

DSA 8070 R Session 11: Discrimination and Classification

Whitney

October 31, 2021

Contents

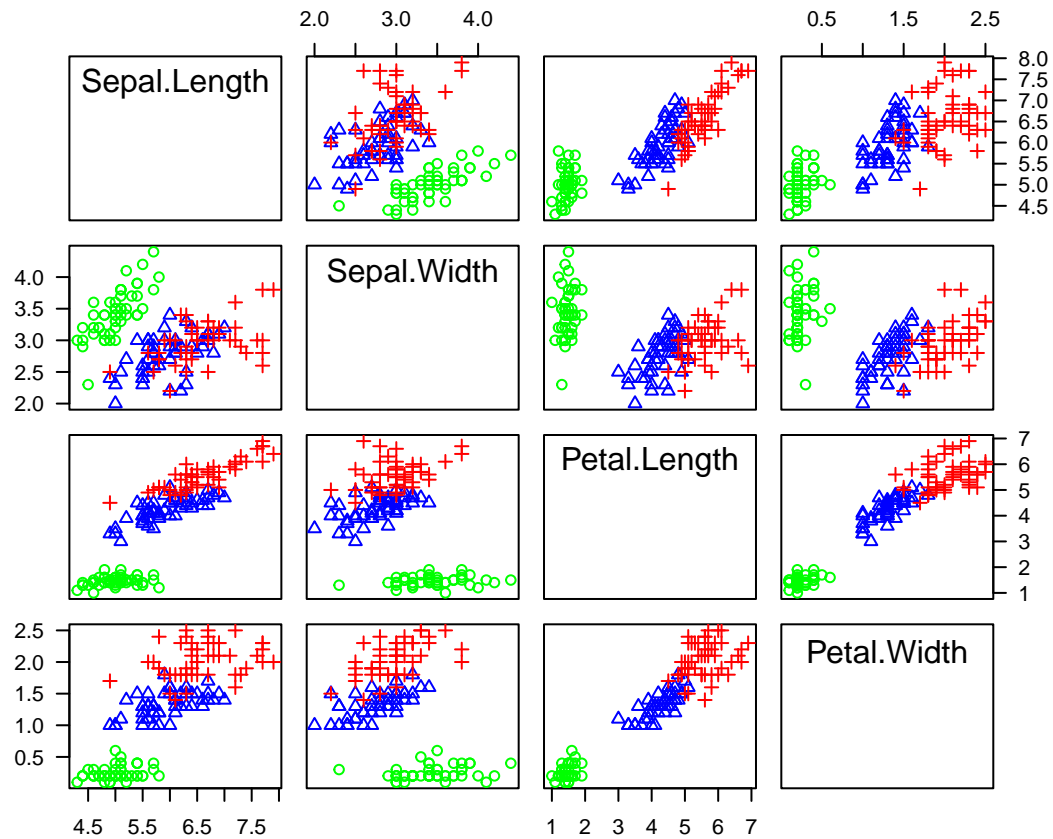
Iris data	1
Binary classification	2
PCA	3
Linear Discriminant Analysis (LDA)	5
Logistic Regression	6
LDA vs. QDA	7

Iris data

```
data(iris)
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

attach(iris)
library(car)
par(las = 1)
scatterplotMatrix(~ Sepal.Length + Sepal.Width + Petal.Length + Petal.Width | Species,
                  col = c("green", "blue", "red"), diagonal = F,
                  smooth = F, regLine = F, legend = F)
```

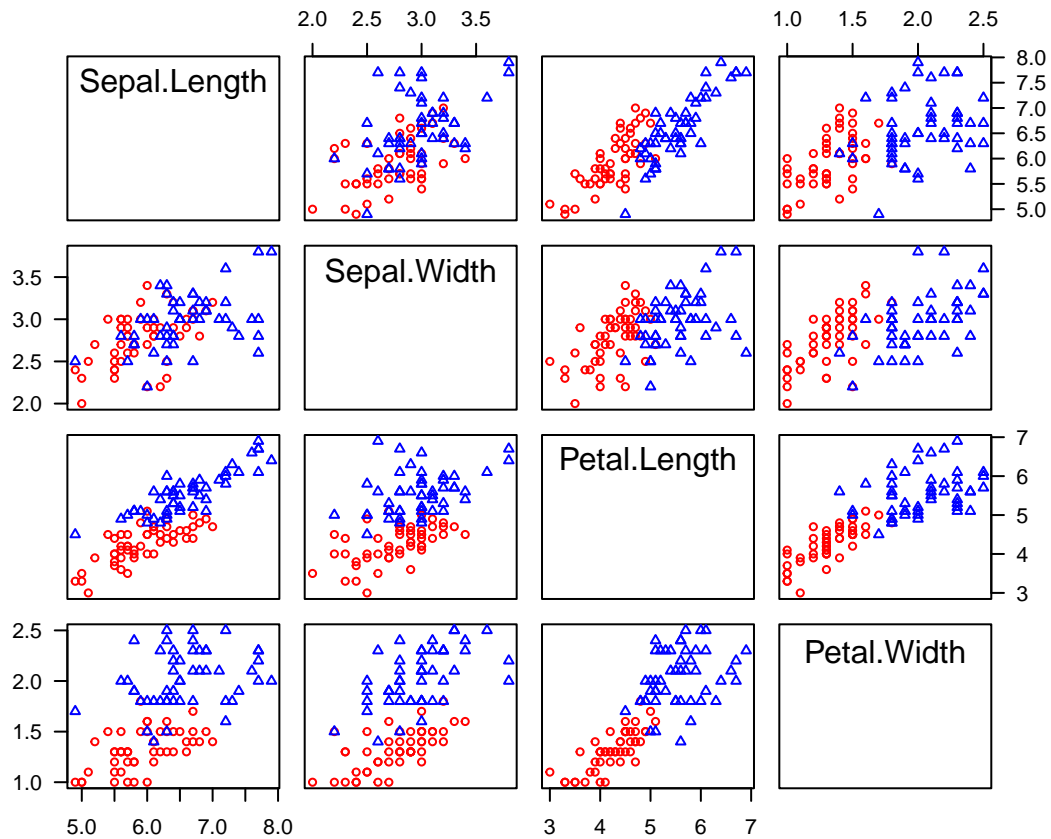


Binary classification

```

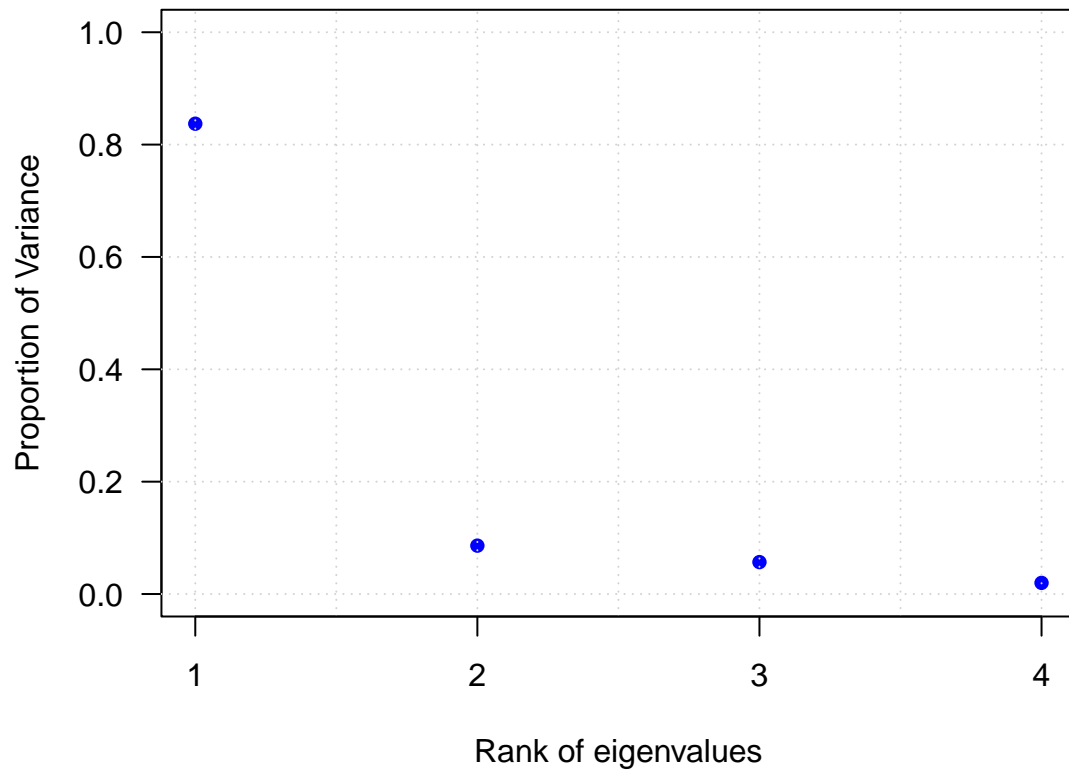
irisv = iris[51:150,]
irisv$Species <- factor(irisv$Species)
attach(irisv)
par(las = 1)
scatterplotMatrix(~ Sepal.Length + Sepal.Width + Petal.Length + Petal.Width | Species,
                  col = c("red", "blue"), diagonal = F,
                  smooth = F, regLine = F, legend = F, cex = 0.75)

```

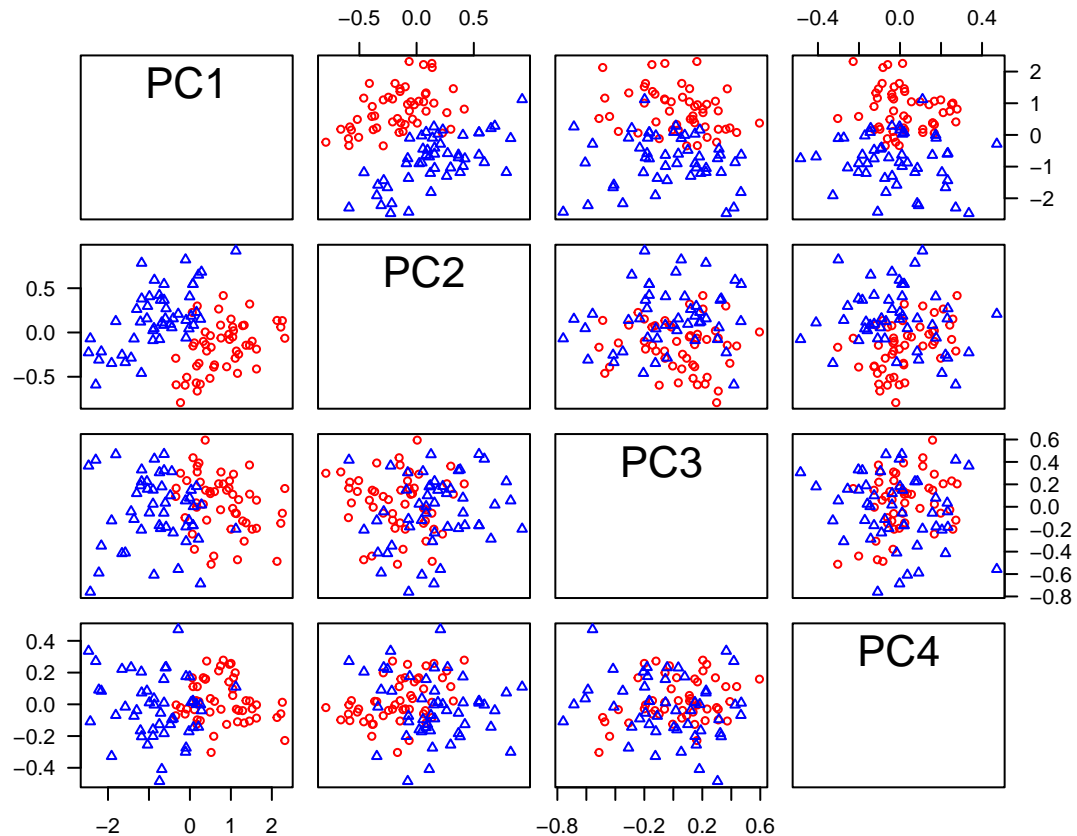


PCA

```
pca <- prcomp(irisv[, 1:4])
Z <- pca$x
lambda <- pca$sdev^2
par(las = 1)
plot(1:4, lambda / sum(lambda), xaxt = "n", las = 1, xlab = "Rank of eigenvalues",
     ylab = "Proportion of Variance", pch = 16, col = "blue", cex = 1, ylim = c(0, 1))
grid(); axis(1, at = 1:4)
```



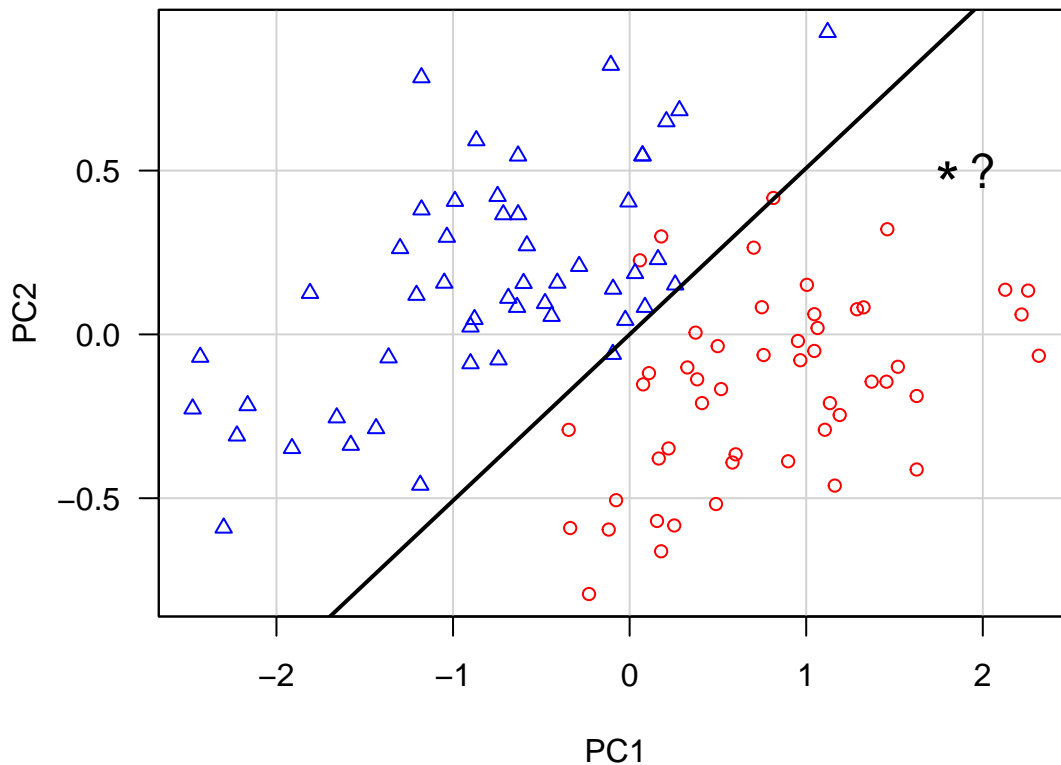
```
scatterplotMatrix(~ Z | Species, col = c("red", "blue"), diagonal = F, smooth = F,
  regLine = F, legend = F, cex = 0.75)
```



Linear Discriminant Analysis (LDA)

```
library(MASS)
par(las = 1)
scatterplot(PC2 ~ PC1 | Species, Z, smooth = F, regLine = F, legend = F, cex = 0.85,
            col = c("red", "blue"))
fit <- lda(Species ~ Z[, 1:2])
fit # show results
```

```
## Call:
## lda(Species ~ Z[, 1:2])
##
## Prior probabilities of groups:
## versicolor virginica
##      0.5      0.5
##
## Group means:
##      Z[, 1:2]PC1 Z[, 1:2]PC2
## versicolor   0.7930189 -0.1607571
## virginica   -0.7930189  0.1607571
##
## Coefficients of linear discriminants:
##      LD1
## Z[, 1:2]PC1 -1.553249
## Z[, 1:2]PC2  3.060560
abline(0, -fit$scaling[1] / fit$scaling[2], pch = 5, lwd = 2)
points(2, 0.5, pch = "?", cex = 1.5)
points(1.8, 0.5, pch = "*", cex = 2)
```

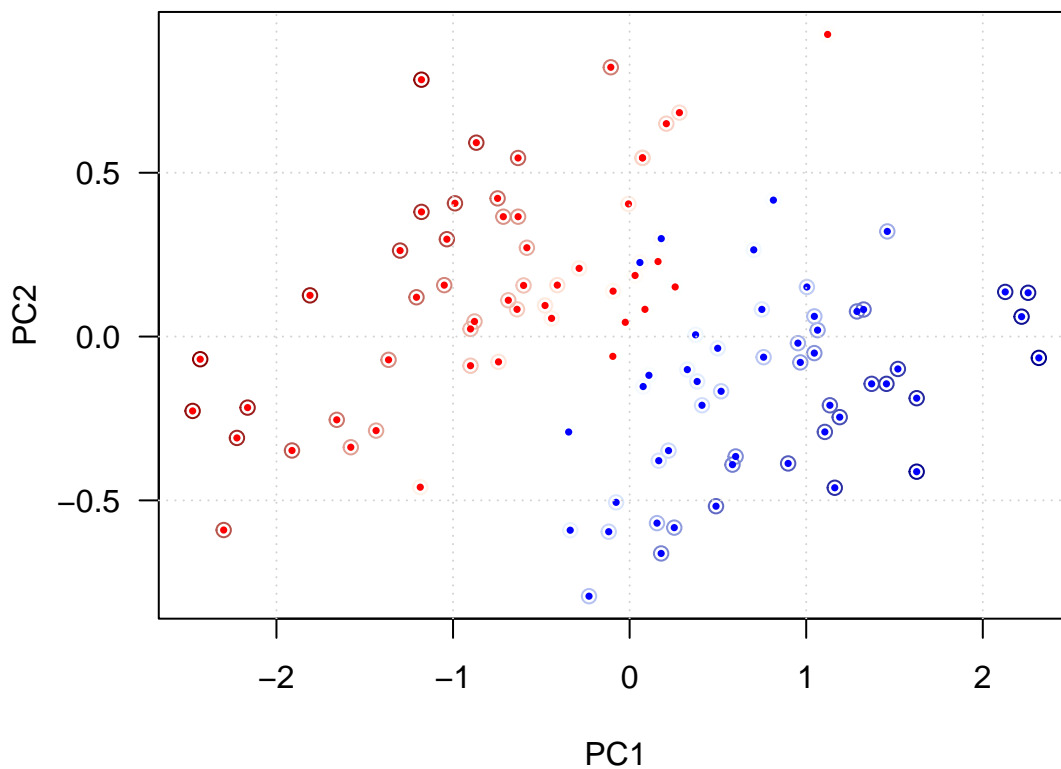


Logistic Regression

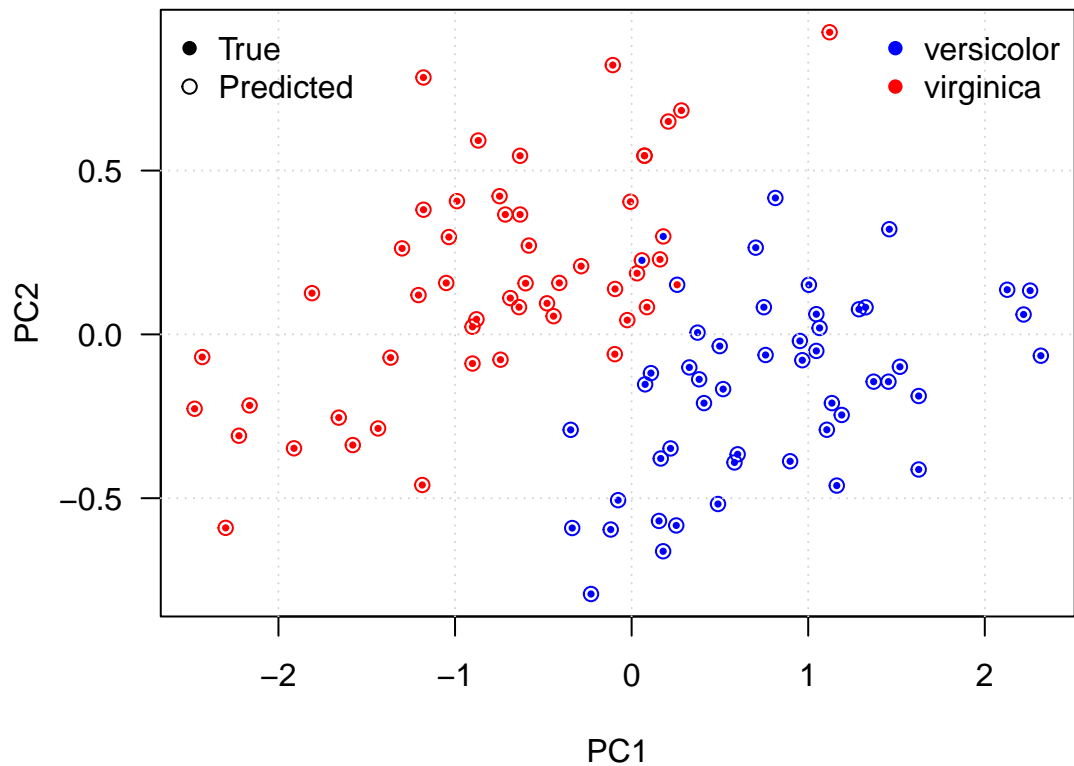
```
logfit <- glm(irisv$Species ~ Z[, 1:2], family = binomial)
logpred <- predict(logfit, type = "response")
library(fields)
cols <- two.colors(n = 100, "darkblue", "darkred")
order <- order(logpred)

predCol <- ifelse(logpred <= 0.5, "blue", "red")
Col <- rep(c("blue", "red"), each = 50)

plot(Z[order, 1:2], col = cols, pch = 1, las = 1)
points(Z[order, 1:2], col = Col[order], pch = 16, cex = 0.5)
grid()
```



```
plot(Z[, 1:2], col = predCol, pch = 1, las = 1)
points(Z[, 1:2], col = Col, pch = 16, cex = 0.5)
grid()
legend("topleft", legend = c("True", "Predicted"), pch = c(16, 1), bty = "n")
legend("topright", legend = c("versicolor", "virginica"),
      col = c("blue", "red"), pch = 16, bty = "n")
```



```
logisticPred <- ifelse(logpred <= 0.5, "versicolor", "virginica")
table(irisv$Species, logisticPred)
```

```
##          logisticPred
##          versicolor virginica
## versicolor         48         2
## virginica          1         49
```

LDA vs. QDA

```
#treat data as matrix
z = as.matrix(Z)

lda <- lda(irisv$Species ~ Z[, 1:2])
qda <- qda(irisv$Species ~ Z[, 1:2])

fit.LDA = predict(lda)$class
table(irisv$Species, fit.LDA)
```

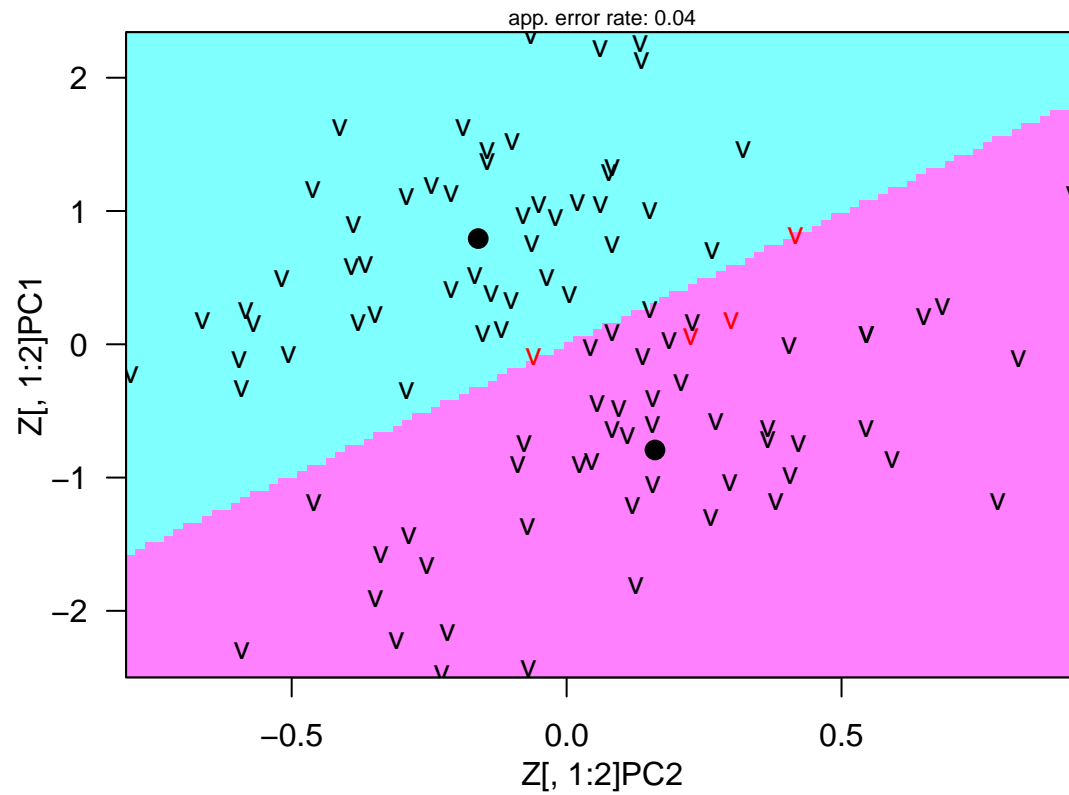
```
##          fit.LDA
##          versicolor virginica
## versicolor         47         3
## virginica          1         49
```

```
fit.QDA = predict(qda)$class
table(irisv$Species, fit.QDA)
```

```
##          fit.QDA
##          versicolor virginica
## versicolor         47         3
```

```
## virginica      2      48
# show results
library(klaR)
par(las = 1, mgp = c(2, 1, 0), mar = c(3.5, 3.5, 2, 1))
partimat(Species ~ Z[, 1:2], method = "lda")
```

Partition Plot



```
partimat(Species ~ Z[, 1:2], method = "qda")
```


Partition Plot

