610 mer ind  
## 22 25 25 9.3190280 9.598848 mer mer  
## 23 26 30 12.4246025 12.595346 mer mer  
## 24 31 29 13.5051668 13.898898 mer mer  
## 25 45 20 13.0856195 14.286943 mer mer  
## 26 37 24 12.6903143 13.463931 mer mer  
## 27 26 26 10.2019397 10.490894 mer mer  
## 28 26 32 13.5359339 13.647572 mer mer  
## 29 31 19 7.9485098 8.637768 mer mer  
## 30 37 34 18.2469713 18.725061 mer mer

X2["type"] == X2["type2"]

## type  
## [1,] TRUE  
## [2,] TRUE  
## [3,] TRUE  
## [4,] TRUE  
## [5,] TRUE  
## [6,] FALSE  
## [7,] TRUE  
## [8,] FALSE  
## [9,] TRUE  
## [10,] FALSE  
## [11,] TRUE  
## [12,] FALSE  
## [13,] FALSE  
## [14,] FALSE  
## [15,] TRUE  
## [16,] FALSE  
## [17,] FALSE  
## [18,] FALSE  
## [19,] FALSE  
## [20,] FALSE  
## [21,] FALSE  
## [22,] TRUE  
## [23,] TRUE  
## [24,] TRUE  
## [25,] TRUE  
## [26,] TRUE  
## [27,] TRUE  
## [28,] TRUE  
## [29,] TRUE  
## [30,] TRUE

confu\_mat2 <- table(X2$type, X2$type2)  
confu\_mat2

##   
## ind mer  
## ind 9 6  
## mer 6 9

error <- 1 - sum(diag(confu\_mat2))/sum(confu\_mat2)  
error

## [1] 0.4