LDA_QDA_Example

STAT380

2023-10-12

Loading the Iris Data

```
### Example 1: Iris data
library(MASS)
library(ggplot2)
data("iris")
str(iris)
```

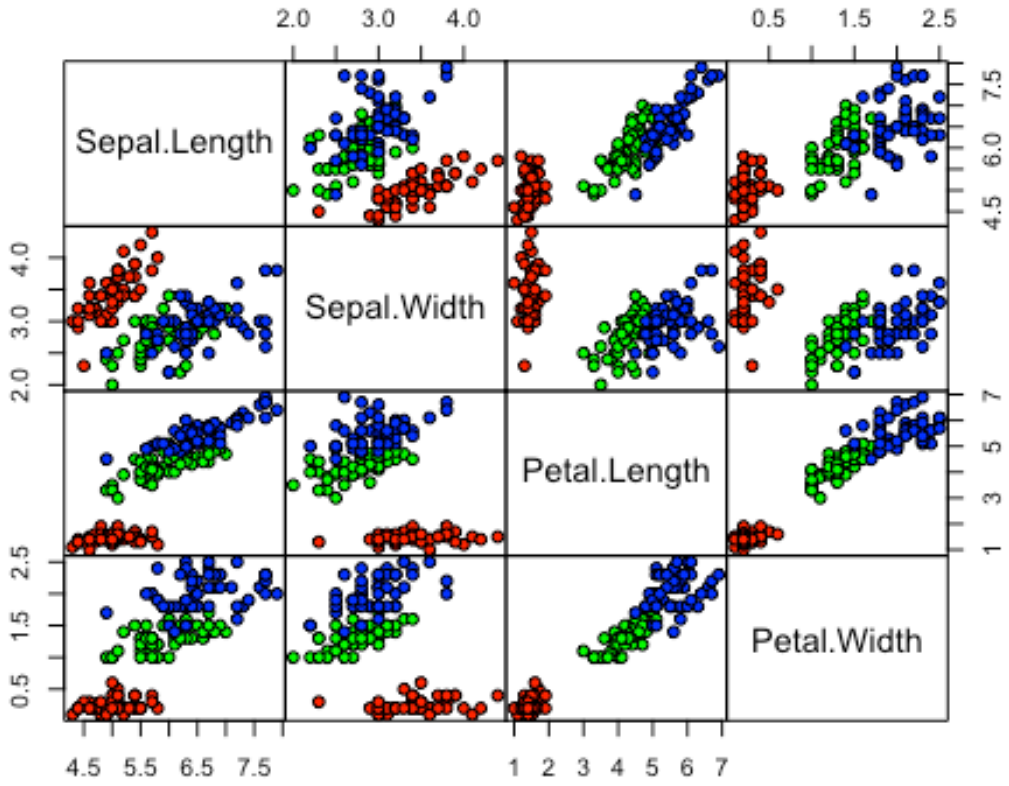
```
## 'data.frame':    150 obs. of  5 variables:
##  $ Sepal.Length: num  5.1 4.9 4.7 4.6 5 5.4 4.6 5 4.4 4.9 ...
##  $ Sepal.Width : num  3.5 3 3.2 3.1 3.6 3.9 3.4 3.4 2.9 3.1 ...
##  $ Petal.Length: num  1.4 1.4 1.3 1.5 1.4 1.7 1.4 1.5 1.4 1.5 ...
##  $ Petal.Width : num  0.2 0.2 0.2 0.2 0.2 0.4 0.3 0.2 0.2 0.1 ...
##  $ Species     : Factor w/ 3 levels "setosa","versicolor",..: 1 1 1 1 1 1
1 1 1 1 ...
```

```
head(iris)
```

```
##   Sepal.Length Sepal.Width Petal.Length Petal.Width Species
## 1          5.1         3.5          1.4         0.2  setosa
## 2          4.9         3.0          1.4         0.2  setosa
## 3          4.7         3.2          1.3         0.2  setosa
## 4          4.6         3.1          1.5         0.2  setosa
## 5          5.0         3.6          1.4         0.2  setosa
## 6          5.4         3.9          1.7         0.4  setosa
```

Basic plots to see the interrelation between the variables

```
pairs(iris[1:4],
              gap = 0,
              bg = c("red", "green", "blue")[iris$Species],
              pch = 21)
```



Splitting the Data in Training and Testing Set

```
set.seed(134)
ind = sample(2, nrow(iris), replace = TRUE, prob = c(0.6, 0.4))
training = iris[ind==1,]
testing = iris[ind==2,]
```

Fitting a Linear Discriminant Analysis

```
iris_lda = lda(Species~., training)
iris_lda

## Call:
## lda(Species ~ ., data = training)
##
## Prior probabilities of groups:
##     setosa versicolor  virginica
##  0.3367347  0.3469388  0.3163265
##
## Group means:
##            Sepal.Length Sepal.Width Petal.Length Petal.Width
## setosa         5.006061    3.457576     1.439394   0.2575758
```

```
## versicolor     5.932353    2.735294     4.223529   1.3147059
## virginica      6.438710    2.935484     5.445161   1.9774194
##
## Coefficients of linear discriminants:
##                     LD1        LD2
## Sepal.Length  0.9583651 -0.6656007
## Sepal.Width   1.1953550  2.4214894
## Petal.Length -2.6930964 -0.4043851
## Petal.Width  -2.1933913  2.4288629
##
## Proportion of trace:
##    LD1    LD2
## 0.9914 0.0086
```

```
attributes(iris_lda); ##or
```

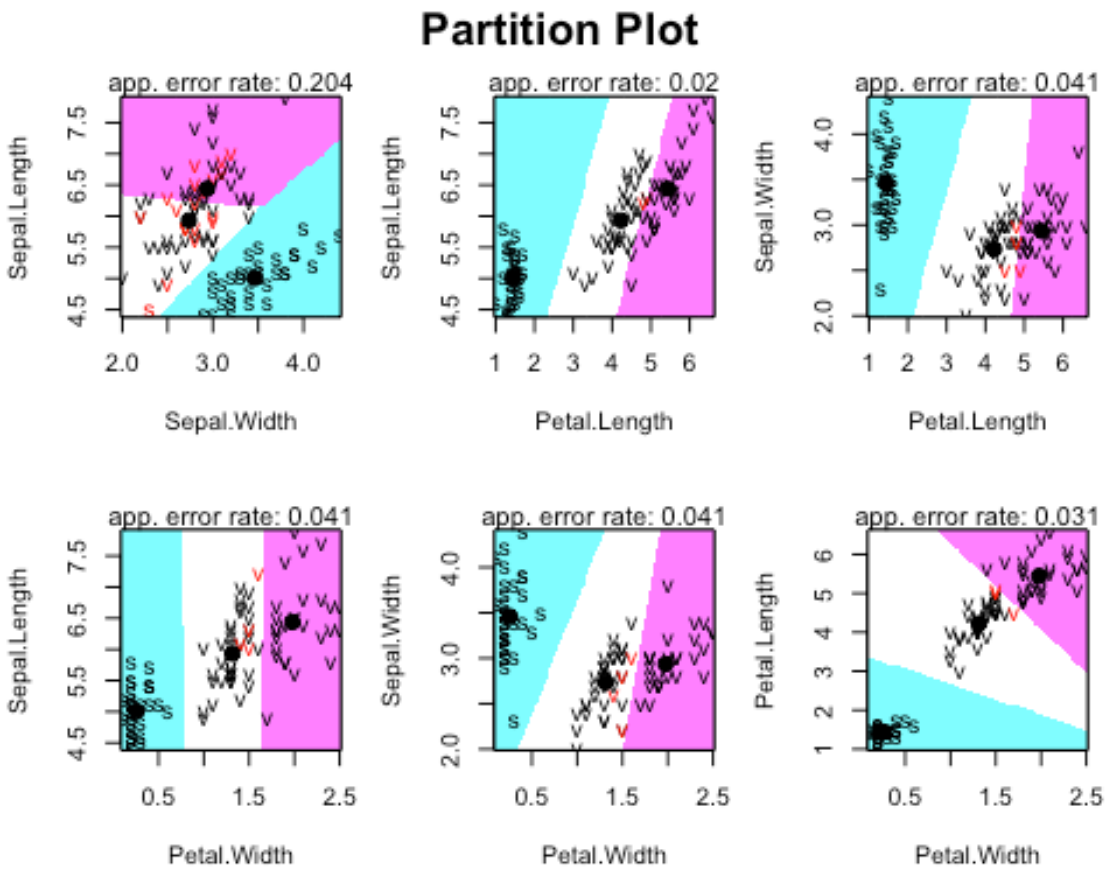
```
## $names
##  [1] "prior"   "counts"  "means"   "scaling" "lev"     "svd"     "N"
##  [8] "call"    "terms"   "xlevels"
##
## $class
## [1] "lda"
```

```
names(iris_lda)
```

```
##  [1] "prior"   "counts"  "means"   "scaling" "lev"     "svd"     "N"
##  [8] "call"    "terms"   "xlevels"
```

Predicting the classes (In Training Set) based on the LDA fit

```
p = predict(iris_lda, training)
library(klaR) # for the function `partimat'
partimat(Species~., data = training, method = "lda")
```




```
### Confusion matrix and accuracy - training data
p1 = predict(iris_lda, training)$class
tab = table(Predicted = p1, Actual = training$Species)
tab
```

```
##             Actual
## Predicted    setosa versicolor virginica
##   setosa         33          0         0
##   versicolor      0         34         0
##   virginica       0          0        31
```

Predicting the classes (In Testing Set) based on the LDA fit

```
p2 = predict(iris_lda, testing)$class
tab1 = table(Predicted = p2, Actual = testing$Species)
tab1
```

```
##             Actual
## Predicted    setosa versicolor virginica
##   setosa         17          0         0
##   versicolor      0         14         0
##   virginica       0          2        19
```

QDA (Quadratic discriminant analysis on the IRIS Data)

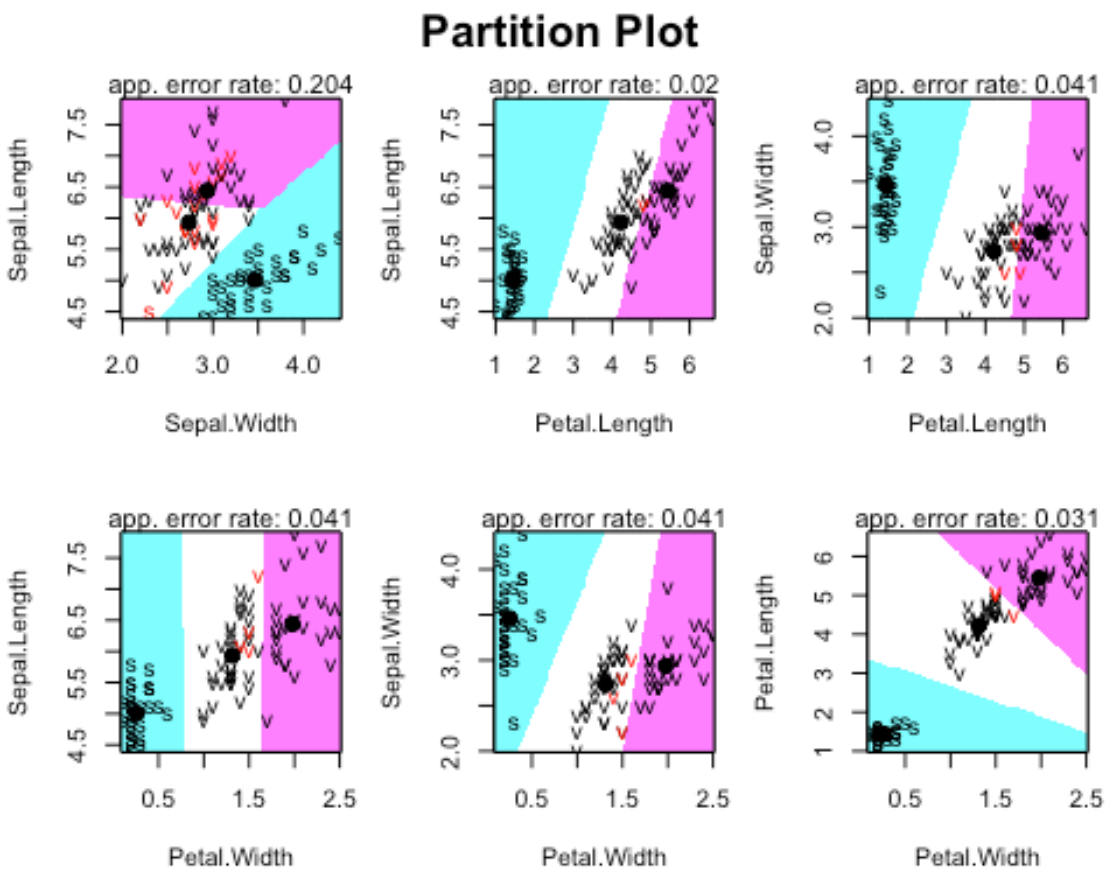
```
##Everything is not linear - quadratic discriminant analysis

iris_qda=qda(Species~.,data=training)
iris_qda
```

```
## Call:
## qda(Species ~ ., data = training)
##
## Prior probabilities of groups:
##     setosa versicolor  virginica
##  0.3367347  0.3469388  0.3163265
##
## Group means:
##            Sepal.Length Sepal.Width Petal.Length Petal.Width
## setosa         5.006061    3.457576     1.439394   0.2575758
## versicolor     5.932353    2.735294     4.223529   1.3147059
## virginica      6.438710    2.935484     5.445161   1.9774194
```

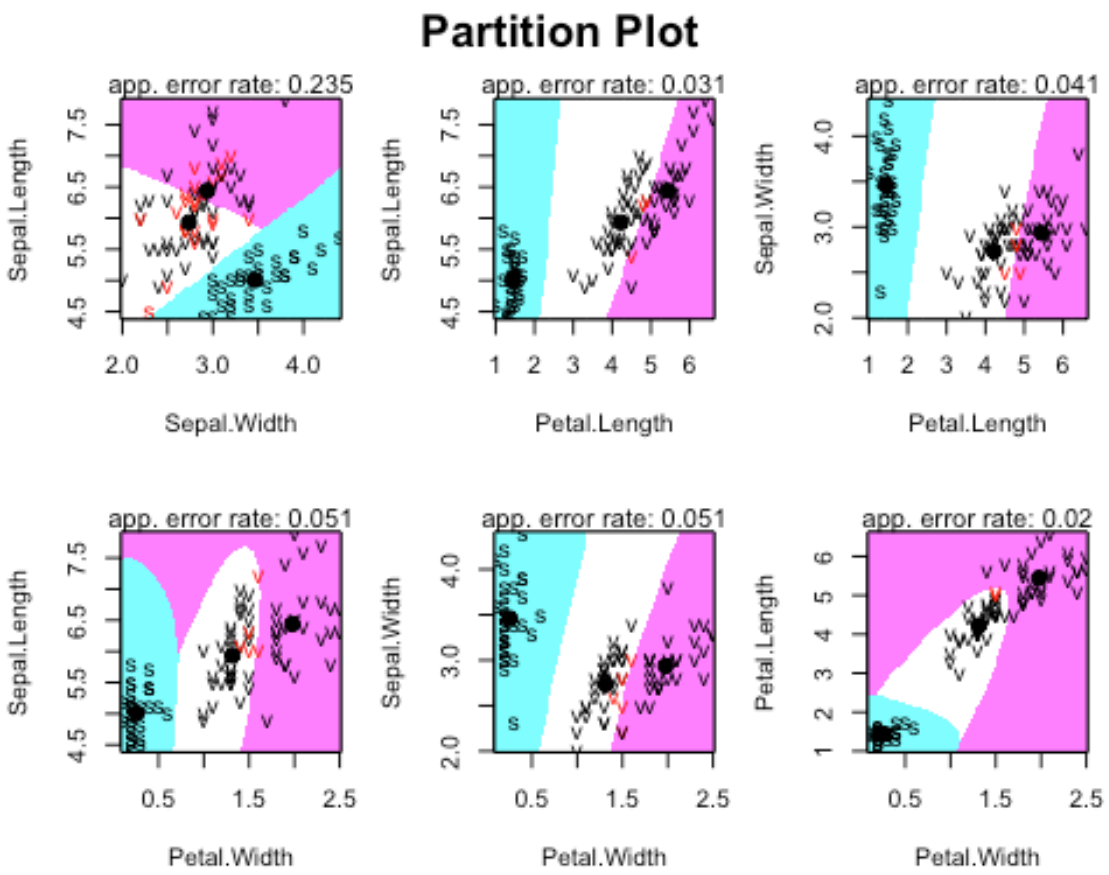
```
summary(iris_qda)
```

```
##         Length Class  Mode
## prior    3     -none- numeric
## counts   3     -none- numeric
## means   12     -none- numeric
## scaling 48     -none- numeric
## ldet     3     -none- numeric
## lev      3     -none- character
## N        1     -none- numeric
## call     3     -none- call
## terms    3     terms  call
## xlevels  0     -none- list
```

```
#library(klaR)
partimat(Species~.,data=training,method="lda")
```




```r
partimat(Species~.,data=training,method="qda")
```



Check the accuracy of our analysis of QDA in Training Set

```
#Check the accuracy of our analysis of qda
Predictions_qda=predict(iris_qda,training)
table(Predictions_qda$class, training$Species)

##
##              setosa versicolor virginica
##   setosa         33          0         0
##   versicolor      0         34         0
##   virginica       0          0        31
```

Check the accuracy of our analysis of QDA in Testing Set

```
#Check the accuracy of our analysis of qda
Predictions_qda=predict(iris_qda,testing)
table(Predictions_qda$class, testing$Species)

##
##              setosa versicolor virginica
##   setosa         17          0         0
##   versicolor      0         14         0
##   virginica       0          2        19
```