

다변량자료분석 및 실습 Homework 4

서울대학교 통계학과 2017-11362 박건도

2021년 12월 01일

Problem 3

(c)

```
# data preparation
new.iris <- iris %>% filter(Species != 'setosa')
species <- as.factor(as.vector(new.iris[,5]))
new.iris <- new.iris[,-5]

# simple estimators for parameters
mu1 <- colMeans(new.iris[species == 'versicolor',])
mu2 <- colMeans(new.iris[species == 'virginica',])
S1 <- cov(new.iris[species == 'versicolor',])
S2 <- cov(new.iris[species == 'virginica',])
pi1 <- (species == 'versicolor') / length(species)
pi2 <- 1 - pi1

# classification
pi1_X <- pi1 * dmvnorm(new.iris, mean = mu1, sigma = S1) # P(X, Label = Versicolor)
pi2_X <- pi2 * dmvnorm(new.iris, mean = mu2, sigma = S2) # P(X, Label = Virginica)
bayes <- pi1_X / pi2_X > pi2 / (10 * pi1)
original_bayes <- pi1_X / pi2_X > pi2 / pi1

predicted <- ifelse(bayes, 'versicolor', 'virginica')
table(predicted, species)
```

##	species		
## predicted	versicolor	virginica	
## versicolor	26	0	
## virginica	24	50	

```
# original Bayes classifier (c(p/n)/c(n/p) = 1)
original_predicted <- ifelse(original_bayes, 'versicolor', 'virginica')
table(original_predicted, species)
```

```
##                species
## original_predicted versicolor virginica
##                versicolor      18         0
##                virginica      32         50
```

주어진 loss table을 이용하면, True Versicolor일 때 predicted Verginica의 loss가 10이다. 1을 찾을 때 보다 True Versicolor, predicted Verginica의 수가 줄어듦을 알 수 있다.

```
pairs(new.iris, col = ifelse(predicted == 'versicolor', 'red', 'blue'))
```

