Trabajo3

Juan Alberto Martinez Lopez / Alberto Armijo Ru?z

27 de mayo de 2017

```
#Librerías utilizadas.
library("caret")
```

```
## Loading required package: lattice
```

```
## Loading required package: ggplot2
```

```
library("leaps")
library("e1071")
```

1. default of credit card clients Data Set (Clasificaci?n)

La base de datos se centra en el caso de los pagos por defecto de los clientes en Taiwán y compara la precisión predictiva de la probabilidad de incumplimiento entre seis métodos de minería de datos. Desde la perspectiva de la gestión de riesgos, el resultado de la precisión predictiva de la probabilidad estimada de incumplimiento ser más valioso que el resultado binario de la clasificación - clientes creíbles o no creíbles. Con la probabilidad real de default como variable de respuesta (Y), y la probabilidad predictiva de default como variable independiente (X), el resultado de regresión lineal simple (Y = A + BX) muestra que el modelo de predicción producido por la red neuronal artificial tiene el mayor coeficiente de determinación. Su intercepción de regresión (A) es cercana a cero, y el coeficiente de regresión (B) a uno. Por lo tanto, entre las seis técnicas de minería de datos, la red neuronal artificial es la única que puede estimar con precisión la probabilidad real de incumplimiento. Dichas caracteristicas son:

SBP: Presión arterial Tobbaco: El tabaco acululado en kg. LDL: Lipoproteina de baja densidad. Adiposity: Acumulacion excesiva de grasa en el organismo. Famhist: Historia familiar sobre casos de enfermedades cardiacas. Typea: Personalidad tipo A. Obesity: Obesidad. Alcohol Consumo de alcohol actual. Age: Edad. Chd: Respuesta a la enfermedad coronaria (la variable que queremos aprender).

```
credit_card =  read.csv("default_of_credict_card_clients.csv"
                        , sep=",", header = TRUE, row.names =1)

summary(credit_card)
```

```
##    LIMIT_BAL          SEX            EDUCATION        MARRIAGE
##  Min.   :  10000   Min.   :1.000   Min.   :0.000   Min.   :0.000
##  1st Qu.:  50000   1st Qu.:1.000   1st Qu.:1.000   1st Qu.:1.000
##  Median : 140000   Median :2.000   Median :2.000   Median :2.000
##  Mean   : 167484   Mean   :1.604   Mean   :1.853   Mean   :1.552
##  3rd Qu.: 240000   3rd Qu.:2.000   3rd Qu.:2.000   3rd Qu.:2.000
##  Max.   :1000000   Max.   :2.000   Max.   :6.000   Max.   :3.000
##       AGE            PAY_0             PAY_2             PAY_3
##  Min.   :21.00   Min.   :-2.0000   Min.   :-2.0000   Min.   :-2.0000
##  1st Qu.:28.00   1st Qu.:-1.0000   1st Qu.:-1.0000   1st Qu.:-1.0000
##  Median :34.00   Median : 0.0000   Median : 0.0000   Median : 0.0000
##  Mean   :35.49   Mean   :-0.0167   Mean   :-0.1338   Mean   :-0.1662
##  3rd Qu.:41.00   3rd Qu.: 0.0000   3rd Qu.: 0.0000   3rd Qu.: 0.0000
##  Max.   :79.00   Max.   : 8.0000   Max.   : 8.0000   Max.   : 8.0000
##      PAY_4             PAY_5             PAY_6           BILL_AMT1
```

```
##  Min.   :-2.0000   Min.   :-2.0000   Min.   :-2.0000   Min.   :-165580
##  1st Qu.:-1.0000   1st Qu.:-1.0000   1st Qu.:-1.0000   1st Qu.:   3559
##  Median : 0.0000   Median : 0.0000   Median : 0.0000   Median :  22382
##  Mean   :-0.2207   Mean   :-0.2662   Mean   :-0.2911   Mean   :  51223
##  3rd Qu.: 0.0000   3rd Qu.: 0.0000   3rd Qu.: 0.0000   3rd Qu.:  67091
##  Max.   : 8.0000   Max.   : 8.0000   Max.   : 8.0000   Max.   : 964511
##    BILL_AMT2         BILL_AMT3         BILL_AMT4         BILL_AMT5
##  Min.   :-69777   Min.   :-157264   Min.   :-170000   Min.   :-81334
##  1st Qu.:  2985   1st Qu.:   2666   1st Qu.:   2327   1st Qu.:  1763
##  Median : 21200   Median :  20089   Median :  19052   Median : 18105
##  Mean   : 49179   Mean   :  47013   Mean   :  43263   Mean   : 40311
##  3rd Qu.: 64006   3rd Qu.:  60165   3rd Qu.:  54506   3rd Qu.: 50191
##  Max.   :983931   Max.   :1664089   Max.   : 891586   Max.   :927171
##    BILL_AMT6          PAY_AMT1          PAY_AMT2          PAY_AMT3
##  Min.   :-339603   Min.   :     0   Min.   :      0   Min.   :     0
##  1st Qu.:   1256   1st Qu.:  1000   1st Qu.:    833   1st Qu.:   390
##  Median :  17071   Median :  2100   Median :   2009   Median :  1800
##  Mean   :  38872   Mean   :  5664   Mean   :   5921   Mean   :  5226
##  3rd Qu.:  49198   3rd Qu.:  5006   3rd Qu.:   5000   3rd Qu.:  4505
##  Max.   : 961664   Max.   :873552   Max.   :1684259   Max.   :896040
##     PAY_AMT4          PAY_AMT5          PAY_AMT6
##  Min.   :     0   Min.   :     0.0   Min.   :     0.0
##  1st Qu.:   296   1st Qu.:   252.5   1st Qu.:   117.8
##  Median :  1500   Median :  1500.0   Median :  1500.0
##  Mean   :  4826   Mean   :  4799.4   Mean   :  5215.5
##  3rd Qu.:  4013   3rd Qu.:  4031.5   3rd Qu.:  4000.0
##  Max.   :621000   Max.   :426529.0   Max.   :528666.0
##  default.payment.next.month
##  Min.   :0.0000
##  1st Qu.:0.0000
##  Median :0.0000
##  Mean   :0.2212
##  3rd Qu.:0.0000
##  Max.   :1.0000
```

```
set.seed(1)
train = sample (nrow(credit_card), round(nrow(credit_card)*0.7))
credit_card.train = credit_card[train,]
credit_card.test = credit_card[-train,]
```

2. Preprocesado de los datos.

Lo primero que queremos hacer es comprobar si hay datos perdidos, y si es así; reemplazaremos el valor perdido.

```
anyNA(credit_card.train)
```

```
## [1] FALSE
```

Como no tenemos ningún dato pérdido, no tendremos que reemplazar los valores. Si hubieramos tenido valores tenido, podríamos haber utilizado la función *knnImputation* para reemplazar los valores perdidos por los k vecinos mÃ¡s cercanos (normalmente k=3). TambiÃ©n podrÃ?amos utilizar la media como sustituto del valor pÃ©rdido.

Lo siguiente que vamos a hacer es modificar aquellas columnas que separan los datos en variables "clases"

como por ejemplo la columna *EDUCATION*, que indica que tipo de estudios tiene cada persona. Por cada tipo en los que los separe, crearemos una nueva columna que indique con 0s y 1s la pertenencia a ese tipo. TambiÃ©n tenemos que realizar este proceso con la columna *MARRIAGE*

```
# Modificamos la columna 2, llamada sex, para dividir los datos en 0=mujer, 1=hombre.
credit_card.train$SEX = ifelse(credit_card.train$SEX == 2, 0, 1)
summary(credit_card.train)
```

```
##    LIMIT_BAL            SEX           EDUCATION        MARRIAGE
##  Min.   :  10000   Min.   :0.0000   Min.   :0.000   Min.   :0.000
##  1st Qu.:  50000   1st Qu.:0.0000   1st Qu.:1.000   1st Qu.:1.000
##  Median : 140000   Median :0.0000   Median :2.000   Median :2.000
##  Mean   : 167948   Mean   :0.3974   Mean   :1.853   Mean   :1.549
##  3rd Qu.: 240000   3rd Qu.:1.0000   3rd Qu.:2.000   3rd Qu.:2.000
##  Max.   :1000000   Max.   :1.0000   Max.   :6.000   Max.   :3.000
##       AGE            PAY_0              PAY_2              PAY_3
##  Min.   :21.0   Min.   :-2.00000   Min.   :-2.0000   Min.   :-2.0000
##  1st Qu.:28.0   1st Qu.:-1.00000   1st Qu.:-1.0000   1st Qu.:-1.0000
##  Median :34.0   Median : 0.00000   Median : 0.0000   Median : 0.0000
##  Mean   :35.5   Mean   :-0.01524   Mean   :-0.1321   Mean   :-0.1697
##  3rd Qu.:41.0   3rd Qu.: 0.00000   3rd Qu.: 0.0000   3rd Qu.: 0.0000
##  Max.   :79.0   Max.   : 8.00000   Max.   : 8.0000   Max.   : 8.0000
##      PAY_4              PAY_5              PAY_6            BILL_AMT1
##  Min.   :-2.000   Min.   :-2.0000   Min.   :-2.000   Min.   :-154973
##  1st Qu.:-1.000   1st Qu.:-1.0000   1st Qu.:-1.000   1st Qu.:   3508
##  Median : 0.000   Median : 0.0000   Median : 0.000   Median :  22143
##  Mean   :-0.221   Mean   :-0.2678   Mean   :-0.289   Mean   :  51238
##  3rd Qu.: 0.000   3rd Qu.: 0.0000   3rd Qu.: 0.000   3rd Qu.:  66175
##  Max.   : 8.000   Max.   : 8.0000   Max.   : 8.000   Max.   : 964511
##    BILL_AMT2          BILL_AMT3          BILL_AMT4          BILL_AMT5
##  Min.   :-69777   Min.   : -46127   Min.   :-170000   Min.   :-81334
##  1st Qu.:  2956   1st Qu.:   2552   1st Qu.:   2276   1st Qu.:  1730
##  Median : 20895   Median :  20012   Median :  19044   Median : 18093
##  Mean   : 49092   Mean   :  47104   Mean   :  43452   Mean   : 40395
##  3rd Qu.: 62795   3rd Qu.:  59674   3rd Qu.:  53536   3rd Qu.: 49719
##  Max.   :983931   Max.   :1664089   Max.   : 891586   Max.   :927171
##    BILL_AMT6           PAY_AMT1          PAY_AMT2            PAY_AMT3
##  Min.   :-339603   Min.   :     0   Min.   :      0.0   Min.   :     0
##  1st Qu.:   1261   1st Qu.:  1000   1st Qu.:    809.8   1st Qu.:   390
##  Median :  17036   Median :  2100   Median :   2000.0   Median :  1794
##  Mean   :  38948   Mean   :  5608   Mean   :   5988.9   Mean   :  5218
##  3rd Qu.:  48740   3rd Qu.:  5006   3rd Qu.:   5000.0   3rd Qu.:  4525
##  Max.   : 961664   Max.   :493358   Max.   :1684259.0   Max.   :896040
##     PAY_AMT4          PAY_AMT5           PAY_AMT6
##  Min.   :     0   Min.   :     0.0   Min.   :     0
##  1st Qu.:   300   1st Qu.:   269.8   1st Qu.:   100
##  Median :  1500   Median :  1500.0   Median :  1500
##  Mean   :  4793   Mean   :  4838.0   Mean   :  5252
##  3rd Qu.:  4010   3rd Qu.:  4037.8   3rd Qu.:  4000
##  Max.   :528897   Max.   :426529.0   Max.   :528666
##  default.payment.next.month
##  Min.   :0.0000
##  1st Qu.:0.0000
##  Median :0.0000
##  Mean   :0.2209
```

```
##  3rd Qu.:0.0000
##  Max.   :1.0000
```

```r
# TambiÃ©n tenemos que modificar la columna EDUCATION, la dividiremos en cuatro columnas diferentes:
# ed.other, ed.university, ed.high_school, ed.school
ed.other = ifelse(credit_card.train$EDUCATION == 4, 1,0)
ed.university = ifelse(credit_card.train$EDUCATION == 2, 1, 0)
ed.high_school = ifelse(credit_card.train$EDUCATION == 3 | credit_card.train$EDUCATION == 2, 1, 0)
ed.school = ifelse(credit_card.train$EDUCATION == 1 | credit_card.train$EDUCATION == 2 | credit_card.tra

credit_card.train = cbind(credit_card.train,ed.other, ed.high_school, ed.school, ed.university)

# Borramos la columna EDUCATION.
credit_card.train = credit_card.train[,-which(colnames(credit_card.train) == "EDUCATION")]
summary(credit_card.train)
```

```
##    LIMIT_BAL           SEX            MARRIAGE          AGE
##  Min.   :  10000   Min.   :0.0000   Min.   :0.000   Min.   :21.0
##  1st Qu.:  50000   1st Qu.:0.0000   1st Qu.:1.000   1st Qu.:28.0
##  Median : 140000   Median :0.0000   Median :2.000   Median :34.0
##  Mean   : 167948   Mean   :0.3974   Mean   :1.549   Mean   :35.5
##  3rd Qu.: 240000   3rd Qu.:1.0000   3rd Qu.:2.000   3rd Qu.:41.0
##  Max.   :1000000   Max.   :1.0000   Max.   :3.000   Max.   :79.0
##      PAY_0              PAY_2              PAY_3              PAY_4
##  Min.   :-2.00000   Min.   :-2.0000   Min.   :-2.0000   Min.   :-2.000
##  1st Qu.:-1.00000   1st Qu.:-1.0000   1st Qu.:-1.0000   1st Qu.:-1.000
##  Median : 0.00000   Median : 0.0000   Median : 0.0000   Median : 0.000
##  Mean   :-0.01524   Mean   :-0.1321   Mean   :-0.1697   Mean   :-0.221
##  3rd Qu.: 0.00000   3rd Qu.: 0.0000   3rd Qu.: 0.0000   3rd Qu.: 0.000
##  Max.   : 8.00000   Max.   : 8.0000   Max.   : 8.0000   Max.   : 8.000
##      PAY_5              PAY_6           BILL_AMT1          BILL_AMT2
##  Min.   :-2.0000   Min.   :-2.000   Min.   :-154973   Min.   :-69777
##  1st Qu.:-1.0000   1st Qu.:-1.000   1st Qu.:   3508   1st Qu.:  2956
##  Median : 0.0000   Median : 0.000   Median :  22143   Median : 20895
##  Mean   :-0.2678   Mean   :-0.289   Mean   :  51238   Mean   : 49092
##  3rd Qu.: 0.0000   3rd Qu.: 0.000   3rd Qu.:  66175   3rd Qu.: 62795
##  Max.   : 8.0000   Max.   : 8.000   Max.   : 964511   Max.   :983931
##     BILL_AMT3          BILL_AMT4          BILL_AMT5         BILL_AMT6
##  Min.   : -46127   Min.   :-170000   Min.   :-81334   Min.   :-339603
##  1st Qu.:   2552   1st Qu.:   2276   1st Qu.:  1730   1st Qu.:   1261
##  Median :  20012   Median :  19044   Median : 18093   Median :  17036
##  Mean   :  47104   Mean   :  43452   Mean   : 40395   Mean   :  38948
##  3rd Qu.:  59674   3rd Qu.:  53536   3rd Qu.: 49719   3rd Qu.:  48740
##  Max.   :1664089   Max.   : 891586   Max.   :927171   Max.   : 961664
##     PAY_AMT1          PAY_AMT2           PAY_AMT3          PAY_AMT4
##  Min.   :     0   Min.   :      0.0   Min.   :     0   Min.   :     0
##  1st Qu.:  1000   1st Qu.:    809.8   1st Qu.:   390   1st Qu.:   300
##  Median :  2100   Median :   2000.0   Median :  1794   Median :  1500
##  Mean   :  5608   Mean   :   5988.9   Mean   :  5218   Mean   :  4793
##  3rd Qu.:  5006   3rd Qu.:   5000.0   3rd Qu.:  4525   3rd Qu.:  4010
##  Max.   :493358   Max.   :1684259.0   Max.   :896040   Max.   :528897
##     PAY_AMT5           PAY_AMT6      default.payment.next.month
##  Min.   :     0.0   Min.   :     0   Min.   :0.0000
##  1st Qu.:   269.8   1st Qu.:   100   1st Qu.:0.0000
##  Median :  1500.0   Median :  1500   Median :0.0000
```

```
##  Mean   :  4838.0   Mean   :  5252   Mean   :0.2209
##  3rd Qu.:  4037.8   3rd Qu.:  4000   3rd Qu.:0.0000
##  Max.   :426529.0   Max.   :528666   Max.   :1.0000
##     ed.other        ed.high_school     ed.school      ed.university
##  Min.   :0.000000   Min.   :0.0000   Min.   :0.0000   Min.   :0.0000
##  1st Qu.:0.000000   1st Qu.:0.0000   1st Qu.:1.0000   1st Qu.:0.0000
##  Median :0.000000   Median :1.0000   Median :1.0000   Median :0.0000
##  Mean   :0.003857   Mean   :0.6332   Mean   :0.9844   Mean   :0.4708
##  3rd Qu.:0.000000   3rd Qu.:1.0000   3rd Qu.:1.0000   3rd Qu.:1.0000
##  Max.   :1.000000   Max.   :1.0000   Max.   :1.0000   Max.   :1.0000
```

```
# TambiÃ©n tenemos que modificar la columna mariage. Introduciremos tres nueva columnas: marriage.marri
marriage.married = ifelse(credit_card.train$MARRIAGE == 1, 1,0)
marriage.single = ifelse(credit_card.train$MARRIAGE == 2, 1,0)
marriage.others = ifelse(credit_card.train$MARRIAGE == 3, 1,0)

# Introducimos los datos.
credit_card.train = cbind(credit_card.train, marriage.married, marriage.single, marriage.others)

# Borramos la variable MARRIAGE.
credit_card.train = credit_card.train[, -which(colnames(credit_card.train) == "MARRIAGE")]
summary(credit_card.train)
```

```
##    LIMIT_BAL          SEX             AGE           PAY_0
##  Min.   :  10000   Min.   :0.0000   Min.   :21.0   Min.   :-2.00000
##  1st Qu.:  50000   1st Qu.:0.0000   1st Qu.:28.0   1st Qu.:-1.00000
##  Median : 140000   Median :0.0000   Median :34.0   Median : 0.00000
##  Mean   : 167948   Mean   :0.3974   Mean   :35.5   Mean   :-0.01524
##  3rd Qu.: 240000   3rd Qu.:1.0000   3rd Qu.:41.0   3rd Qu.: 0.00000
##  Max.   :1000000   Max.   :1.0000   Max.   :79.0   Max.   : 8.00000
##      PAY_2             PAY_3             PAY_4             PAY_5
##  Min.   :-2.0000   Min.   :-2.0000   Min.   :-2.000   Min.   :-2.0000
##  1st Qu.:-1.0000   1st Qu.:-1.0000   1st Qu.:-1.000   1st Qu.:-1.0000
##  Median : 0.0000   Median : 0.0000   Median : 0.000   Median : 0.0000
##  Mean   :-0.1321   Mean   :-0.1697   Mean   :-0.221   Mean   :-0.2678
##  3rd Qu.: 0.0000   3rd Qu.: 0.0000   3rd Qu.: 0.000   3rd Qu.: 0.0000
##  Max.   : 8.0000   Max.   : 8.0000   Max.   : 8.000   Max.   : 8.0000
##      PAY_6          BILL_AMT1         BILL_AMT2        BILL_AMT3
##  Min.   :-2.000   Min.   :-154973   Min.   :-69777   Min.   : -46127
##  1st Qu.:-1.000   1st Qu.:   3508   1st Qu.:  2956   1st Qu.:   2552
##  Median : 0.000   Median :  22143   Median : 20895   Median :  20012
##  Mean   :-0.289   Mean   :  51238   Mean   : 49092   Mean   :  47104
##  3rd Qu.: 0.000   3rd Qu.:  66175   3rd Qu.: 62795   3rd Qu.:  59674
##  Max.   : 8.000   Max.   : 964511   Max.   :983931   Max.   :1664089
##    BILL_AMT4         BILL_AMT5        BILL_AMT6          PAY_AMT1
##  Min.   :-170000   Min.   :-81334   Min.   :-339603   Min.   :     0
##  1st Qu.:   2276   1st Qu.:  1730   1st Qu.:   1261   1st Qu.:  1000
##  Median :  19044   Median : 18093   Median :  17036   Median :  2100
##  Mean   :  43452   Mean   : 40395   Mean   :  38948   Mean   :  5608
##  3rd Qu.:  53536   3rd Qu.: 49719   3rd Qu.:  48740   3rd Qu.:  5006
##  Max.   : 891586   Max.   :927171   Max.   : 961664   Max.   :493358
##     PAY_AMT2           PAY_AMT3          PAY_AMT4          PAY_AMT5
##  Min.   :      0.0   Min.   :     0   Min.   :     0   Min.   :     0.0
##  1st Qu.:    809.8   1st Qu.:   390   1st Qu.:   300   1st Qu.:   269.8
##  Median :   2000.0   Median :  1794   Median :  1500   Median :  1500.0
```

```
##  Mean   :   5988.9   Mean   :  5218   Mean   :  4793   Mean   :  4838.0
##  3rd Qu.:   5000.0   3rd Qu.:  4525   3rd Qu.:  4010   3rd Qu.:  4037.8
##  Max.   :1684259.0   Max.   :896040   Max.   :528897   Max.   :426529.0
##     PAY_AMT6      default.payment.next.month    ed.other
##  Min.   :     0   Min.   :0.0000             Min.   :0.000000
##  1st Qu.:   100   1st Qu.:0.0000             1st Qu.:0.000000
##  Median :  1500   Median :0.0000             Median :0.000000
##  Mean   :  5252   Mean   :0.2209             Mean   :0.003857
##  3rd Qu.:  4000   3rd Qu.:0.0000             3rd Qu.:0.000000
##  Max.   :528666   Max.   :1.0000             Max.   :1.000000
##  ed.high_school     ed.school      ed.university    marriage.married
##  Min.   :0.0000   Min.   :0.0000   Min.   :0.0000   Min.   :0.000
##  1st Qu.:0.0000   1st Qu.:1.0000   1st Qu.:0.0000   1st Qu.:0.000
##  Median :1.0000   Median :1.0000   Median :0.0000   Median :0.000
##  Mean   :0.6332   Mean   :0.9844   Mean   :0.4708   Mean   :0.458
##  3rd Qu.:1.0000   3rd Qu.:1.0000   3rd Qu.:1.0000   3rd Qu.:1.000
##  Max.   :1.0000   Max.   :1.0000   Max.   :1.0000   Max.   :1.000
##  marriage.single  marriage.others
##  Min.   :0.0000   Min.   :0.00000
##  1st Qu.:0.0000   1st Qu.:0.00000
##  Median :1.0000   Median :0.00000
##  Mean   :0.5296   Mean   :0.01076
##  3rd Qu.:1.0000   3rd Qu.:0.00000
##  Max.   :1.0000   Max.   :1.00000
```

TambiÃ©n vamos a cambiar el nombre de la columna *PAY0* por *PAY1*.

```
colnames(credit_card.train)[which(colnames(credit_card.train)=="PAY_0")]="PAY_1"
summary(credit_card.train)
```

```
##    LIMIT_BAL           SEX             AGE           PAY_1
##  Min.   :  10000   Min.   :0.0000   Min.   :21.0   Min.   :-2.00000
##  1st Qu.:  50000   1st Qu.:0.0000   1st Qu.:28.0   1st Qu.:-1.00000
##  Median : 140000   Median :0.0000   Median :34.0   Median : 0.00000
##  Mean   : 167948   Mean   :0.3974   Mean   :35.5   Mean   :-0.01524
##  3rd Qu.: 240000   3rd Qu.:1.0000   3rd Qu.:41.0   3rd Qu.: 0.00000
##  Max.   :1000000   Max.   :1.0000   Max.   :79.0   Max.   : 8.00000
##      PAY_2             PAY_3             PAY_4            PAY_5
##  Min.   :-2.0000   Min.   :-2.0000   Min.   :-2.000   Min.   :-2.0000
##  1st Qu.:-1.0000   1st Qu.:-1.0000   1st Qu.:-1.000   1st Qu.:-1.0000
##  Median : 0.0000   Median : 0.0000   Median : 0.000   Median : 0.0000
##  Mean   :-0.1321   Mean   :-0.1697   Mean   :-0.221   Mean   :-0.2678
##  3rd Qu.: 0.0000   3rd Qu.: 0.0000   3rd Qu.: 0.000   3rd Qu.: 0.0000
##  Max.   : 8.0000   Max.   : 8.0000   Max.   : 8.000   Max.   : 8.0000
##      PAY_6          BILL_AMT1         BILL_AMT2        BILL_AMT3
##  Min.   :-2.000   Min.   :-154973   Min.   :-69777   Min.   : -46127
##  1st Qu.:-1.000   1st Qu.:   3508   1st Qu.:  2956   1st Qu.:   2552
##  Median : 0.000   Median :  22143   Median : 20895   Median :  20012
##  Mean   :-0.289   Mean   :  51238   Mean   : 49092   Mean   :  47104
##  3rd Qu.: 0.000   3rd Qu.:  66175   3rd Qu.: 62795   3rd Qu.:  59674
##  Max.   : 8.000   Max.   : 964511   Max.   :983931   Max.   :1664089
##    BILL_AMT4         BILL_AMT5         BILL_AMT6          PAY_AMT1
##  Min.   :-170000   Min.   :-81334   Min.   :-339603   Min.   :     0
##  1st Qu.:   2276   1st Qu.:  1730   1st Qu.:   1261   1st Qu.:  1000
##  Median :  19044   Median : 18093   Median :  17036   Median :  2100
```

```
##  Mean   :  43452   Mean   : 40395   Mean   :  38948   Mean   :  5608
##  3rd Qu.:  53536   3rd Qu.: 49719   3rd Qu.:  48740   3rd Qu.:  5006
##  Max.   : 891586   Max.   :927171   Max.   : 961664   Max.   :493358
##     PAY_AMT2           PAY_AMT3          PAY_AMT4          PAY_AMT5
##  Min.   :      0.0   Min.   :     0   Min.   :     0   Min.   :     0.0
##  1st Qu.:    809.8   1st Qu.:   390   1st Qu.:   300   1st Qu.:   269.8
##  Median :   2000.0   Median :  1794   Median :  1500   Median :  1500.0
##  Mean   :   5988.9   Mean   :  5218   Mean   :  4793   Mean   :  4838.0
##  3rd Qu.:   5000.0   3rd Qu.:  4525   3rd Qu.:  4010   3rd Qu.:  4037.8
##  Max.   :1684259.0   Max.   :896040   Max.   :528897   Max.   :426529.0
##     PAY_AMT6      default.payment.next.month    ed.other
##  Min.   :     0   Min.   :0.0000             Min.   :0.000000
##  1st Qu.:   100   1st Qu.:0.0000             1st Qu.:0.000000
##  Median :  1500   Median :0.0000             Median :0.000000
##  Mean   :  5252   Mean   :0.2209             Mean   :0.003857
##  3rd Qu.:  4000   3rd Qu.:0.0000             3rd Qu.:0.000000
##  Max.   :528666   Max.   :1.0000             Max.   :1.000000
##  ed.high_school     ed.school      ed.university    marriage.married
##  Min.   :0.0000   Min.   :0.0000   Min.   :0.0000   Min.   :0.000
##  1st Qu.:0.0000   1st Qu.:1.0000   1st Qu.:0.0000   1st Qu.:0.000
##  Median :1.0000   Median :1.0000   Median :0.0000   Median :0.000
##  Mean   :0.6332   Mean   :0.9844   Mean   :0.4708   Mean   :0.458
##  3rd Qu.:1.0000   3rd Qu.:1.0000   3rd Qu.:1.0000   3rd Qu.:1.000
##  Max.   :1.0000   Max.   :1.0000   Max.   :1.0000   Max.   :1.000
##  marriage.single  marriage.others
##  Min.   :0.0000   Min.   :0.00000
##  1st Qu.:0.0000   1st Qu.:0.00000
##  Median :1.0000   Median :0.00000
##  Mean   :0.5296   Mean   :0.01076
##  3rd Qu.:1.0000   3rd Qu.:0.00000
##  Max.   :1.0000   Max.   :1.00000
```

Por Ãºltimo.

```
# Por Ãºltimo, utilizamos la funciÃ³n preprocess.
trans = preProcess(credit_card.train, c("BoxCox") )
trainTransformado = predict(trans, credit_card.train)
summary(trainTransformado)
```

```
##    LIMIT_BAL           SEX              AGE            PAY_1
##  Min.   : 49.50   Min.   :0.0000   Min.   :1.564   Min.   :-2.00000
##  1st Qu.: 82.29   1st Qu.:0.0000   1st Qu.:1.622   1st Qu.:-1.00000
##  Median :113.27   Median :0.0000   Median :1.657   Median : 0.00000
##  Mean   :111.23   Mean   :0.3974   Mean   :1.656   Mean   :-0.01524
##  3rd Qu.:133.74   3rd Qu.:1.0000   3rd Qu.:1.688   3rd Qu.: 0.00000
##  Max.   :206.99   Max.   :1.0000   Max.   :1.775   Max.   : 8.00000
##      PAY_2             PAY_3             PAY_4             PAY_5
##  Min.   :-2.0000   Min.   :-2.0000   Min.   :-2.000   Min.   :-2.0000
##  1st Qu.:-1.0000   1st Qu.:-1.0000   1st Qu.:-1.000   1st Qu.:-1.0000
##  Median : 0.0000   Median : 0.0000   Median : 0.000   Median : 0.0000
##  Mean   :-0.1321   Mean   :-0.1697   Mean   :-0.221   Mean   :-0.2678
##  3rd Qu.: 0.0000   3rd Qu.: 0.0000   3rd Qu.: 0.000   3rd Qu.: 0.0000
##  Max.   : 8.0000   Max.   : 8.0000   Max.   : 8.000   Max.   : 8.0000
##      PAY_6          BILL_AMT1         BILL_AMT2        BILL_AMT3
##  Min.   :-2.000   Min.   :-154973   Min.   :-69777   Min.   : -46127
```

```
##  1st Qu.:-1.000   1st Qu.:   3508   1st Qu.:  2956   1st Qu.:   2552
##  Median : 0.000   Median :  22143   Median : 20895   Median :  20012
##  Mean   :-0.289   Mean   :  51238   Mean   : 49092   Mean   :  47104
##  3rd Qu.: 0.000   3rd Qu.:  66175   3rd Qu.: 62795   3rd Qu.:  59674
##  Max.   : 8.000   Max.   : 964511   Max.   :983931   Max.   :1664089
##    BILL_AMT4         BILL_AMT5         BILL_AMT6          PAY_AMT1
##  Min.   :-170000   Min.   :-81334   Min.   :-339603   Min.   :     0
##  1st Qu.:   2276   1st Qu.:  1730   1st Qu.:   1261   1st Qu.:  1000
##  Median :  19044   Median : 18093   Median :  17036   Median :  2100
##  Mean   :  43452   Mean   : 40395   Mean   :  38948   Mean   :  5608
##  3rd Qu.:  53536   3rd Qu.: 49719   3rd Qu.:  48740   3rd Qu.:  5006
##  Max.   : 891586   Max.   :927171   Max.   : 961664   Max.   :493358
##     PAY_AMT2           PAY_AMT3          PAY_AMT4          PAY_AMT5
##  Min.   :      0.0   Min.   :     0   Min.   :     0   Min.   :     0.0
##  1st Qu.:    809.8   1st Qu.:   390   1st Qu.:   300   1st Qu.:   269.8
##  Median :   2000.0   Median :  1794   Median :  1500   Median :  1500.0
##  Mean   :   5988.9   Mean   :  5218   Mean   :  4793   Mean   :  4838.0
##  3rd Qu.:   5000.0   3rd Qu.:  4525   3rd Qu.:  4010   3rd Qu.:  4037.8
##  Max.   :1684259.0   Max.   :896040   Max.   :528897   Max.   :426529.0
##     PAY_AMT6      default.payment.next.month    ed.other
##  Min.   :     0   Min.   :0.0000             Min.   :0.000000
##  1st Qu.:   100   1st Qu.:0.0000             1st Qu.:0.000000
##  Median :  1500   Median :0.0000             Median :0.000000
##  Mean   :  5252   Mean   :0.2209             Mean   :0.003857
##  3rd Qu.:  4000   3rd Qu.:0.0000             3rd Qu.:0.000000
##  Max.   :528666   Max.   :1.0000             Max.   :1.000000
##  ed.high_school     ed.school      ed.university    marriage.married
##  Min.   :0.0000   Min.   :0.0000   Min.   :0.0000   Min.   :0.000
##  1st Qu.:0.0000   1st Qu.:1.0000   1st Qu.:0.0000   1st Qu.:0.000
##  Median :1.0000   Median :1.0000   Median :0.0000   Median :0.000
##  Mean   :0.6332   Mean   :0.9844   Mean   :0.4708   Mean   :0.458
##  3rd Qu.:1.0000   3rd Qu.:1.0000   3rd Qu.:1.0000   3rd Qu.:1.000
##  Max.   :1.0000   Max.   :1.0000   Max.   :1.0000   Max.   :1.000
##  marriage.single  marriage.others
##  Min.   :0.0000   Min.   :0.00000
##  1st Qu.:0.0000   1st Qu.:0.00000
##  Median :1.0000   Median :0.00000
##  Mean   :0.5296   Mean   :0.01076
##  3rd Qu.:1.0000   3rd Qu.:0.00000
##  Max.   :1.0000   Max.   :1.00000
```

```r
# Para hacer mÃ¡s sencillo hacer las transformaciones al conjunto de test, se crearÃ¡n funciones para r

# FunciÃ³n para comprobar si hay datos pÃ©rdidos y reemplazarlos.
reemplazarCol = function(x){
  # Columna SEX
  x$SEX = ifelse(x$SEX == 2, 0, 1)

  # Columna EDUCATION.
  x.other = ifelse(x$EDUCATION == 4, 1,0)
  x.university = ifelse(x$EDUCATION == 2, 1, 0)
  x.high_school = ifelse(x$EDUCATION == 3 | x$EDUCATION == 2, 1, 0)
  x.school = ifelse(x$EDUCATION == 1 | x$EDUCATION == 2 | x$EDUCATION == 3, 1, 0)
  x = cbind(x,x.other, x.high_school, x.school, x.university)
```

```
  # Borramos la columna EDUCATION.
  x = x[,-which(colnames(x) == "EDUCATION")]

  # Columna MARRIAGE.
  x.married = ifelse(x$MARRIAGE == 1, 1,0)
  x.single = ifelse(x$MARRIAGE == 2, 1,0)
  x.others = ifelse(x$MARRIAGE == 3, 1,0)

  # Introducimos los datos.
  x = cbind(x.train, x.married, x.single, x.others)

  # Borramos la variable MARRIAGE.
  x = x[, -which(colnames(x) == "MARRIAGE")]

  # Cambiamos el nombre de la variables PAY_0.
  colnames(x)[which(colnames(x)=="PAY_0")]="PAY_1"
  x
}
preprocesar = function(x,pred=trans){
  transTest = predict(pred, x)
  transTest
}
```