Pattern recognition on spatial data

Predicting structure : discriminant analysis

Yves Brostaux

June 2017

Contents

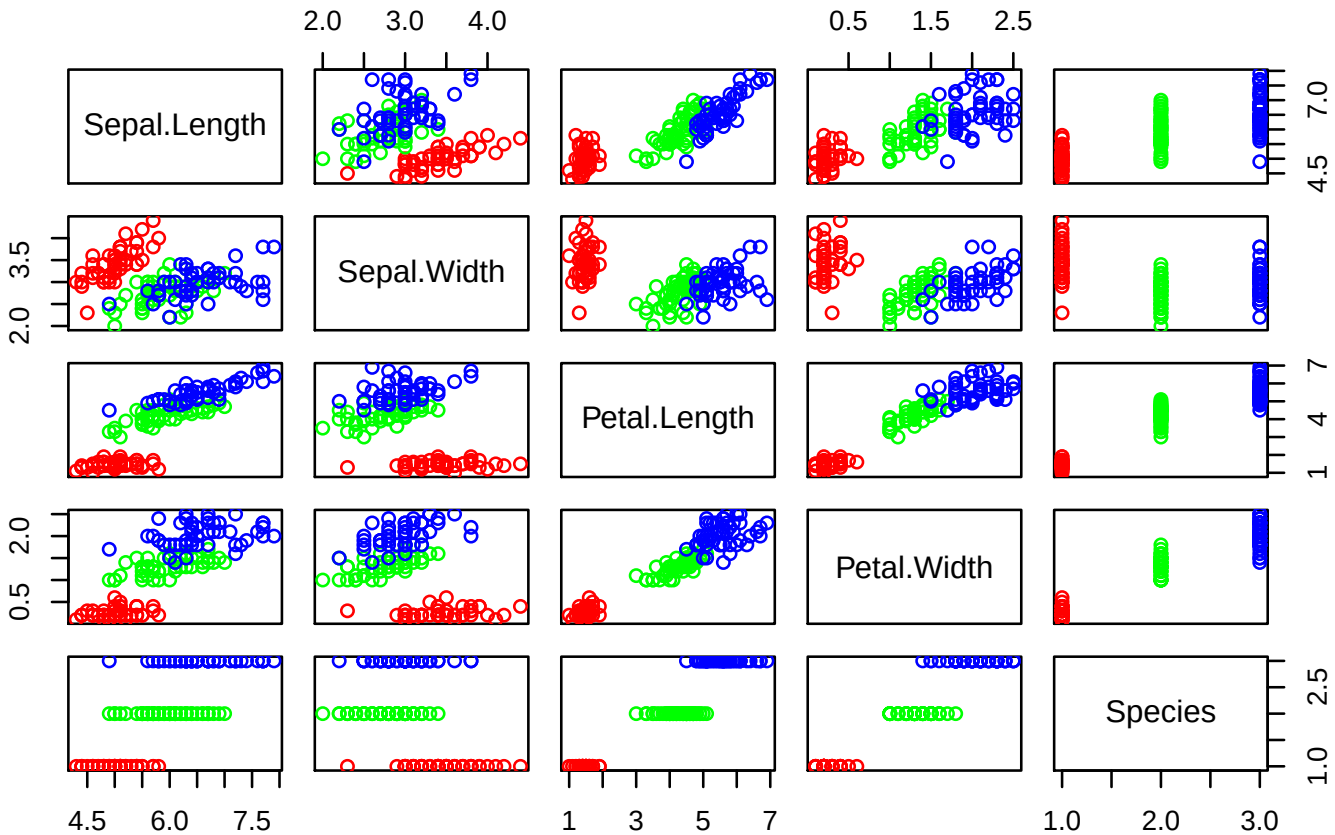
Introduction

- allocation of n observations into k pre existing groups

Example data

Iris data (FISHER, 1936)

- 3 species (50 obs/sp)
  - I. setosa
  - I. versicolor
  - I. virginica
- 4 variables
  - sepal length
  - sepal width
  - petal length
  - petal width


Sepal.Length
Sepal.Width
Petal.Length
Petal.Width
Species
2.0
3.0
4.0
0.5
1.5
2.5
4.5
7.0
2.0
3.5
1
4
7
0.5
2.0
1.0
2.5
4.5
6.0
7.5
1
3
5
7
1.0
2.0
3.0

Contents

Linear predictive discriminant analysis

- Two populations/one variable
- Two populations/two variables
- g populations/p variables
- Error rates

Two populations/one variable

Example

- I. versicolor and I. virginica
- petal length

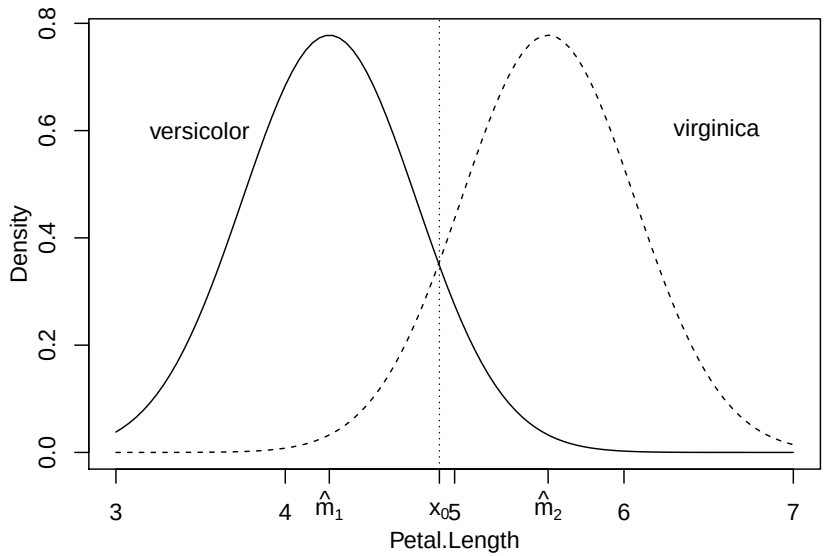
Species	I. versicolor	I. virginica
mean	4.260	5.552
standard-deviation	0.470	0.552

Assumptions

- normality
- homogeneity of within group variance

Two populations/one variable

$$\sigma^2 = \frac{(49)(0.470)^2 + (49)(0.552)^2}{49 + 49} = 0.263$$



Classification rules

- **Threshold**

$$x_0 = (\hat{m}_1 + \hat{m}_2)/2$$

- **Distance**

$$d_{1i}^2 = \left(\frac{x_i - \hat{m}_1}{\hat{\sigma}}\right)^2 \quad \text{et} \quad d_{2i}^2 = \left(\frac{x_i - \hat{m}_2}{\hat{\sigma}}\right)^2$$

- **Density**

$$f_1(x_i) = \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{x_i - \hat{m}_1}{\hat{\sigma}}\right)^2\right]$$

$$f_2(x_i) = \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{x_i - \hat{m}_2}{\hat{\sigma}}\right)^2\right]$$

Classification rules

- **Posterior probability**

$$P(A1 \mid x_i) = \frac{f_1(x_i)}{f_1(x_i) + f_2(x_i)} = \frac{\exp\left(-\frac{1}{2}d_{1i}^2\right)}{\sum_{k=1}^{2} \exp\left(-\frac{1}{2}d_{ki}^2\right)}$$

$$P(A2 \mid x_i) = \frac{f_2(x_i)}{f_1(x_i) + f_2(x_i)} = \frac{\exp\left(-\frac{1}{2}d_{2i}^2\right)}{\sum_{k=1}^{2} \exp\left(-\frac{1}{2}d_{ki}^2\right)}$$

Classification rules

Assign unit i to population 1 if :

- $x_i < x_0$
- $d_{1i}^2 < d_{2i}^2$
- $f_1(x_i) > f_2(x_i)$
- $\mathrm{P}(\mathrm{A1} \mid x_i) > \mathrm{P}(\mathrm{A2} \mid x_i)$

Example

$x_i = 4.7$

- $x_i = 4.7 < x_0 = 4.91$
- $d_{1i}^2 = 0.74 < d_{2i}^2 = 2.76$
- $f_1(x_i) = 0.54 > f_2(x_i) = 0.20$
- $$P(A1 \mid x_i) = \frac{0.54}{0.54 + 0.20} = 0.73$$
  $$P(A2 \mid x_i) = \frac{0.20}{0.54 + 0.20} = 0.27$$

$\Rightarrow$ x_i is allocated to population 1

Linear predictive discriminant analysis

- Two populations/one variable
- Two populations/two variables
- g populations/p variables
- Error rates

Two populations/two variables

Example

- I. versicolor and I. virginica
- petal length and petal width

Assumptions

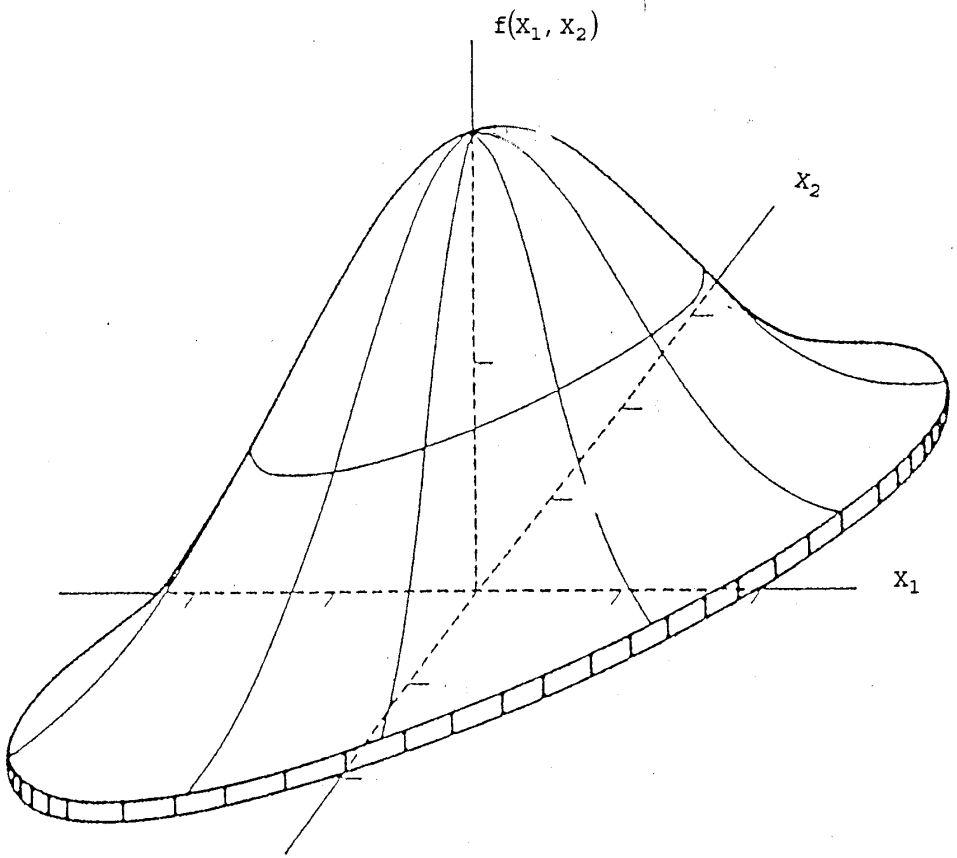
- normality
- homogeneity of within group covariance matrix

Two populations/two variables

	I. versicolor	I. virginica
mean length	4.260	5.552
mean width	1.326	2.026
covariance matrix	$\begin{bmatrix} 0.2208 & 0.0731 \\ 0.0731 & 0.0391 \end{bmatrix}$	$\begin{bmatrix} 0.3046 & 0.0488 \\ 0.0488 & 0.0754 \end{bmatrix}$

$$\hat{\sum} = \left(49\widehat{\sum}_1 + 49\widehat{\sum}_2\right)/98 = \begin{bmatrix} 0.2627 & 0.0610 \\ 0.0610 & 0.0573 \end{bmatrix}$$

Distribution normale à 2 dimensions



Classification rules

- **Density**

$$f_1(x_{1i}, x_{2i}) = \frac{1}{2\pi\hat{\sigma}_{x_1}\hat{\sigma}_{x_2}\sqrt{(1-\hat{\rho}^2)}} \exp\left[-\frac{1}{2}d_{1i}^2\right]$$

$$f_2(x_{1i}, x_{2i}) = \frac{1}{2\pi\hat{\sigma}_{x_1}\hat{\sigma}_{x_2}\sqrt{(1-\hat{\rho}^2)}} \exp\left[-\frac{1}{2}d_{2i}^2\right]$$

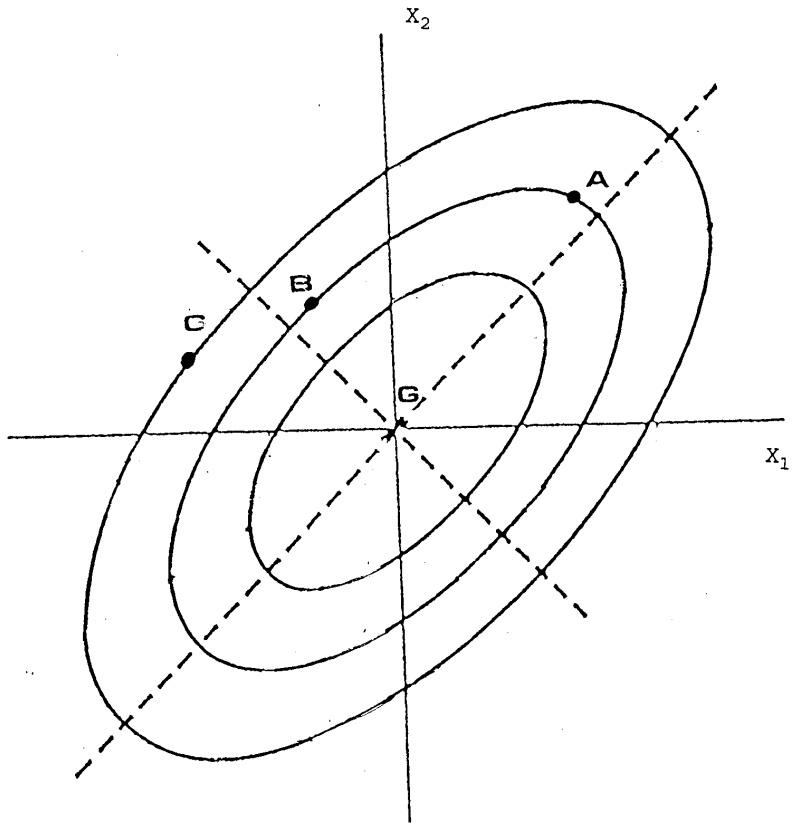
Classification rules

- **Distance (Mahalanobis)**

$$d_{1i}^2 = \frac{1}{1-\hat{\rho}^2}\left[\left(\frac{x_{1i}-\hat{m}_{11}}{\hat{\sigma}_{x_1}}\right)^2 - 2\hat{\rho}\left(\frac{x_{1i}-\hat{m}_{11}}{\hat{\sigma}_{x_1}}\right)\left(\frac{x_{2i}-\hat{m}_{12}}{\hat{\sigma}_{x_2}}\right) + \left(\frac{x_{2i}-\hat{m}_{12}}{\hat{\sigma}_{x_2}}\right)^2\right]$$

$$d_{2i}^2 = \frac{1}{1-\hat{\rho}^2}\left[\left(\frac{x_{1i}-\hat{m}_{21}}{\hat{\sigma}_{x_1}}\right)^2 - 2\hat{\rho}\left(\frac{x_{1i}-\hat{m}_{21}}{\hat{\sigma}_{x_1}}\right)\left(\frac{x_{2i}-\hat{m}_{22}}{\hat{\sigma}_{x_2}}\right) + \left(\frac{x_{2i}-\hat{m}_{22}}{\hat{\sigma}_{x_2}}\right)^2\right]$$

Mahalanobis' distance



Classification rules

- **Distance (Mahalanobis)**

$$d_{1i}^2 = [\mathbf{x}_i - \hat{\mathbf{m}}_1]' \hat{\mathbf{\Sigma}}^{-1} [\mathbf{x}_i - \hat{\mathbf{m}}_1]$$
$$d_{2i}^2 = [\mathbf{x}_i - \hat{\mathbf{m}}_2]' \hat{\mathbf{\Sigma}}^{-1} [\mathbf{x}_i - \hat{\mathbf{m}}_2]$$

$$\mathbf{x}_i = \begin{bmatrix} x_{1i} \\ x_{2i} \end{bmatrix} \hat{\mathbf{m}}_1 = \begin{bmatrix} \hat{m}_{11} \\ \hat{m}_{12} \end{bmatrix} \hat{\mathbf{m}}_2 = \begin{bmatrix} \hat{m}_{21} \\ \hat{m}_{22} \end{bmatrix} \hat{\mathbf{\Sigma}} = \begin{bmatrix} \hat{\sigma}_{x_1}^2 & \hat{\mu}_{11} \\ \hat{\mu}_{11} & \hat{\sigma}_{x_2}^2 \end{bmatrix}$$

Classification rules

- **Posterior probability**

$$P(A1 \mid x_i) = \frac{f_1(x_i)}{f_1(x_i) + f_2(x_i)} = \frac{\exp\left(-\frac{1}{2} d_{1i}^2\right)}{\sum_{k=1}^{2} \exp\left(-\frac{1}{2} d_{ki}^2\right)}$$

$$P(A2 \mid x_i) = \frac{f_2(x_i)}{f_1(x_i) + f_2(x_i)} = \frac{\exp\left(-\frac{1}{2} d_{2i}^2\right)}{\sum_{k=1}^{2} \exp\left(-\frac{1}{2} d_{ki}^2\right)}$$

Classification rules

Assign unit i to population 1 if :

- $d_{1i}^2 < d_{2i}^2$
- $f_1(x_i) > f_2(x_i)$
- $\mathrm{P}(\mathrm{A1} \mid x_i) > \mathrm{P}(\mathrm{A2} \mid x_i)$

Example

$$x_i = \begin{bmatrix} 4.7 \\ 1.6 \end{bmatrix}$$

- $d^2_{1i} = 1.422 < d^2_{2i} = 3.972$
- $f_1(x_i) = 0.734 > f_2(x_i) = 0.205$
- $$\mathrm{P}(\mathrm{A1} \mid x_i) = \frac{0.734}{0.734 + 0.205} = 0.782$$
  $$\mathrm{P}(\mathrm{A2} \mid x_i) = \frac{0.205}{0.734 + 0.205} = 0.218$$

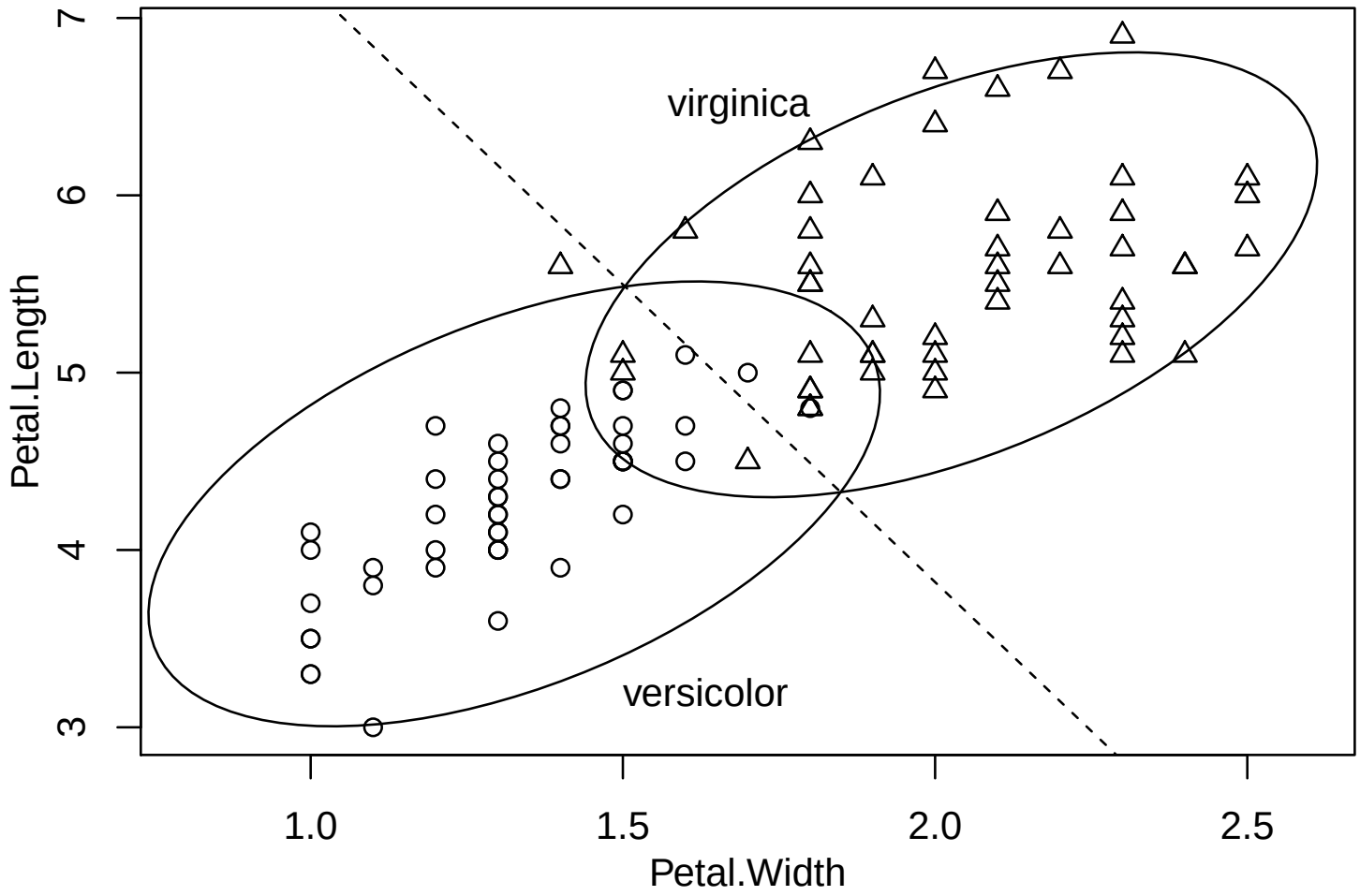
$\Rightarrow$ x_i is allocated to population 1

Geometric interpretation

The limit between the two populations is defined by the set of points which are at **equal (Mahalanobis') distance** from the centroïds of the populations.

This set of points draw a **straight line** between the populations, which pass through the **intersection of the ellipses** of equal Mahalanobis' distance, giving its name to the method (*linear* discriminant analysis).

Geometric interpretation



Linear predictive discriminant analysis

- Two populations/one variable
- Two populations/two variables
- g populations/p variables
- Error rates

g populations/p variables

Example

- I. versicolor, I. virginica, I. setosa
- petal length and petal width

Assumptions

- normality
- homogeneity of within group covariance matrix

Classification rules

- **Distance (for population h)**

$$d_{hi}^2 = [\mathbf{x}_i - \hat{\mathbf{m}}_h]' \, \hat{\mathbf{\Sigma}}^{-1} \, [\mathbf{x}_i - \hat{\mathbf{m}}_h]$$

- **Density (for population h)**

$$f_h(\mathbf{x}_i) = \frac{1}{\sqrt{(2\pi)^p |\hat{\mathbf{\Sigma}}|}} \exp\left[-\frac{1}{2} d_{hi}^2\right]$$

- **Posterior probability (for population h)**

$$\mathrm{P}(\mathrm{A}h \mid \mathbf{x}_i) = \frac{\exp\left(-\frac{1}{2} d_{hi}^2\right)}{\sum_{k=1}^{g} \exp\left(-\frac{1}{2} d_{ki}^2\right)}$$

Classification rules

- **Likelihood ratio**

$$\frac{f_h(\mathbf{x})}{f_l(\mathbf{x})} = \frac{\left(1/\sqrt{(2\pi)^p|\hat{\mathbf{\Sigma}}|}\right)\exp\left[-\frac{1}{2}(x-\hat{m}_h)'\hat{\mathbf{\Sigma}}^{-1}(x-\hat{m}_h)\right]}{\left(1/\sqrt{(2\pi)^p|\hat{\mathbf{\Sigma}}|}\right)\exp\left[-\frac{1}{2}(x-\hat{m}_l)'\hat{\mathbf{\Sigma}}^{-1}(x-\hat{m}_l)\right]}$$

- **Log-likelihood ratio**

$$\log_e(L_{hl} = \left(\hat{\mathbf{m}}_h\hat{\mathbf{\Sigma}}^{-1}\mathbf{x} - \frac{1}{2}\hat{\mathbf{m}}_h'\hat{\mathbf{\Sigma}}^{-1}\hat{\mathbf{m}}_h\right) - \left(\hat{\mathbf{m}}_l\hat{\mathbf{\Sigma}}^{-1}\mathbf{x} - \frac{1}{2}\hat{\mathbf{m}}_l'\hat{\mathbf{\Sigma}}^{-1}\hat{\mathbf{m}}_l\right)$$

Example with R

```r
# load data (internal)
data(iris)

# select only petal length and width
iris4 <- subset(iris, select=3:5)

# load package
library(MASS)

# adjust the LDA
iris.lda <- lda(Species~., data=iris4)
```

Example with R

```
## Call:
## lda(Species ~ ., data = iris4)
##
## Prior probabilities of groups:
##     setosa versicolor  virginica
##  0.3333333  0.3333333  0.3333333
##
## Group means:
##            Petal.Length Petal.Width
## setosa            1.462       0.246
## versicolor        4.260       1.326
## virginica         5.552       2.026
##
## Coefficients of linear discriminants:
##                    LD1       LD2
## Petal.Length 1.544371 -2.161222
## Petal.Width  2.402394  5.042599
##
## Proportion of trace:
##    LD1    LD2
## 0.9947 0.0053
```

Canonical discriminant analysis

Linear discriminant analysis can also be seen as a **factor analysis** (like PCA), which aims at creating linear combinations of the original variables that gives the best possible separation between the groups.

Canonical variables are then calculated by an similar procedure to PCA, but the criteria of maximum variance of the resulting components is replaced by the **maximum separation between the groups**.

$$F = \frac{\sigma^2_{Between}}{\sigma^2_{Within}}$$

Linear discriminant functions

```
# Linear discriminants coefficients
iris.lda$scaling
```

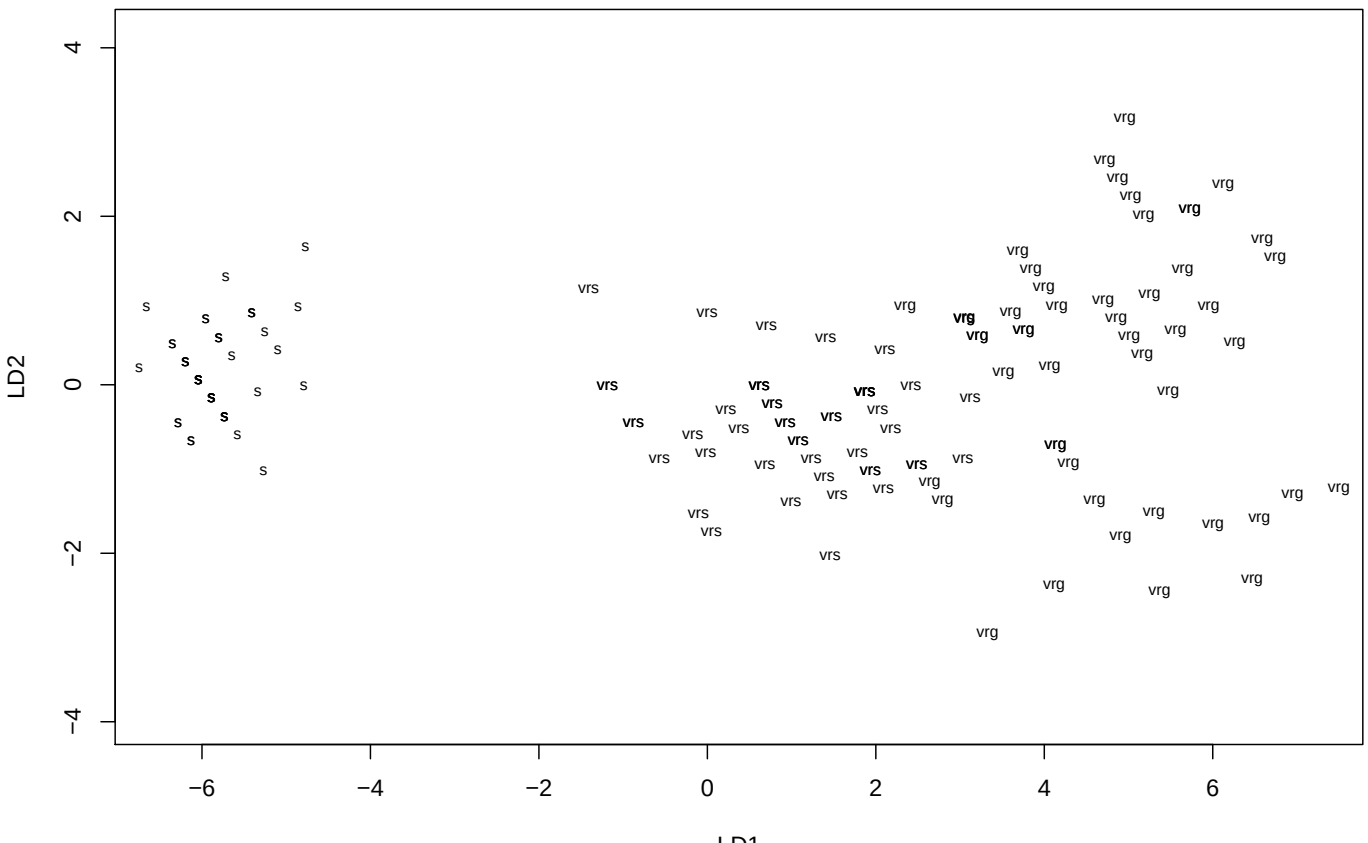
```
##                    LD1        LD2
## Petal.Length 1.544371 -2.161222
## Petal.Width  2.402394  5.042599
```

```
# Separation between populations
iris.lda$svd^2/sum(iris.lda$svd^2)
```

```
## [1] 0.99470499 0.00529501
```

Linear discriminant functions

```
plot(iris.lda, abbrev=1)
```



Linear discriminant scores

```
iris.pred <- predict(iris.lda)
# class prediction (class with maximum post prob)
head(iris.pred$class, n=5)
```

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

```
# posterior probability of each class
head(iris.pred$posterior, n=5)
```

```
##   setosa   versicolor    virginica
## 1      1 8.750491e-12 4.742801e-26
## 2      1 8.750491e-12 4.742801e-26
## 3      1 2.640992e-12 9.514213e-27
## 4      1 2.899331e-11 2.364269e-25
## 5      1 8.750491e-12 4.742801e-26
```

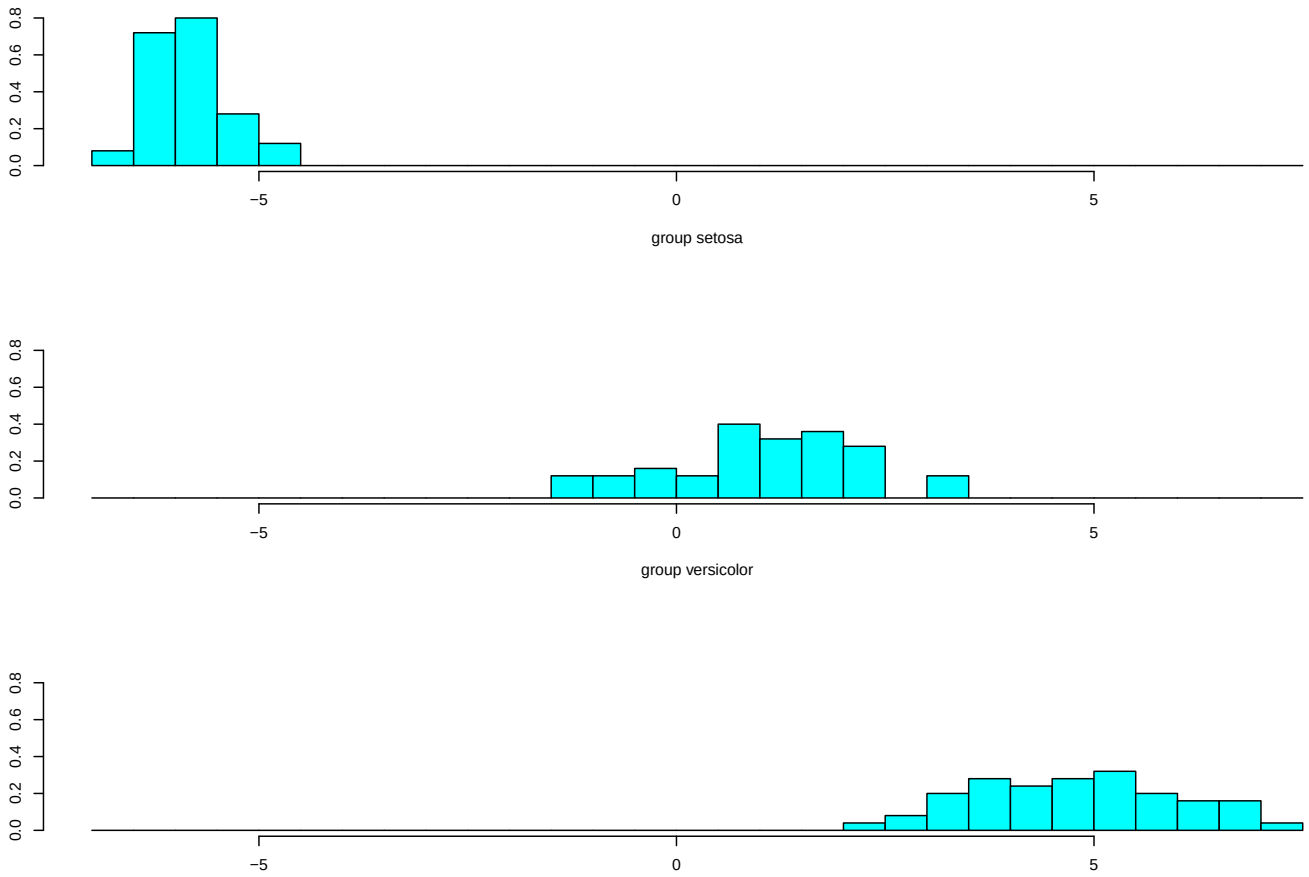
Linear discriminant scores

```
# Canonical scores
head(iris.pred$x, n=5)
```

```
##          LD1         LD2
## 1 -6.042418  0.05692487
## 2 -6.042418  0.05692487
## 3 -6.196856  0.27304711
## 4 -5.887981 -0.15919736
## 5 -6.042418  0.05692487
```

Linear discriminant scores

```
ldahist(iris.pred$x[,1], iris$Species)
```



Linear predictive discriminant analysis

- Two populations/one variable
- Two populations/two variables
- g populations/p variables
- Error rates

Definitions

Optimal error rate

Theoretical error rate when affectation rule is based on real population parameters. Function of Mahalanobis' distance between centroids of populations.

Actual error rate

Observed error rate when affecting new individuals from the same mix of populations used to create affectation rules

Expected actual error rate

Mathematical expectation of the actual error rate

Parametric estimators

Only for some situations

- function of the classification rule
- function of the (unknown) parameters of the populations

Example : LCF, $g = 2, \quad p_1 = p_2$
Optimal error rate : $eo = \Phi(-\Delta/2)$

Non parametric estimators

Percent of misclassified observations

- resubstitution
- holdout
  - training sample
  - test sample
- leave-n-out
  - K-cross validation
  - jackknife
- bootstrap

Prediction error

Resubstitution confusion matrix

```
iris.err <- table(iris4$Species, iris.pred$class)
iris.err
```

```
##
##                setosa versicolor virginica
##    setosa          50          0         0
##    versicolor       0         48         2
##    virginica        0          4        46
```

```
# resubstitution error rate
1 - sum(diag(iris.err))/sum(iris.err)
```

```
## [1] 0.04
```

Prediction error

Cross validated confusion matrix

```
# compute lda with cross validation
iris.cv <- lda(Species~., data=iris4, CV=TRUE)

iris.ecv <- table(iris4$Species, iris.cv$class)
iris.ecv
```

```
##
##                setosa versicolor virginica
##    setosa          50          0         0
##    versicolor       0         48         2
##    virginica        0          4        46
```