HW4-LianZuo

Lian Zuo

2023-11-13

# 1.

Box's M-test for Homogeneity of Covariance Matrices  
  
data: new.pengu[, 2:5]  
Chi-Sq (approx.) = 76.795, df = 20, p-value = 1.365e-08

Conclusion:The test statistic value is found to be 76.795 with a p-value = 1.365e-08. Therefore, we reject the null hypothesis and conclude that the population covariance matrices are not significantly different,so the data shows evidence of heterogeneity of covariance matrices at 5% level of significance (p-value = 1.365e-08).

# 2.

Conclusion: Based on the conclusion in part 1, we can apply quadratic discriminant analysis instead of linear discriminant analysis for the data.

# 3.

a. Split the data into training set and testing set in the ratio 80:20. Use seed as 1234.

set.seed(1234)  
index=sample(1:nrow(new.pengu), floor(0.8\*nrow(new.pengu)), replace=F)  
train.data=new.pengu[index,]  
head(train.data,3)

test.data=new.pengu[-index,]  
head(test.data,3)

b. Scale the variables

library(tidyverse)   
train.trs=data.frame(scale(select(train.data, where(is.numeric))))   
test.trs=data.frame(scale(select(test.data, where(is.numeric))))  
train.trs$Species=train.data$species  
test.trs$Species=test.data$species

c. Perform the appropriate discriminant analysis.

qda.pengu=MASS::qda(Species~., data=train.trs)  
qda.pengu

Call:  
qda(Species ~ ., data = train.trs)  
  
Prior probabilities of groups:  
 Adelie Chinstrap Gentoo   
0.4688645 0.1978022 0.3333333   
  
Group means:  
 bill\_length\_mm bill\_depth\_mm flipper\_length\_mm body\_mass\_g  
Adelie -0.8944932 0.5662161 -0.7547519 -0.5838012  
Chinstrap 0.9421106 0.5867969 -0.2805355 -0.5468856  
Gentoo 0.6991336 -1.1446449 1.2281006 1.1456964

d.

accuracy

[1] 0.9275362

Comment: The model efficacy is found to be about 92.75%, which is reasonably a high percentage.

e.

Confusion Matrix and Statistics  
  
 Reference  
Prediction Adelie Chinstrap Gentoo  
 Adelie 23 5 0  
 Chinstrap 0 9 0  
 Gentoo 0 0 32

comment: The model has an overall accuracy of 92.75%, indicating good predictive performance. Adelie and Gentoo penguins are well-predicted, with high counts on the diagonal. However, Chinstrap penguins are occasionally confused with Adelie, as indicated by 5 counts in the cell.

f.

Comment: Obs.1 belongs to Gentoo group and Obs.2 belongs to Adelie group, respectively with posterior probability of 1 or very close 1.