Multiple_Discriminant_Analysis.R

2020-04-16

```
#Multiple Discriminant Analysis

##Calling library
library(data.table)
library(data.table) # fast file reading
library(gridExtra)  # arranging ggplot in grid
library(rmarkdown)

library(tinytex)

library(latexpdf)
library(latex2exp)

library(MASS)
#install.packages("ROCR",
lib="/Library/Frameworks/R.framework/Versions/3.5/Resources/library")
library(ROCR)

## Attaching package: 'gplots'

## The following object is masked from 'package:stats':
##
##     lowess

bank=read.csv("C:/Users/Deepal/Desktop/MVA/bank.csv")
#View(bank)

#Convert the data frame to data table
setDT(bank)
#Describe the columns and their data types
str(bank)

## Classes 'data.table' and 'data.frame':   11162 obs. of  17 variables:
##  $ age      : int  59 56 41 55 54 42 56 60 37 28 ...
##  $ job      : Factor w/ 12 levels "admin.","blue-collar",..: 1 1 10 8 1 5
5 6 10 8 ...
##  $ marital  : Factor w/ 3 levels "divorced","married",..: 2 2 2 2 2 3 2 1
2 3 ...
##  $ education: Factor w/ 4 levels "primary","secondary",..: 2 2 2 2 3 3 3 2
2 2 ...
##  $ default  : Factor w/ 2 levels "no","yes": 1 1 1 1 1 1 1 1 1 1 ...
##  $ balance  : int  2343 45 1270 2476 184 0 830 545 1 5090 ...
##  $ housing  : Factor w/ 2 levels "no","yes": 2 1 2 2 1 2 2 2 2 2 ...
##  $ loan     : Factor w/ 2 levels "no","yes": 1 1 1 1 1 2 2 1 1 1 ...
##  $ contact  : Factor w/ 3 levels "cellular","telephone",..: 3 3 3 3 3 3 3
```

```
3 3 3 ...
##  $ day      : int  5 5 5 5 5 5 6 6 6 6 ...
##  $ month    : Factor w/ 12 levels "apr","aug","dec",..: 9 9 9 9 9 9 9 9 9
9 ...
##  $ duration : int  1042 1467 1389 579 673 562 1201 1030 608 1297 ...
##  $ campaign : int  1 1 1 1 2 2 1 1 1 3 ...
##  $ pdays    : int  -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 ...
##  $ previous : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ poutcome : Factor w/ 4 levels "failure","other",..: 4 4 4 4 4 4 4 4 4 4
...
##  $ deposit  : Factor w/ 2 levels "no","yes": 2 2 2 2 2 2 2 2 2 ...
##  - attr(*, ".internal.selfref")=<externalptr>

#By head we get to know first n rows to get grasp of the data
head(bank)

##    age        job marital education default balance housing loan contact
## 1:  59     admin. married secondary      no    2343     yes   no unknown
## 2:  56     admin. married secondary      no      45      no   no unknown
## 3:  41 technician married secondary      no    1270     yes   no unknown
## 4:  55   services married secondary      no    2476     yes   no unknown
## 5:  54     admin. married  tertiary      no     184      no   no unknown
## 6:  42 management  single  tertiary      no       0     yes  yes unknown
##    day month duration campaign pdays previous poutcome deposit
## 1:   5   may     1042        1    -1        0  unknown     yes
## 2:   5   may     1467        1    -1        0  unknown     yes
## 3:   5   may     1389        1    -1        0  unknown     yes
## 4:   5   may      579        1    -1        0  unknown     yes
## 5:   5   may      673        2    -1        0  unknown     yes
## 6:   5   may      562        2    -1        0  unknown     yes

#Find NA In the data table.

table(is.na(bank))

##
##  FALSE
## 189754

#Find NA in Columns.
bank[is.na(age),NROW(age)]

## [1] 0

bank[is.na(job),NROW(job)]

## [1] 0

bank[is.na(education),NROW(education)]

## [1] 0
```

```r
grep('NA',bank)
```

```
## integer(0)
```

```r
#Find different elements in the column
unique(bank$job)
```

```
##  [1] admin.        technician    services      management    retired
##  [6] blue-collar   unemployed    entrepreneur  housemaid     unknown
## [11] self-employed student
## 12 Levels: admin. blue-collar entrepreneur housemaid ... unknown
```

```r
unique(bank$marital)
```

```
## [1] married  single   divorced
## Levels: divorced married single
```

```r
#Summary of dataset
summary(bank)
```

```
##       age                 job           marital         education
##  Min.   :18.00   management :2566   divorced:1293   primary  :1500
##  1st Qu.:32.00   blue-collar:1944   married :6351   secondary:5476
##  Median :39.00   technician :1823   single  :3518   tertiary :3689
##  Mean   :41.23   admin.     :1334                   unknown  : 497
##  3rd Qu.:49.00   services   : 923
##  Max.   :95.00   retired    : 778
##                  (Other)    :1794
##  default        balance      housing     loan           contact
##  no :10994   Min.   :-6847   no :5881   no :9702   cellular :8042
##  yes:  168   1st Qu.:  122   yes:5281   yes:1460   telephone: 774
##              Median :  550                         unknown  :2346
##              Mean   : 1529
##              3rd Qu.: 1708
##              Max.   :81204
##
##       day            month         duration       campaign
##  Min.   : 1.00   may    :2824   Min.   :   2   Min.   : 1.000
##  1st Qu.: 8.00   aug    :1519   1st Qu.: 138   1st Qu.: 1.000
##  Median :15.00   jul    :1514   Median : 255   Median : 2.000
##  Mean   :15.66   jun    :1222   Mean   : 372   Mean   : 2.508
##  3rd Qu.:22.00   nov    : 943   3rd Qu.: 496   3rd Qu.: 3.000
##  Max.   :31.00   apr    : 923   Max.   :3881   Max.   :63.000
##                  (Other):2217
##      pdays          previous          poutcome    deposit
##  Min.   : -1.00   Min.   : 0.0000   failure:1228   no :5873
##  1st Qu.: -1.00   1st Qu.: 0.0000   other  : 537   yes:5289
##  Median : -1.00   Median : 0.0000   success:1071
##  Mean   : 51.33   Mean   : 0.8326   unknown:8326
##  3rd Qu.: 20.75   3rd Qu.: 1.0000
```

```
##  Max.   :854.00   Max.   :58.0000
##
```

```
#Take sample of 1000 from the dataset.
bankdata=bank[sample(.N,1000)]
bank=bankdata
#View(bank)
dim(bank)
```

```
## [1] 1000   17
```

```
#we need to convert to matrix to facilitate distance measurement
bank.data <- as.matrix(bank[,c(0:16)])
dim(bank.data)
```

```
## [1] 1000   16
```

```
dim(bank)
```

```
## [1] 1000   17
```

```
bank_raw <- cbind(bank.data, as.numeric(bank$deposit)-1)
dim(bank_raw)
```

```
## [1] 1000   17
```

```
colnames(bank_raw)[17] <- "deposit"

#View(bank_raw)
# Lets cut the data into two parts
smp_size_raw <- floor(0.75 * nrow(bank_raw))
train_ind_raw <- sample(nrow(bank_raw), size = smp_size_raw)
train_raw.df <- as.data.frame(bank_raw[train_ind_raw, ])
test_raw.df <- as.data.frame(bank_raw[-train_ind_raw, ])
# We now have a training and a test set. Training is 75% and test is 25%

bank_raw.lda <- lda(formula = train_raw.df$deposit ~
default+loan+job+marital+age+education+housing, data = train_raw.df)
bank_raw.lda
```

```
## Call:
## lda(train_raw.df$deposit ~ default + loan + job + marital + age +
##     education + housing, data = train_raw.df)
##
## Prior probabilities of groups:
##     0     1
## 0.496 0.504
##
## Group means:
##   defaultyes   loanyes jobblue-collar jobentrepreneur jobhousemaid
## 0 0.01344086 0.1639785      0.2177419      0.02419355   0.03225806
## 1 0.01058201 0.1084656      0.1375661      0.02116402   0.01587302
```

```
##   jobmanagement jobretired jobself-employed jobservices jobstudent
## 0     0.2634409 0.03763441       0.03763441  0.08870968 0.01344086
## 1     0.2486772 0.08994709       0.02645503  0.09259259 0.04761905
##   jobtechnician jobunemployed  jobunknown maritalmarried maritalsingle
## 0     0.1505376    0.02419355 0.002688172      0.6263441     0.2741935
## 1     0.1613757    0.03174603 0.005291005      0.5052910     0.3650794
##         age19       age20       age21       age22       age23       age24
## 0 0.000000000 0.000000000 0.002688172 0.002688172 0.008064516 0.002688172
## 1 0.002645503 0.005291005 0.000000000 0.005291005 0.007936508 0.010582011
##         age25      age26      age27      age28      age29      age30
## 0 0.005376344 0.01344086 0.01881720 0.02150538 0.03225806 0.06451613
## 1 0.010582011 0.01851852 0.02380952 0.03703704 0.03439153 0.04232804
##        age31      age32      age33      age34      age35      age36
## 0 0.05107527 0.03763441 0.04301075 0.04032258 0.03763441 0.03763441
## 1 0.03174603 0.05026455 0.05026455 0.02910053 0.03439153 0.03703704
##        age37      age38      age39      age40      age41      age42
## 0 0.04301075 0.01881720 0.02419355 0.03494624 0.03763441 0.04301075
## 1 0.04761905 0.02645503 0.02910053 0.02645503 0.03439153 0.01587302
##        age43      age44       age45       age46      age47      age48
## 0 0.02150538 0.01881720 0.010752688 0.013440860 0.03225806 0.02688172
## 1 0.01587302 0.01851852 0.007936508 0.007936508 0.01851852 0.02645503
##        age49       age50       age51      age52      age53      age54
## 0 0.02688172 0.008064516 0.026881720 0.02150538 0.02688172 0.01075269
## 1 0.02645503 0.015873016 0.005291005 0.01587302 0.02116402 0.01587302
##        age55      age56      age57       age58      age59      age60
## 0 0.04301075 0.02150538 0.01344086 0.008064516 0.01881720 0.01612903
## 1 0.01058201 0.01851852 0.01322751 0.023809524 0.01851852 0.01058201
##         age61       age62       age63       age64       age65       age66
## 0 0.000000000 0.000000000 0.000000000 0.000000000 0.000000000 0.000000000
## 1 0.007936508 0.005291005 0.002645503 0.007936508 0.005291005 0.007936508
##         age67       age68       age69       age70       age71       age72
## 0 0.002688172 0.000000000 0.000000000 0.002688172 0.000000000 0.002688172
## 1 0.000000000 0.002645503 0.005291005 0.002645503 0.005291005 0.005291005
##        age73       age74       age75       age76       age77       age80
## 0 0.00000000 0.000000000 0.002688172 0.002688172 0.000000000 0.000000000
## 1 0.01058201 0.002645503 0.000000000 0.002645503 0.005291005 0.007936508
##         age84       age86 educationsecondary educationtertiary
## 0 0.000000000 0.000000000          0.4677419         0.3172043
## 1 0.002645503 0.002645503          0.4523810         0.3677249
##   educationunknown housingyes
## 0       0.04301075  0.5403226
## 1       0.05555556  0.3835979
##
## Coefficients of linear discriminants:
##                              LD1
## defaultyes         -0.05931809
## loanyes            -0.35278046
## jobblue-collar     -0.46799478
## jobentrepreneur    -0.98020065
## jobhousemaid       -1.63084077
```

```
## jobmanagement      -0.69243836
## jobretired         -0.92940385
## jobself-employed   -0.77520619
## jobservices        -0.00788011
## jobstudent          0.30871584
## jobtechnician      -0.25414123
## jobunemployed       0.53559616
## jobunknown         -0.70596420
## maritalmarried     -0.66275699
## maritalsingle      -0.14770753
## age19               0.20210218
## age20               1.07544567
## age21              -5.08456210
## age22              -1.57974834
## age23              -2.05842948
## age24              -0.16825404
## age25              -0.67497694
## age26              -1.28032111
## age27              -1.35151412
## age28              -0.86250344
## age29              -1.39454582
## age30              -2.06591216
## age31              -2.07963865
## age32              -0.95749501
## age33              -1.14196134
## age34              -1.82248295
## age35              -1.44966575
## age36              -1.47026421
## age37              -1.07988450
## age38              -0.71014165
## age39              -0.79430625
## age40              -1.66408450
## age41              -1.27598501
## age42              -2.71016493
## age43              -1.86097916
## age44              -1.04635757
## age45              -1.96847295
## age46              -2.28775591
## age47              -1.80521681
## age48              -1.19925495
## age49              -1.24639360
## age50              -0.33599589
## age51              -2.81904848
## age52              -1.42868342
## age53              -1.55969912
## age54              -0.59578437
## age55              -3.04870891
## age56              -1.54559866
## age57              -1.29493150
## age58              -0.13453693
```

```
## age59               -0.82897419
## age60               -1.72920903
## age61                1.69078920
## age62                2.09231530
## age63                2.00606813
## age64                1.71962967
## age65                1.43242227
## age66                2.09271012
## age67               -4.02287657
## age68                1.58038366
## age69                1.43242227
## age70               -0.89389469
## age71                1.58038366
## age72                0.15178375
## age73                2.17117332
## age74                1.95527133
## age75               -4.31074590
## age76               -0.87455473
## age77                1.88643507
## age80                2.28504007
## age84                1.58038366
## age86                1.58038366
## educationsecondary   0.28786932
## educationtertiary    0.67081045
## educationunknown     0.48997151
## housingyes          -0.65702909
```

```
summary(bank_raw.lda)
```

```
##         Length Class  Mode
## prior      2    -none- numeric
## counts     2    -none- numeric
## means    162    -none- numeric
## scaling   81    -none- numeric
## lev        2    -none- character
## svd        1    -none- numeric
## N          1    -none- numeric
## call       3    -none- call
## terms      3    terms  call
## xlevels    7    -none- list
```

```
print(bank_raw.lda)
```

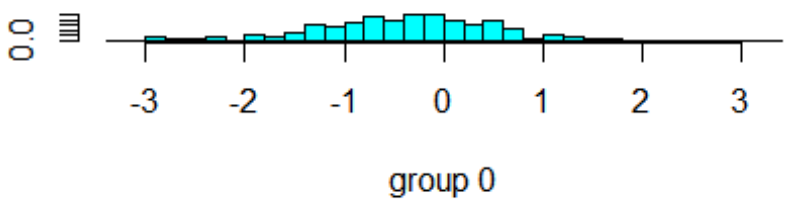
```
## Call:
## lda(train_raw.df$deposit ~ default + loan + job + marital + age +
##     education + housing, data = train_raw.df)
##
## Prior probabilities of groups:
##     0     1
## 0.496 0.504
##
```

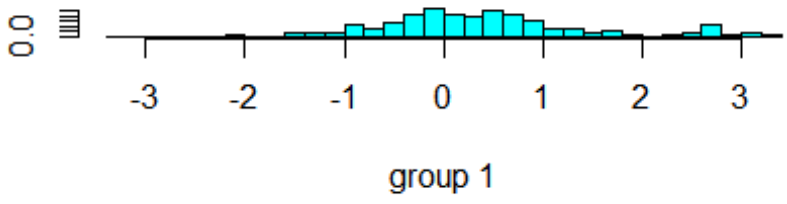
```
## Group means:
##   defaultyes   loanyes jobblue-collar jobentrepreneur jobhousemaid
## 0 0.01344086 0.1639785      0.2177419      0.02419355   0.03225806
## 1 0.01058201 0.1084656      0.1375661      0.02116402   0.01587302
##   jobmanagement jobretired jobself-employed jobservices jobstudent
## 0     0.2634409 0.03763441       0.03763441  0.08870968 0.01344086
## 1     0.2486772 0.08994709       0.02645503  0.09259259 0.04761905
##   jobtechnician jobunemployed  jobunknown maritalmarried maritalsingle
## 0     0.1505376    0.02419355 0.002688172      0.6263441     0.2741935
## 1     0.1613757    0.03174603 0.005291005      0.5052910     0.3650794
##         age19       age20       age21       age22       age23       age24
## 0 0.000000000 0.000000000 0.002688172 0.002688172 0.008064516 0.002688172
## 1 0.002645503 0.005291005 0.000000000 0.005291005 0.007936508 0.010582011
##         age25      age26      age27      age28      age29      age30
## 0 0.005376344 0.01344086 0.01881720 0.02150538 0.03225806 0.06451613
## 1 0.010582011 0.01851852 0.02380952 0.03703704 0.03439153 0.04232804
##        age31      age32      age33      age34      age35      age36
## 0 0.05107527 0.03763441 0.04301075 0.04032258 0.03763441 0.03763441
## 1 0.03174603 0.05026455 0.05026455 0.02910053 0.03439153 0.03703704
##        age37      age38      age39      age40      age41      age42
## 0 0.04301075 0.01881720 0.02419355 0.03494624 0.03763441 0.04301075
## 1 0.04761905 0.02645503 0.02910053 0.02645503 0.03439153 0.01587302
##        age43      age44       age45       age46      age47      age48
## 0 0.02150538 0.01881720 0.010752688 0.013440860 0.03225806 0.02688172
## 1 0.01587302 0.01851852 0.007936508 0.007936508 0.01851852 0.02645503
##        age49       age50       age51      age52      age53      age54
## 0 0.02688172 0.008064516 0.026881720 0.02150538 0.02688172 0.01075269
## 1 0.02645503 0.015873016 0.005291005 0.01587302 0.02116402 0.01587302
##        age55      age56      age57       age58      age59      age60
## 0 0.04301075 0.02150538 0.01344086 0.008064516 0.01881720 0.01612903
## 1 0.01058201 0.01851852 0.01322751 0.023809524 0.01851852 0.01058201
##         age61       age62       age63       age64       age65       age66
## 0 0.000000000 0.000000000 0.000000000 0.000000000 0.000000000 0.000000000
## 1 0.007936508 0.005291005 0.002645503 0.007936508 0.005291005 0.007936508
##         age67       age68       age69       age70       age71       age72
## 0 0.002688172 0.000000000 0.000000000 0.002688172 0.000000000 0.002688172
## 1 0.000000000 0.002645503 0.005291005 0.002645503 0.005291005 0.005291005
##        age73       age74       age75       age76       age77       age80
## 0 0.00000000 0.000000000 0.002688172 0.002688172 0.000000000 0.000000000
## 1 0.01058201 0.002645503 0.000000000 0.002645503 0.005291005 0.007936508
##         age84       age86 educationsecondary educationtertiary
## 0 0.000000000 0.000000000          0.4677419         0.3172043
## 1 0.002645503 0.002645503          0.4523810         0.3677249
##   educationunknown housingyes
## 0       0.04301075  0.5403226
## 1       0.05555556  0.3835979
##
## Coefficients of linear discriminants:
##                             LD1
## defaultyes         -0.05931809
```

```
## loanyes            -0.35278046
## jobblue-collar     -0.46799478
## jobentrepreneur    -0.98020065
## jobhousemaid       -1.63084077
## jobmanagement      -0.69243836
## jobretired         -0.92940385
## jobself-employed   -0.77520619
## jobservices        -0.00788011
## jobstudent          0.30871584
## jobtechnician      -0.25414123
## jobunemployed       0.53559616
## jobunknown         -0.70596420
## maritalmarried     -0.66275699
## maritalsingle      -0.14770753
## age19               0.20210218
## age20               1.07544567
## age21              -5.08456210
## age22              -1.57974834
## age23              -2.05842948
## age24              -0.16825404
## age25              -0.67497694
## age26              -1.28032111
## age27              -1.35151412
## age28              -0.86250344
## age29              -1.39454582
## age30              -2.06591216
## age31              -2.07963865
## age32              -0.95749501
## age33              -1.14196134
## age34              -1.82248295
## age35              -1.44966575
## age36              -1.47026421
## age37              -1.07988450
## age38              -0.71014165
## age39              -0.79430625
## age40              -1.66408450
## age41              -1.27598501
## age42              -2.71016493
## age43              -1.86097916
## age44              -1.04635757
## age45              -1.96847295
## age46              -2.28775591
## age47              -1.80521681
## age48              -1.19925495
## age49              -1.24639360
## age50              -0.33599589
## age51              -2.81904848
## age52              -1.42868342
## age53              -1.55969912
## age54              -0.59578437
```

```
## age55               -3.04870891
## age56               -1.54559866
## age57               -1.29493150
## age58               -0.13453693
## age59               -0.82897419
## age60               -1.72920903
## age61                1.69078920
## age62                2.09231530
## age63                2.00606813
## age64                1.71962967
## age65                1.43242227
## age66                2.09271012
## age67               -4.02287657
## age68                1.58038366
## age69                1.43242227
## age70               -0.89389469
## age71                1.58038366
## age72                0.15178375
## age73                2.17117332
## age74                1.95527133
## age75               -4.31074590
## age76               -0.87455473
## age77                1.88643507
## age80                2.28504007
## age84                1.58038366
## age86                1.58038366
## educationsecondary   0.28786932
## educationtertiary    0.67081045
## educationunknown     0.48997151
## housingyes          -0.65702909

plot(bank_raw.lda)
```






```
bank_raw.lda.predict <- lda(formula = train_raw.df$deposit ~
default+loan+job+marital+education+housing, data = train_raw.df)
bank_raw.lda.predict$class

## NULL

bank_raw.lda.predict$x

## $default
## [1] "no"  "yes"
##
## $loan
## [1] "no"  "yes"
##
## $job
##  [1] "admin."        "blue-collar"   "entrepreneur"  "housemaid"
##  [5] "management"    "retired"       "self-employed" "services"
##  [9] "student"       "technician"    "unemployed"    "unknown"
##
## $marital
## [1] "divorced" "married"  "single"
##
## $education
## [1] "primary"   "secondary" "tertiary"  "unknown"
##
## $housing
## [1] "no"  "yes"
```

```r
# Get the deposit as a dataframe.
bank_raw.lda.predict.deposit <- as.data.frame(bank_raw.lda.predict$deposit)
dim(bank_raw.lda.predict.deposit)
```

```
## [1] 0 0
```

```r
# As LDA predict gives output as NULL, we can't further apply LDA on our
dataset.
```