Homework9

Yi Chen

4/17/2020

## Homework 9

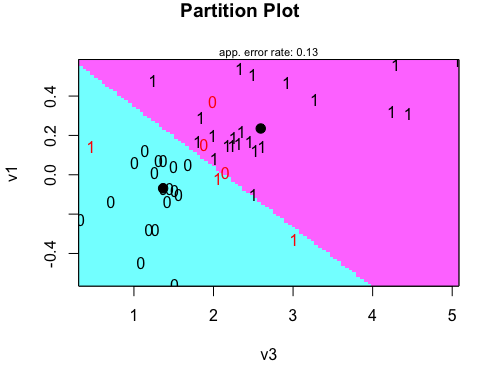
data <- read.table("T11-4.DAT")  
colnames(data) <- c("v1","v2","v3","v4","population")

#### question 1

library(MASS)  
library(klaR)  
data.lda = lda(population~v1+v3, data= data)  
data.lda

## Call:  
## lda(population ~ v1 + v3, data = data)  
##   
## Prior probabilities of groups:  
## 0 1   
## 0.4565217 0.5434783   
##   
## Group means:  
## v1 v3  
## 0 -0.06904762 1.366667  
## 1 0.23520000 2.593600  
##   
## Coefficients of linear discriminants:  
## LD1  
## v1 2.6645609  
## v3 0.8156689

data$population <- as.factor(data$population)  
partimat(population~v1+v3, data= data,method="lda")



#### question 2

data.lda.values = predict(data.lda)  
table(Predicted= data.lda.values$class, Actual= data$population)

## Actual  
## Predicted 0 1  
## 0 18 3  
## 1 3 22

mean(data$population==data.lda.values$class)

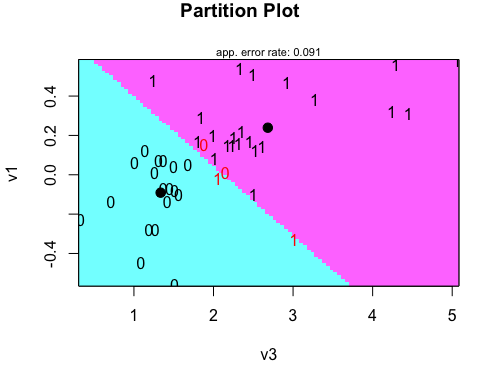
## [1] 0.8695652

#### question 3

data\_copy <- data[c(-16,-34),]  
data.lda = lda(population~v1+v3, data= data\_copy)  
data.lda

## Call:  
## lda(population ~ v1 + v3, data = data\_copy)  
##   
## Prior probabilities of groups:  
## 0 1   
## 0.4545455 0.5454545   
##   
## Group means:  
## v1 v3  
## 0 -0.0910000 1.3355  
## 1 0.2391667 2.6825  
##   
## Coefficients of linear discriminants:  
## LD1  
## v1 2.620905  
## v3 0.926678

data\_copy$population <- as.factor(data\_copy$population)  
partimat(population~v1+v3, data= data\_copy,method="lda")



#### question 4

data.lda.values = predict(data.lda)  
table(Predicted= data.lda.values$class, Actual= data\_copy$population)

## Actual  
## Predicted 0 1  
## 0 18 2  
## 1 2 22

mean(data\_copy$population==data.lda.values$class)

## [1] 0.9090909

#### question 5

data.qda = qda(population~v1+v3, data= data)  
data.qda

## Call:  
## qda(population ~ v1 + v3, data = data)  
##   
## Prior probabilities of groups:  
## 0 1   
## 0.4565217 0.5434783   
##   
## Group means:  
## v1 v3  
## 0 -0.06904762 1.366667  
## 1 0.23520000 2.593600

data.qda.values = predict(data.qda)  
table(Predicted= data.qda.values$class, Actual= data$population)

## Actual  
## Predicted 0 1  
## 0 19 1  
## 1 2 24

mean(data$population==data.qda.values$class)

## [1] 0.9347826

#### question 6

Conclusions:

1. Removing the outlier (influential points) help to improve the classifcation accuracy.
2. Removing the equal variance assumption (using qudratic discrimination analysis) improve the classifcation accuracy.
3. The overall accuracy for all models are acceptable.