Analyse discriminante sous R

Tanagra

02/12/2020

Importation des données

```
#modifier le dossier de travail
setwd("C:/Users/Zatovo/Desktop/demo")

#chargement données de modélisation
library(xlsx)
D1 <- read.xlsx("Data_LDA_R.xlsx",sheetName = "DATA_TRAIN")
str(D1)
```

```
## 'data.frame':    52 obs. of  9 variables:
##  $ MEOH: num  336 442 373 418 84 ...
##  $ ACET: num  225 338 356 62 65 ...
##  $ BU1 : num  1 1.9 0 0.8 2 2.2 1.9 0.4 0.9 0.8 ...
##  $ BU2 : num  1 10 29 0 2 52.1 46 3 36 12 ...
##  $ ISOP: num  92 91 83 89 2 123 85 6 84 7 ...
##  $ MEPR: num  37 30 27 24 0 38.2 33 9 36 9 ...
##  $ PRO1: num  177 552 814 342 288 ...
##  $ ACAL: num  0 31 11 7 6 13.3 35 4 4.8 2 ...
##  $ TYPE: chr  "KIRSCH" "KIRSCH" "KIRSCH" "KIRSCH" ...
```

Modélisation

```
#lda avec dicriminR - modélisation sans sélection
library(discriminR)
mdl <- discriminR::lda(data=D1,target="TYPE",stepdisc=FALSE)

#affichage standard
print(mdl)
```

```
##                   Count
## Total Sample Size    52
## Variables             8
## Classes               3
##
## Class Level Information :
##        frequency proportion
## KIRSCH        17  0.3269231
## MIRAB         15  0.2884615
## POIRE         20  0.3846154
##
##
## Classification functions coefficients:
##                KIRSCH         MIRAB         POIRE
```

```
## _CONST_ -5.0164526431 -18.840685373 -24.764879274
## MEOH     0.0034281727   0.029028464   0.033390239
## ACET     0.0063904459   0.016412816   0.007513488
## BU1     -0.0636813168   0.405389956   0.318047131
## BU2     -0.0008831867   0.071352049   0.114992814
## ISOP     0.0230821919   0.029763415  -0.008486278
## MEPR     0.0374935009  -0.128941667   0.061779984
## PRO1     0.0019711377  -0.005412661  -0.008318109
## ACAL     0.0661839107  -0.226423790  -0.130331850
```

Prédiction

```
#chargement données de prédiction
D2 <- read.xlsx("Data_LDA_R.xlsx",sheetName = "DATA_PREDICT")
str(D2)

## 'data.frame':    50 obs. of  8 variables:
##  $ MEOH: num  3 475 186 371 583 0 421 557 167 523 ...
##  $ ACET: num  15 172 101 414 226 25 142 447 86 367 ...
##  $ BU1 : num  0.2 1.9 0 1.2 2.3 0.1 1.6 0 0 2.6 ...
##  $ BU2 : num  30 7 1.6 0 19 8 8 34 0 30 ...
##  $ ISOP: num  9 113 36 97 120 0 75 107 32 116 ...
##  $ MEPR: num  9 33 11 39 46 6 24 39 10 45 ...
##  $ PRO1: num  350 546 128 502 656 253 128 162 114 787 ...
##  $ ACAL: num  9 14 8 9 11 7 31 94 8 25 ...

#prediction
pred <- predict(model=mdl,test_X=D2)
print(pred)

##  [1] "KIRSCH" "KIRSCH" "KIRSCH" "KIRSCH" "KIRSCH" "KIRSCH" "KIRSCH" "KIRSCH"
##  [9] "KIRSCH" "KIRSCH" "KIRSCH" "KIRSCH" "KIRSCH" "KIRSCH" "POIRE"  "MIRAB"
## [17] "MIRAB"  "MIRAB"  "MIRAB"  "MIRAB"  "MIRAB"  "MIRAB"  "MIRAB"  "MIRAB"
## [25] "MIRAB"  "MIRAB"  "MIRAB"  "POIRE"  "MIRAB"  "MIRAB"  "POIRE"  "MIRAB"
## [33] "MIRAB"  "MIRAB"  "POIRE"  "KIRSCH" "POIRE"  "POIRE"  "MIRAB"  "POIRE"
## [41] "POIRE"  "POIRE"  "POIRE"  "POIRE"  "POIRE"  "POIRE"  "POIRE"  "MIRAB"
## [49] "POIRE"  "POIRE"
## attr(,"class")
## [1] "pred"

#fréquence absolue
print(table(pred))

## pred
## KIRSCH  MIRAB  POIRE
##     15     19     16
```

Sélection avec “discriminR”

```
#modélisation avec sélection
mdlSel <- discriminR::lda(data=D1, target="TYPE", stepdisc=TRUE, method= "F",
pval_cut=0.01)
```

```
##
##  ************ Step  1  , forward ************
##
##      partial R2    F-value      p-value
## MEOH 0.71737116 62.1861291 3.586020e-14
## ACET 0.02814533  0.7095305 4.968583e-01
## BU1  0.71382712 61.1125850 4.862777e-14
## BU2  0.08541186  2.2880141 1.122087e-01
## ISOP 0.11226921  3.0984570 5.406192e-02
## MEPR 0.30814621 10.9121065 1.203236e-04
## PRO1 0.16453476  4.8249781 1.222491e-02
## ACAL 0.02035764  0.5091269 6.041644e-01
##
##  Variable  MEOH  will be entered
##
##  Variable(s) that have been entered :
## [1] "MEOH"
##
##  ************ Step  2  , forward ************
##
##      partial R2   F-value      p-value
## ACET 0.10265978  2.745708 7.429700e-02
## BU1  0.31872862 11.228252 9.992715e-05
## BU2  0.13632009  3.788072 2.968026e-02
## ISOP 0.06569721  1.687604 1.957508e-01
## MEPR 0.21728681  6.662572 2.795527e-03
## PRO1 0.09536500  2.530037 9.023214e-02
## ACAL 0.16605431  4.778852 1.280284e-02
##
##  Variable  BU1  will be entered
##
##  Variable(s) that have been entered :
## [1] "MEOH" "BU1"
##
##  ************ Step  3  , forward ************
##
##      partial R2  F-value     p-value
## ACET 0.07178629 1.817445 0.173671095
## BU2  0.11558499 3.071236 0.055772294
## ISOP 0.09450247 2.452583 0.097017683
## MEPR 0.23246765 7.117602 0.001994216
## PRO1 0.08541879 2.194821 0.122665772
## ACAL 0.09894342 2.580493 0.086431774
##
##  Variable  MEPR  will be entered
##
##  Variable(s) that have been entered :
## [1] "MEOH" "BU1"  "MEPR"
##
##  ************ Step  4  , forward ************
##
##      partial R2  F-value    p-value
## ACET 0.06606904 1.627088 0.20760596
## BU2  0.11034031 2.852582 0.06794383
## ISOP 0.12157018 3.183082 0.05072973
```

```
## PRO1 0.07587867 1.888507 0.16284217
## ACAL 0.13818038 3.687719 0.03270218
##
##
##  No more variables can be entered
##
##      Step Entered Removed Partial R2          F-value
## [1,] "1"  "MEOH"  ""      "0.717371161327467" "62.1861291122058"
## [2,] "2"  "BU1"   ""      "0.318728619231166" "11.2282521730405"
## [3,] "3"  "MEPR"  ""      "0.23246764713538"  "7.11760186693394"
##      P-value                Wilk's lambda
## [1,] "3.58602036953926e-14" "0.282628838672533"
## [2,] "9.99271464310336e-05" "0.192546939167529"
## [3,] "0.00199421617469675"  "0.147786005256134"
##
##  3 features selected with forward approach :
##
## [1] "MEOH" "BU1"  "MEPR"
## attr(,"class")
## [1] "Selected Features"
```

```
#affichage du modèle définitif
print(mdlSel)
```

```
##                   Count
## Total Sample Size    52
## Variables             3
## Classes               3
##
## Class Level Information :
##        frequency proportion
## KIRSCH        17  0.3269231
## MIRAB         15  0.2884615
## POIRE         20  0.3846154
##
##
## Classification functions coefficients:
##               KIRSCH        MIRAB        POIRE
## _CONST_ -3.610716965 -14.77537494 -18.37109920
## MEOH     0.006922254   0.02132093   0.02261872
## BU1     -0.076617322   0.40104435   0.37352181
## MEPR     0.086677674  -0.03247434   0.04674687
```

Sélection avec “klaR”

```
#ou le package klaR
library(klaR)
```

```
## Loading required package: MASS
```

```
##
## Attaching package: 'MASS'
```

```
## The following object is masked from 'package:discriminR':
##
##     lda

selKla <- klaR::greedy.wilks(TYPE ~ ., data=D1, niveau=0.01)
print(selKla)

## Formula containing included variables:
##
## TYPE ~ MEOH + BU1 + MEPR
## <environment: 0x0000000018e7f038>
##
##
## Values calculated in each step of the selection procedure:
##
##   vars Wilks.lambda F.statistics.overall p.value.overall F.statistics.diff
## 1 MEOH    0.2826288             62.18613    3.587597e-14         62.186129
## 2  BU1    0.1925469             30.69441    1.883938e-16         11.228252
## 3 MEPR    0.1477860             25.08637    1.401659e-17          7.117602
##   p.value.diff
## 1 3.587597e-14
## 2 9.677890e-05
## 3 1.963012e-03

#formule après sélection
print(selKla$formula)

## TYPE ~ MEOH + BU1 + MEPR
## <environment: 0x0000000018e7f038>

#extraction du sous-ensemble de données (data.frame)
D1Sel <- model.frame(selKla$formula, data = D1)
print(head(D1Sel))

##     TYPE  MEOH BU1 MEPR
## 1 KIRSCH 336.0 1.0 37.0
## 2 KIRSCH 442.0 1.9 30.0
## 3 KIRSCH 373.0 0.0 27.0
## 4 KIRSCH 418.0 0.8 24.0
## 5 KIRSCH  84.0 2.0  0.0
## 6 KIRSCH 632.5 2.2 38.2

#modélisation sur la sélection
mdlSelKla <- discriminR::lda(data=D1Sel,target="TYPE",stepdisc=FALSE)
print(mdlSelKla)

##                   Count
## Total Sample Size    52
## Variables             3
## Classes               3
##
## Class Level Information :
##        frequency proportion
## KIRSCH        17  0.3269231
## MIRAB         15  0.2884615
## POIRE         20  0.3846154
```

```
##
##
## Classification functions coefficients:
##               KIRSCH        MIRAB        POIRE
## _CONST_ -3.610716965 -14.77537494 -18.37109920
## MEOH     0.006922254   0.02132093   0.02261872
## BU1     -0.076617322   0.40104435   0.37352181
## MEPR     0.086677674  -0.03247434   0.04674687
```