Hw6_201511646_나여영

1. The following data were recorded for two species of irises. The objective is to develop a rule for classifying a new flower based on the four variables sl (Sepal Length), sw (Sepal Width), pl (Petal Length), pw (Petal Width). Provide necessary computer output to solve the following problems 1) and 2).

R에 내장되어 있는 Iris 데이터를 재구성하여 문제와 같이 복원했다.

```
newiris=iris[51:150,]

sl<-newiris$Sepal.Length*10
sw<-newiris$Sepal.Width*10
pl<-newiris$Petal.Length*10
pw<-newiris$Petal.Width*10
sp<-newiris$Species

niris=data.frame(sl,sw,pl,pw,sp)
niris
```

```
##     sl sw pl pw         sp
## 1   70 32 47 14 versicolor
## 2   64 32 45 15 versicolor
## 3   69 31 49 15 versicolor
## 4   55 23 40 13 versicolor
## 5   65 28 46 15 versicolor
## 6   57 28 45 13 versicolor
## 7   63 33 47 16 versicolor
## 8   49 24 33 10 versicolor
## 9   66 29 46 13 versicolor
## 10  52 27 39 14 versicolor
## 11  50 20 35 10 versicolor
## 12  59 30 42 15 versicolor
## 13  60 22 40 10 versicolor
## 14  61 29 47 14 versicolor
## 15  56 29 36 13 versicolor
## 16  67 31 44 14 versicolor
## 17  56 30 45 15 versicolor
## 18  58 27 41 10 versicolor
## 19  62 22 45 15 versicolor
## 20  56 25 39 11 versicolor
## 21  59 32 48 18 versicolor
## 22  61 28 40 13 versicolor
## 23  63 25 49 15 versicolor
## 24  61 28 47 12 versicolor
## 25  64 29 43 13 versicolor
## 26  66 30 44 14 versicolor
## 27  68 28 48 14 versicolor
## 28  67 30 50 17 versicolor
## 29  60 29 45 15 versicolor
## 30  57 26 35 10 versicolor
## 31  55 24 38 11 versicolor
## 32  55 24 37 10 versicolor
## 33  58 27 39 12 versicolor
## 34  60 27 51 16 versicolor
## 35  54 30 45 15 versicolor
## 36  60 34 45 16 versicolor
## 37  67 31 47 15 versicolor
## 38  63 23 44 13 versicolor
## 39  56 30 41 13 versicolor
## 40  55 25 40 13 versicolor
## 41  55 26 44 12 versicolor
## 42  61 30 46 14 versicolor
## 43  58 26 40 12 versicolor
## 44  50 23 33 10 versicolor
## 45  56 27 42 13 versicolor
## 46  57 30 42 12 versicolor
## 47  57 29 42 13 versicolor
## 48  62 29 43 13 versicolor
## 49  51 25 30 11 versicolor
## 50  57 28 41 13 versicolor
## 51  63 33 60 25  virginica
## 52  58 27 51 19  virginica
## 53  71 30 59 21  virginica
## 54  63 29 56 18  virginica
## 55  65 30 58 22  virginica
## 56  76 30 66 21  virginica
```

```
## 57  49 25 45 17  virginica
## 58  73 29 63 18  virginica
## 59  67 25 58 18  virginica
## 60  72 36 61 25  virginica
## 61  65 32 51 20  virginica
## 62  64 27 53 19  virginica
## 63  68 30 55 21  virginica
## 64  57 25 50 20  virginica
## 65  58 28 51 24  virginica
## 66  64 32 53 23  virginica
## 67  65 30 55 18  virginica
## 68  77 38 67 22  virginica
## 69  77 26 69 23  virginica
## 70  60 22 50 15  virginica
## 71  69 32 57 23  virginica
## 72  56 28 49 20  virginica
## 73  77 28 67 20  virginica
## 74  63 27 49 18  virginica
## 75  67 33 57 21  virginica
## 76  72 32 60 18  virginica
## 77  62 28 48 18  virginica
## 78  61 30 49 18  virginica
## 79  64 28 56 21  virginica
## 80  72 30 58 16  virginica
## 81  74 28 61 19  virginica
## 82  79 38 64 20  virginica
## 83  64 28 56 22  virginica
## 84  63 28 51 15  virginica
## 85  61 26 56 14  virginica
## 86  77 30 61 23  virginica
## 87  63 34 56 24  virginica
## 88  64 31 55 18  virginica
## 89  60 30 48 18  virginica
## 90  69 31 54 21  virginica
## 91  67 31 56 24  virginica
## 92  69 31 51 23  virginica
## 93  58 27 51 19  virginica
## 94  68 32 59 23  virginica
## 95  67 33 57 25  virginica
## 96  67 30 52 23  virginica
## 97  63 25 50 19  virginica
## 98  65 30 52 20  virginica
## 99  62 34 54 23  virginica
## 100 59 30 51 18  virginica
```

1. Develop such a rule using the proc discrim. Try linear and quadratic discriminant functions.

LDA

```
library(MASS)
```

```
## Warning: package 'MASS' was built under R version 3.4.2
```

```
index<-sample(2,size=nrow(niris), replace=T, prob=c(0.7,0.3))
train<-niris[index==1,]
test<-niris[index==2,]

#LDA
flda<-lda(sp~.,data=train)
```

```
## Warning in lda.default(x, grouping, ...): group setosa is empty
```

```
fts1<-predict(flda, newdata=test)
tab1 <- table(test$sp, fts1$class)
mclda <- 1-sum(diag(tab1))/sum(tab1)
flda
```

```
## Call:
## lda(sp ~ ., data = train)
##
## Prior probabilities of groups:
## versicolor  virginica
##  0.4776119  0.5223881
##
## Group means:
##                  sl       sw       pl       pw
## versicolor 59.71875 27.59375 42.43750 13.37500
## virginica  66.05714 30.00000 55.42857 20.28571
##
## Coefficients of linear discriminants:
##            LD1
## sl -0.12069859
## sw -0.08773144
## pl  0.19812021
## pw  0.29858376
```

```
tab1
```

```
##
##             setosa versicolor virginica
##   setosa         0          0         0
##   versicolor     0         16         2
##   virginica      0          0        15
```

```
mclda
```

```
## [1] 0.06060606
```

QDA

```
sp=as.numeric(newiris$Species)
niris=data.frame(sl,sw,pl,pw,sp)
index<-sample(2,size=nrow(niris), replace=T, prob=c(0.7,0.3))
train<-niris[index==1,]
test<-niris[index==2,]

fqda <- qda(sp~.,data=train)
fts2 <- predict(fqda,newdata=test)
tab2 <- table(test$sp,fts2$class)
mcqda <- 1-sum(diag(tab2))/sum(tab2)
fqda
```

```
## Call:
## qda(sp ~ ., data = train)
##
## Prior probabilities of groups:
##         2         3
## 0.4864865 0.5135135
##
## Group means:
##         sl       sw       pl       pw
## 2 59.38889 27.69444 42.72222 13.22222
## 3 66.21053 29.92105 55.65789 19.94737
```

```
tab2
```

```
##
##      2  3
##   2 13  1
##   3  1 11
```

```
mcqda
```

```
## [1] 0.07692308
```

2. A new iris is discovered use your results to classify the new species into one of these two groups. The new data are sl=60, sw=25, pl=40, pw=11. Use linear discriminant functions with c(1|2)=c(2|1) and P1=P2.

```
newd=data.frame(sl=60,sw=25,pl=40,pw=11)
predict(flda, newdata = newd)
```

```
## $class
## [1] versicolor
## Levels: setosa versicolor virginica
##
## $posterior
##   versicolor    virginica
## 1  0.9999653 3.474131e-05
##
## $x
##         LD1
## 1 -2.910953
```