Untitled

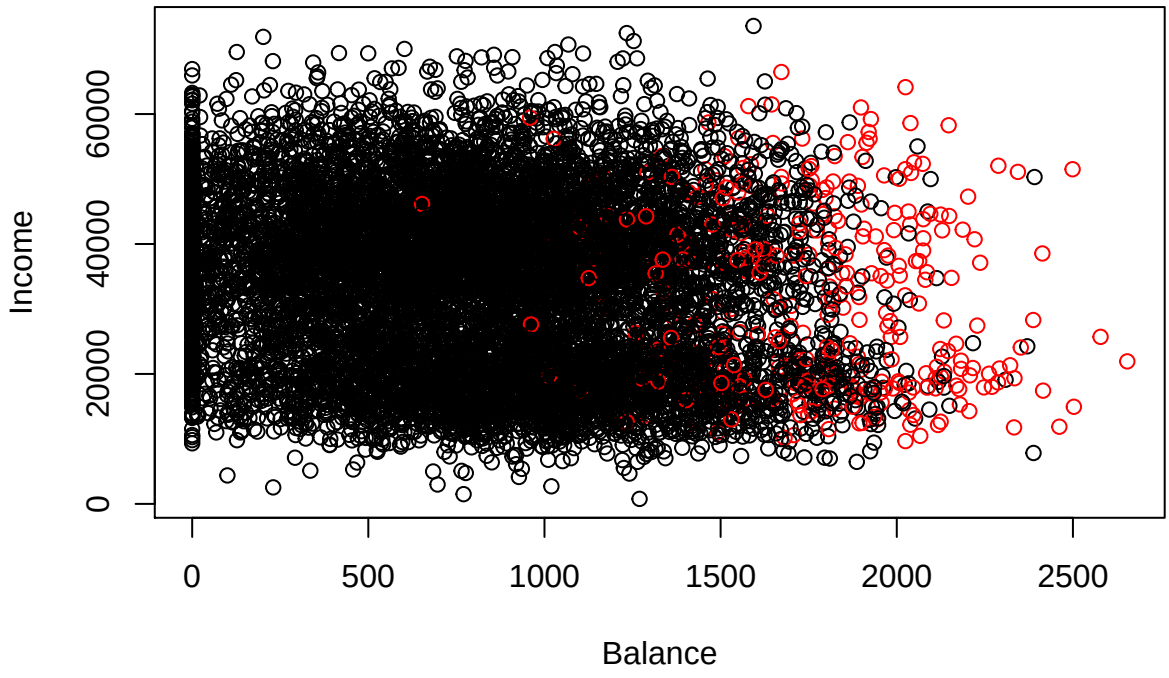
Luc A.

19/10/2016

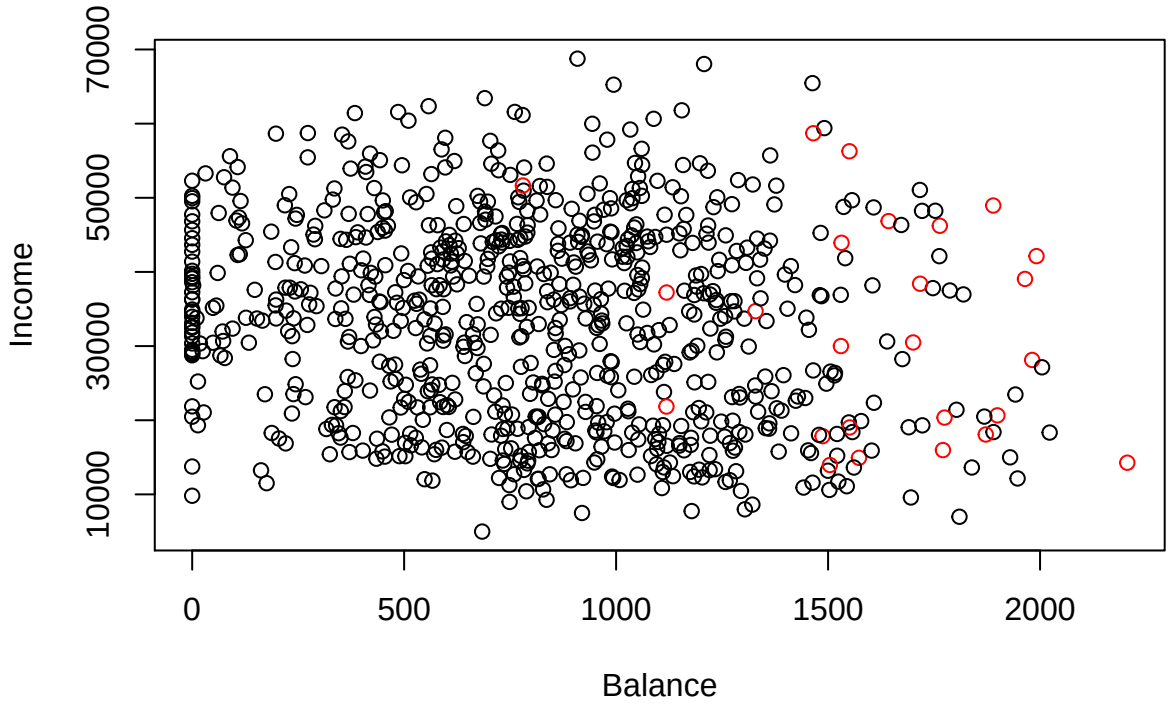
```
# Classification : Logistique-Lda-Qda
# Données : sur les défauts de paiement
rm(list=ls())
library(ISLR)
attach(Default)
str(Default)
```

```
## 'data.frame':    10000 obs. of  4 variables:
##  $ default: Factor w/ 2 levels "No","Yes": 1 1 1 1 1 1 1 1 1 1 ...
##  $ student: Factor w/ 2 levels "No","Yes": 1 2 1 1 1 2 1 2 1 1 ...
##  $ balance: num  730 817 1074 529 786 ...
##  $ income : num  44362 12106 31767 35704 38463 ...
```

```
# données d'entraînement
don_entr<-Default[1:800,]
# données de test
don_test<-Default[801:1000,]
#Graphique
plot(balance,income,col=c(default),xlab="Balance",ylab="Income")
```




```
#
plot(don_entr$balance,don_entr$income,col=c(don_entr$default),xlab="Balance",ylab="Income")
```




```
#
fit.log<-glm(default~balance+income, family=binomial,data=don_entr)
summary(fit.log)
```

```
##
## Call:
## glm(formula = default ~ balance + income, family = binomial,
##     data = don_entr)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.3623  -0.1635  -0.0702  -0.0292   3.3117
##
## Coefficients:
##               Estimate Std. Error z value Pr(>|z|)
## (Intercept) -1.112e+01  1.461e+00  -7.607 2.81e-14 ***
## balance      5.380e-03  7.699e-04   6.988 2.79e-12 ***
## income       2.788e-05  1.717e-05   1.624    0.104
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 222.50  on 799  degrees of freedom
## Residual deviance: 133.48  on 797  degrees of freedom
## AIC: 139.48
##
## Number of Fisher Scoring iterations: 8
```

```r
# création d'un ensemble de valeurs pour les prédictions
# newd<-data.frame(balance=c(1000))
# prev1<-predict(fit.g,newd,type="response",se.fit=TRUE)
prev_log<-predict(fit.log,don_test,type="response")
#
# conversion des probabilités : prob>0.5  <=> Defaut
#
test_y<-don_test$default
prev_y<-rep("No",length(test_y))
prev_y[prev_log > 0.1] = "Yes"
table(prev_y)
```

```
## prev_y
##  No Yes
## 186  14
```

```r
# Matrice de confusion
table(prev_y,test_y)
```

```
##       test_y
## prev_y  No Yes
##    No  185   1
##    Yes   7   7
```

```r
# taux d'erreur de la classification par regression logistique
taux_err_log = mean(prev_y != test_y)
#
#
### LDA
library(MASS)
fit.lda <-lda(default~balance+income, data=don_entr)
prev_lda <-predict(fit.lda,don_test)
prev_lda_y <-prev_lda$class
# Matrice de confusion
table(prev_lda_y,test_y)
```

```
##           test_y
## prev_lda_y  No Yes
##        No  191   5
##        Yes   1   3
```

```r
# taux d'erreur de la classification par regression logistique
taux_err_lda <- mean(prev_lda_y != test_y)
#
## QDA ###
#
fit.qda<-qda(default~balance+income, data=don_entr)
prev_qda <-predict(fit.qda,don_test)
prev_qda_y <-prev_qda$class
# Matrice de confusion
table(prev_qda_y,test_y)
```

```
##           test_y
## prev_qda_y  No Yes
##        No  191   5
##        Yes   1   3
```

```r
# taux d'erreur de la classification par regression logistique
taux_err_qda <- mean(prev_qda_y != test_y)
#
```