

Problem 1: Linear Discriminant Analysis

Install the package `MASS` and use the function `lda` to perform Fisher's linear discriminant analysis to the Fisher's iris data set. The data set can be accessed with the command `data(iris)`. It contains measurements of sepal width, sepal length, petal width and petal length. In your analysis, consider the two species that are the most difficult to separate: `versicolor` and `virginica`, i.e., leave the first 50 observations out.

- Use `lda` to find the vector a . Verify that a is equal to the eigenvector of $W^{-1}B$ that corresponds to the largest eigenvalue.
- Suppose we have a new flower with the following measurements: (`Sepal.Length`, `Sepal.Width`, `Petal.Length`, `Petal.Width`) = (6.08, 2.76, 4.6, 1.44). In which group will this flower be classified based on Fisher's discriminant analysis?
- Use the leave-one-out method to determine the missclassification rate for Fisher's discriminant analysis.

Solution

Install the package `MASS` if you haven't yet.

```
install.packages("MASS")
```

Import the package `MASS` and read the data. Leave only species `versicolor` and `virginica`, i.e., drop the first 50 observations. We also drop the redundant level `setosa` with the function `droplevels`. Note that the column `Species` is of a class factor not character.

```
library(MASS)
data(iris)
```

```
iris <- droplevels(iris[-(1:50), ])
str(iris)
```

```
## 'data.frame':    100 obs. of  5 variables:
## $ Sepal.Length: num  7 6.4 6.9 5.5 6.5 5.7 6.3 4.9 6.6 5.2 ...
## $ Sepal.Width : num  3.2 3.2 3.1 2.3 2.8 2.8 3.3 2.4 2.9 2.7 ...
## $ Petal.Length: num  4.7 4.5 4.9 4 4.6 4.5 4.7 3.3 4.6 3.9 ...
## $ Petal.Width : num  1.4 1.5 1.5 1.3 1.5 1.3 1.6 1 1.3 1.4 ...
## $ Species      : Factor w/ 2 levels "versicolor","virginica": 1 1 1 1 1 1 1 1 1 1 ...
```

Firstly, we visualize the data in Figure 1.

```
pairs(iris[, 1:4], pch = c(16, 17)[iris$Species], gap = 0, upper.panel = NULL,
      col = c(rgb(0, 0, 1, 0.5), rgb(1, 0, 0, 0.5))[iris$Species])
par(xpd = TRUE)
legend(0.75, 0.75, legend = levels(iris$Species), pch = c(16, 17),
      col = c(rgb(0, 0, 1, 0.5), rgb(1, 0, 0, 0.5)), cex = 1)
```

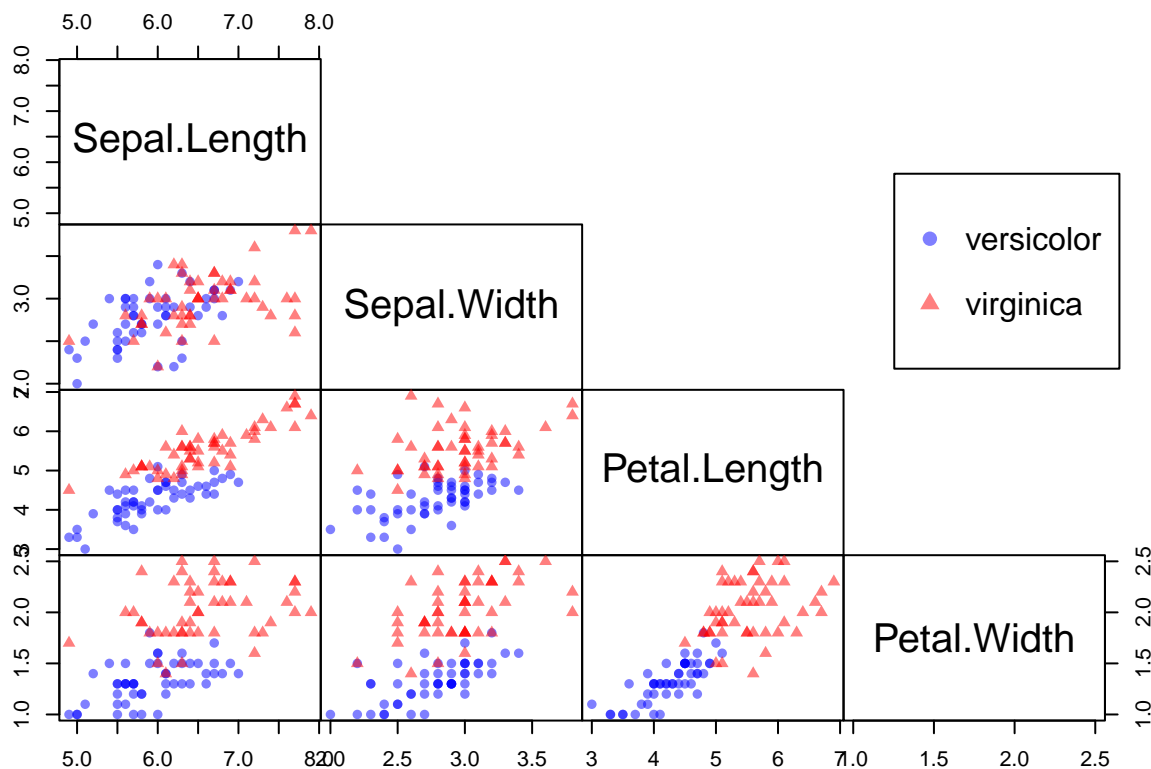


Figure 1: Pairwise scatter plots of the variables.

- a) With the function `lda` from the package `MASS` we can perform Fisher's linear discriminant analysis. Function `lda` can be used quite similarly compared to the function `lm`. That is, the input can be given as a formula

$$y \sim x_1 + x_2 + \dots + x_g,$$

where response y is the grouping factor and x_1, x_2, \dots, x_g are the discriminators. In this case,

$y = \text{Species}$,
 $x_1 = \text{Sepal.Length}$,
 $x_2 = \text{Sepal.Width}$,
 $x_3 = \text{Petal.Length}$ and
 $x_4 = \text{Petal.Width}$.

Below we perform Fisher's linear discriminant analysis and give the vector a .

```
# Species ~ . is a shorthand for
# Species ~ Sepal.Length + Sepal.Width + Petal.Length + Petal.Width
iris_lda <- lda(Species ~ ., data = iris)
a_lda <- iris_lda$scaling
a_lda
```

```
##                LD1
## Sepal.Length -0.9431178
## Sepal.Width -1.4794287
## Petal.Length  1.8484510
## Petal.Width  3.2847304
```

Next we perform Fisher's linear discriminant analysis manually. Let

$$W = \sum_{i=1}^g (n_i - 1) \text{Cov}(X_i) \quad \text{and} \quad B = \sum_{i=1}^g n_i (\bar{x}_i - \bar{x})(\bar{x}_i - \bar{x})^\top.$$

In the case when $g = 2$, we can express B as

$$B = \frac{n_1 n_2}{n} d d^\top,$$

where $d = \bar{x}_1 - \bar{x}_2$. Then vector a is obtained by computing

$$a = W^{-1}d.$$

In general case $g \in \{2, 3, \dots\}$, solution a is equal to the eigenvector corresponding to the largest eigenvalue of $W^{-1}B$.

```
# cov(X_1)
s1 <- cov(iris[1:50, 1:4])

# cov(X_2)
s2 <- cov(iris[51:100, 1:4])

d1 <- colMeans(iris[1:50, 1:4])
d2 <- colMeans(iris[51:100, 1:4])
d <- as.matrix(d1 - d2, ncol = 1)

b <- (50 * 50) / 100 * d %*% t(d)
w <- 49 * (s1 + s2)
l <- solve(w) %*% b

# a is the eigenvector corresponding to the largest eigenvalue of W^(-1)B
a_manual <- eigen(l)$vectors[, 1]

# When g = 2, a = W^(-1)d
a_manual2 <- solve(w) %*% d
```

By comparing `a_lda`, `a_manual` and `a_manual2` it can be seen that they are equal up to scale.

```
c(norm(a_lda, type = "2"), norm(a_manual, type = "2"),
  norm(a_manual2, type = "2"))

## [1] 4.1574518 1.0000000 0.1599683

a_lda <- a_lda / norm(a_lda, type = "2")
a_manual2 <- a_manual2 / norm(a_manual2, type = "2")

data.frame(a_lda = a_lda, a_manual = a_manual, a_manual2 = a_manual2)

##           LD1    a_manual  a_manual2
## Sepal.Length -0.2268500 -0.2268500  0.2268500
## Sepal.Width  -0.3558499 -0.3558499  0.3558499
## Petal.Length  0.4446115  0.4446115 -0.4446115
## Petal.Width   0.7900826  0.7900826 -0.7900826
```

b) As below snippet of code shows, the new flower will be classified as `versicolor`.

```
newobs <- data.frame(Sepal.Length = 6.08, Sepal.Width = 2.76,
                    Petal.Length = 4.6, Petal.Width = 1.44)
predict(iris_lda, newdata = newobs)$class
```

```
## [1] versicolor
## Levels: versicolor virginica
```

We can also predict the group of the new observation x manually. New observation x is allocated to the population whose mean score is closest to the $a^T x$. That is, x is allocated to group j if

$$|a^T x - a^T \bar{x}_j| < |a^T x - a^T \bar{x}_i|, \quad \text{for all } i \neq j.$$

```
# The new observation is classified as versicolor
abs(t(a_lda) %*% (as.numeric(newobs) - d1)) <
  abs(t(a_lda) %*% (as.numeric(newobs) - d2))
```

```
##      [,1]
## LD1 TRUE
```

c) Below we perform leave-one-out cross-validation.

```
d_cv <- lda(Species ~ ., data = iris, CV = TRUE)
```

Results of the leave-one-out cross-validation can be cross tabulated. Here two **versicolor** are classified as **virginica** and one **virginica** is classified as **versicolor**.

```
result <- data.frame(est = d_cv$class, truth = iris[, 5])
table(result)
```

```
##           truth
## est      versicolor virginica
## versicolor      48         1
## virginica       2         49
```

Thus the misclassification rate is $3/100 = 0.03$.

We can also perform the leave-one-out cross-validation manually.

```
predicted <- rep(NA, 100)
for (i in 1:100) {
  train <- iris[-i, ]
  test <- iris[i, ]
  predicted[i] <- predict(lda(Species ~ ., data = train),
                        newdata = test)$class
}
predicted <- factor(predicted, levels = c(1, 2),
                  labels = c("versicolor", "virginica"))
sum(predicted != iris$Species) / nrow(iris)

## [1] 0.03
```

Demo Problem 2: Classification Based on Halfspace Depth

Let us continue with the same data set as in Problem 1. That is, we consider species **versicolor** and **virginica** from Fisher's iris data set. However, this time we use the halfspace depth for classification.

- For each data point compute the sample halfspace depth with respect to the point cloud given by all observations of the same species.

- b) Suppose we have a new flower with the following measurements: (Sepal.Length, Sepal.Width, Petal.Length, Petal.Width) = (6.08, 2.76, 4.6, 1.44). In which group will this flower be classified based on halfspace depth?
- c) Use the leave-one-out method to determine the misclassification rate for classification based on halfspace depth.

Solution

First, install the package `ddalpha` if you haven't. It is an R package for depth-based classification.

```
install.packages("ddalpha")
```

Also, make sure that the data set `iris` is cleaned similarly as in Problem 1 (remove species `setosa`).

- a) Function `depth.halfspace` from package `ddalpha` can be used for computing the sample halfspace depth. By default the function `depth.halfspace` only approximates the sample halfspace depth since computation of the halfspace depth is demanding.

```
library(ddalpha)
depths_versicolor <- depth.halfspace(iris[1:50, 1:4], iris[1:50, 1:4],
                                     exact = TRUE)
depths_virginica <- depth.halfspace(iris[51:100, 1:4], iris[51:100, 1:4],
                                    exact = TRUE)
```

Figures 2 and 3 show that the most central observations have the highest depth value.

```
pal <- colorRampPalette(c("blue", "red"))
n_col <- length(unique(depths_versicolor))
ranks <- as.numeric(as.factor(depths_versicolor))

pairs(iris[1:50, 1:4], pch = 16, gap = 0, upper.panel = NULL,
      col = pal(n_col)[ranks])
par(xpd = TRUE)
legend(0.75, 0.75, legend = range(depths_versicolor), pch = 16,
      col = pal(2), cex = 1)
```

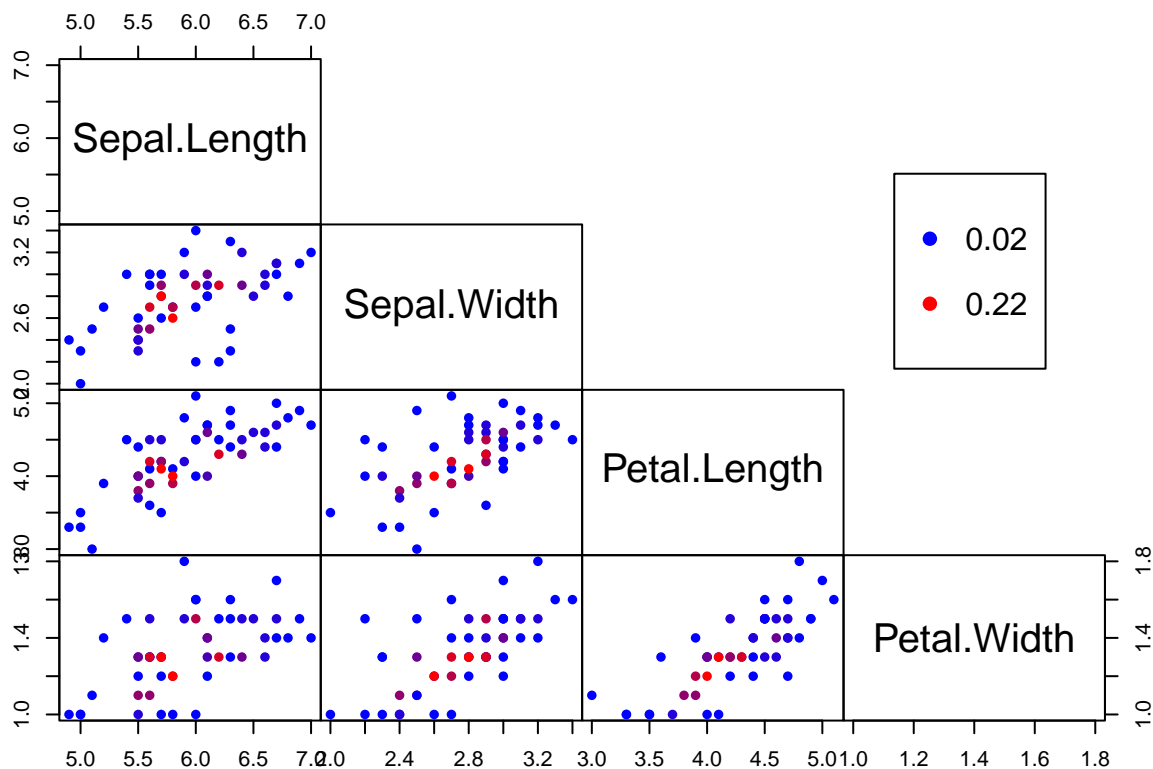


Figure 2: Pairwise scatter plots of variables for the species *versicolor*. Points are colored based on their values of the sample halfspace depth. Red indicates high value for the depth and blue indicates low value.

```
n_col <- length(unique(depths_virginica))
ranks <- as.numeric(as.factor(depths_virginica))

pairs(iris[51:100, 1:4], pch = 16, gap = 0, upper.panel = NULL,
      col = pal(n_col)[ranks])
par(xpd = TRUE)
legend(0.75, 0.75, legend = range(depths_virginica), pch = 16,
      col = pal(2), cex = 1)
```

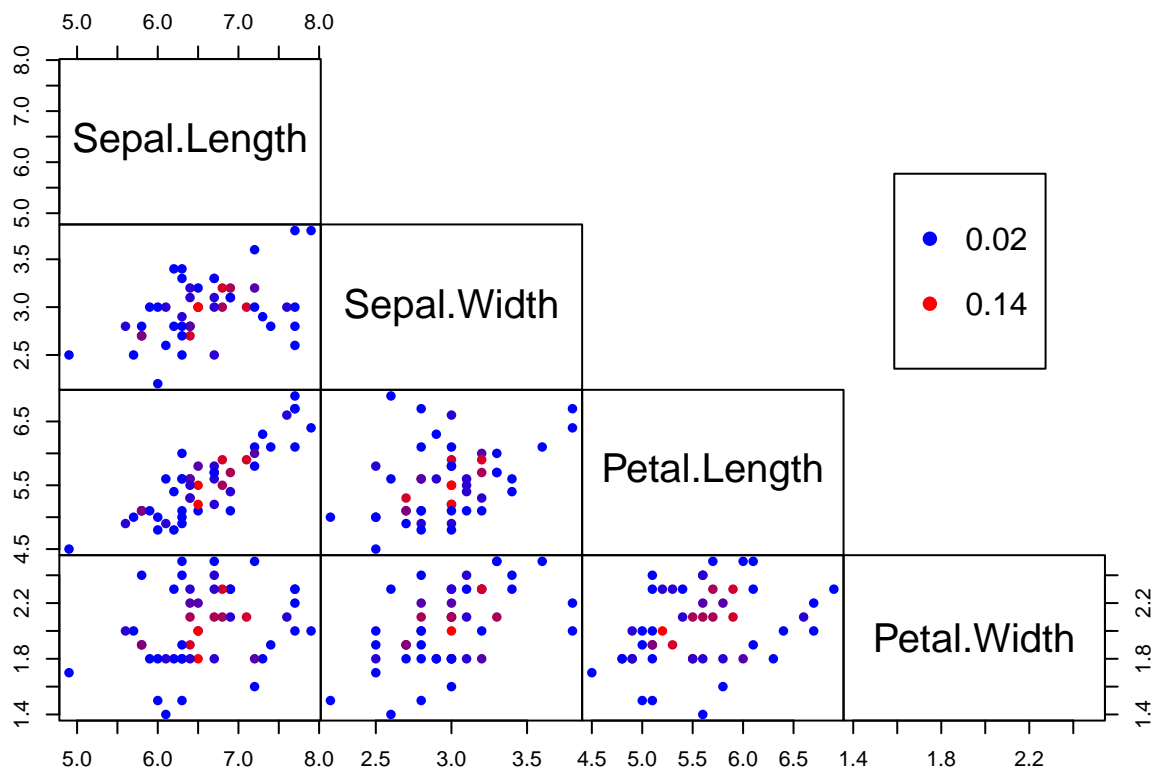


Figure 3: Pairwise scatter plots of variables for the species *virginica*. Points are colored based on their values of the sample halfspace depth. Red indicates high value for the depth and blue indicates low value.

- b) The new flower will be classified to a certain class if the sample halfspace depth of the new flower with respect to the pointcloud corresponding to the class is maximal. As in Problem 1, the new flower is classified as *versicolor* based on classification with halfspace depth.

```
depth.halfspace(newobs, iris[1:50, 1:4], exact = TRUE) >
depth.halfspace(newobs, iris[51:100, 1:4], exact = TRUE)
```

```
## [1] TRUE
```

- c) As below snippet of code shows the misclassification rate for depth based classification is 0.26 that is much higher than the misclassification rate for Fisher's linear discriminant analysis.

```
predicted <- rep(NA, 100)
for (i in 1:100) {
  train <- iris[-i, ]
  test <- iris[i, 1:4]
  train_versicolor <- train[train$Species == "versicolor", 1:4]
  train_virginica <- train[train$Species == "virginica", 1:4]

  cond <- (depth.halfspace(test, train_versicolor, exact = TRUE) >
    depth.halfspace(test, train_virginica, exact = TRUE))

  predicted[i] <- ifelse(cond, 1, 2)
}
predicted <- factor(predicted, levels = c(1, 2),
  labels = c("versicolor", "virginica"))
```

```
sum(predicted != iris$Species) / nrow(iris)
```

```
## [1] 0.26
```

Demo Problem 3: Fisher's Linear Discriminant Function

Show that the solution for the problem:

$$\max_a \left\{ \frac{a^\top B a}{a^\top W a} \right\},$$

is obtained by setting a equal to the eigenvector of $W^{-1}B$ that corresponds to the largest eigenvalue.

Solution

Since the vector a can be scaled arbitrarily without affecting the ratio, we can formulate the problem as follows:

$$\max_a \{a^\top B a\} \quad \text{s.t.} \quad a^\top W a = 1.$$

Note that matrix W is symmetric, and let $W^{1/2}$ be the symmetric square root of W . Let $z = W^{1/2}a$ (and then $a = W^{-1/2}z$). Then

$$\begin{aligned} a^\top B a &= (W^{-1/2}z)^\top B (W^{-1/2}z) = z^\top W^{-1/2} B W^{-1/2} z \quad \text{and} \\ a^\top W a &= (W^{-1/2}z)^\top W (W^{-1/2}z) = z^\top \underbrace{W^{-1/2} W W^{-1/2}}_{=I} z = z^\top z. \end{aligned}$$

Note that $W^{-1/2} B W^{-1/2}$ is symmetric and hereby the spectral decomposition exists ($W^{-1/2} B W^{-1/2} = \Gamma \Lambda \Gamma^\top$ and $\Gamma^\top \Gamma = I$). Denote $w = \Gamma^\top z$. Then

$$\begin{aligned} z^\top W^{-1/2} B W^{-1/2} z &= z^\top \Gamma \Lambda \Gamma^\top z = w^\top \Lambda w, \quad \text{and} \\ z^\top z &= z^\top \Gamma \Gamma^\top z = w^\top w. \end{aligned}$$

Now we can reformulate the problem as

$$\max_w \{w^\top \Lambda w\} = \max_w \left\{ \sum_{i=1}^p \lambda_i w_i^2 \right\} \quad \text{s.t.} \quad w^\top w = 1.$$

Since $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p$, we choose the first element of w to be one and the rest to be zero. This means that $z = \Gamma w = \gamma_1$, where γ_1 is the first eigenvector of $W^{-1/2} B W^{-1/2}$ and $a = W^{-1/2}z = W^{-1/2}\gamma_1$.

Note that for any two matrices $A \in \mathbb{R}^{n \times p}$ and $C \in \mathbb{R}^{p \times n}$, the non-zero eigenvalues of AC and CA are the same and have the same multiplicity Theorem A.6.2 of (Mardia, Kent, and Bibby 1979). Now let $A = W^{-1/2}B$ and $C = W^{-1/2}$, this means that the non-zero eigenvalues of $CA = W^{-1}B$ are the same as $AC = W^{-1/2} B W^{-1/2}$. Hence, λ_1 is the largest eigenvalue of $W^{-1}B$. Since γ_1 is the eigenvector corresponding to the largest λ_1 of $W^{-1/2} B W^{-1/2}$, we have that

$$W^{-1}B(W^{-1/2}\gamma_1) = W^{-1/2}(W^{-1/2} B W^{-1/2}\gamma_1) = W^{-1/2}\lambda_1\gamma_1 = \lambda_1(W^{-1/2}\gamma_1).$$

This shows that $a = W^{-1/2}\gamma_1$ is the eigenvector of $W^{-1}B$ corresponding to its largest eigenvalue λ_1 .

Homework Problem 1: Linear Discriminant Analysis

Perform the Fisher's linear discriminant analysis to the data set `alcohol.txt`. Suppose that we have a new drink with the following measurements: MECH = 500, ACET = 400, BU1 = 3, MEPR = 30, ACAL = 20 and LNPR01 = 10. Answer the following questions and provide the requested figure for part b).

- a) Give the vector a , such that $\|a\|_2 = 1$.
- b) Visualize the original data using a pairwise scatterplot. Add the new drink to the plot. Furthermore, use different colors for KIRSCH, MIRAB, POIRE and the new drink.
- c) In which group will the new drink be classified to?

References

Mardia, K. V., J. T. Kent, and J. M. Bibby (1979). *Multivariate analysis*. London: Academic press. ISBN: 0-12-471250-9.