SVM

Cristopher Barrios, Carlos Daniel Estrada

2023-04-21

```
library(e1071)
library(caret)
library(corrplot)
library(labelled)
library(plotly)
library(ggplot2)
```

1. Use los mismos conjuntos de entrenamiento y prueba de las hojas de trabajo pasadas para probar el algoritmo.

```
set.seed(123)
train = read.csv("./train.csv")
```

2. Explore los datos y explique las transformaciones que debe hacerle para generar un modelo de máquinas vectoriales de soporte.

```
train[is.na(train)] <- 0
train$tipoDeCasa = as.numeric(as.character( cut(train$SalePrice,c(0,145000,205000,410000),
train[sapply(train, is.character)] <- lapply(train[sapply(train, is.character)],as.factor)

#columnas con NA
completeFun <- function(data, desiredCols) {
   completeVec <- complete.cases(data[, desiredCols])
   return(data[completeVec, ])
}
train <- completeFun(train, "tipoDeCasa") #variable respuesta, variable categorica</pre>
```

```
1436 obs. of 82 variables:
## 'data.frame':
## $ Id
                  : int 1 2 3 4 5 6 7 8 9 10 ...
## $ MSSubClass : int 60 20 60 70 60 50 20 60 50 190 ...
## $ MSZoning : Factor w/ 5 levels "C (all)","FV",..: 4 4 4 4 4 4 4 5 4 ...
## $ LotFrontage : num 65 80 68 60 84 85 75 0 51 50 ...
                 : int 8450 9600 11250 9550 14260 14115 10084 10382 6120 7420 ...
## $ LotArea
## $ Street
                  : Factor w/ 2 levels "Grvl", "Pave": 2 2 2 2 2 2 2 2 2 ...
                  : Factor w/ 3 levels "0", "Grvl", "Pave": 1 1 1 1 1 1 1 1 1 1 ...
## $ Alley
## $ LotShape : Factor w/ 4 levels "IR1", "IR2", "IR3", ...: 4 4 1 1 1 1 4 1 4 4 ...
## $ LandContour : Factor w/ 4 levels "Bnk", "HLS", "Low", ...: 4 4 4 4 4 4 4 4 4 ...
## $ Utilities : Factor w/ 2 levels "AllPub", "NoSeWa": 1 1 1 1 1 1 1 1 1 1 ...
                  : Factor w/ 5 levels "Corner", "CulDSac", ...: 5 3 5 1 3 5 5 1 5 1 ...
## $ LotConfig
```

str(train)

```
## $ LandSlope
                   : Factor w/ 3 levels "Gtl", "Mod", "Sev": 1 1 1 1 1 1 1 1 1 1 ...
## $ Neighborhood : Factor w/ 25 levels "Blmngtn", "Blueste",..: 6 25 6 7 14 12 21 17 18 4 ...
                  : Factor w/ 9 levels "Artery", "Feedr", ...: 3 2 3 3 3 3 3 5 1 1 ...
                   : Factor w/ 8 levels "Artery", "Feedr", ...: 3 3 3 3 3 3 3 3 1 ...
## $ Condition2
   $ BldgType
                   : Factor w/ 5 levels "1Fam", "2fmCon", ...: 1 1 1 1 1 1 1 1 2 ...
## $ HouseStyle
                   : Factor w/ 8 levels "1.5Fin", "1.5Unf", ...: 6 3 6 6 6 1 3 6 1 2 ...
## $ OverallQual : int 7 6 7 7 8 5 8 7 7 5 ...
   $ OverallCond : int 5 8 5 5 5 5 6 5 6 ...
##
##
   $ YearBuilt
                   : int
                         2003 1976 2001 1915 2000 1993 2004 1973 1931 1939 ...
## $ YearRemodAdd : int 2003 1976 2002 1970 2000 1995 2005 1973 1950 1950 ...
## $ RoofStyle
                  : Factor w/ 6 levels "Flat", "Gable",...: 2 2 2 2 2 2 2 2 2 ...
                   : Factor w/ 8 levels "ClyTile", "CompShg", ...: 2 2 2 2 2 2 2 2 2 2 ...
## $ RoofMatl
   $ Exterior1st : Factor w/ 15 levels "AsbShng", "AsphShn",..: 13 9 13 14 13 13 13 7 4 9 ...
##
## $ Exterior2nd : Factor w/ 16 levels "AsbShng", "AsphShn", ...: 14 9 14 16 14 14 14 7 16 9 ...
## $ MasVnrType
                  : Factor w/ 5 levels "0", "BrkCmn", "BrkFace", ...: 3 4 3 4 3 4 5 5 4 4 ....
##
   $ MasVnrArea
                  : num 196 0 162 0 350 0 186 240 0 0 ...
## $ ExterQual
                  : Factor w/ 4 levels "Ex", "Fa", "Gd", ...: 3 4 3 4 3 4 3 4 4 4 ...
## $ ExterCond
                   : Factor w/ 5 levels "Ex", "Fa", "Gd", ...: 5 5 5 5 5 5 5 5 5 5 ...
                   : Factor w/ 6 levels "BrkTil", "CBlock", ...: 3 2 3 1 3 6 3 2 1 1 ...
## $ Foundation
## $ BsmtQual
                   : Factor w/ 5 levels "0", "Ex", "Fa", ...: 4 4 4 5 4 4 2 4 5 5 ...
## $ BsmtCond
                   : Factor w/ 5 levels "0", "Fa", "Gd", ...: 5 5 5 3 5 5 5 5 5 5 ...
## $ BsmtExposure : Factor w/ 5 levels "0", "Av", "Gd", ...: 5 3 4 5 2 5 2 4 5 5 ...
## $ BsmtFinType1 : Factor w/ 7 levels "0", "ALQ", "BLQ", ...: 4 2 4 2 4 4 4 2 7 4 ...
                  : int 706 978 486 216 655 732 1369 859 0 851 ...
   $ BsmtFinSF1
## $ BsmtFinType2 : Factor w/ 7 levels "0", "ALQ", "BLQ", ...: 7 7 7 7 7 7 7 3 7 7 ...
## $ BsmtFinSF2
                  : int 0000003200...
## $ BsmtUnfSF
                   : int
                         150 284 434 540 490 64 317 216 952 140 ...
   $ TotalBsmtSF : int 856 1262 920 756 1145 796 1686 1107 952 991 ...
## $ Heating
                   : Factor w/ 6 levels "Floor", "GasA", ...: 2 2 2 2 2 2 2 2 2 2 ...
                   : Factor w/ 5 levels "Ex", "Fa", "Gd", ...: 1 1 1 3 1 1 1 1 3 1 ....
## $ HeatingQC
                   : Factor w/ 2 levels "N", "Y": 2 2 2 2 2 2 2 2 2 2 ...
##
   $ CentralAir
##
   $ Electrical
                  : Factor w/ 6 levels "0", "FuseA", "FuseF",..: 6 6 6 6 6 6 6 6 3 6 ...
## $ X1stFlrSF
                   : int 856 1262 920 961 1145 796 1694 1107 1022 1077 ...
                   : int 854 0 866 756 1053 566 0 983 752 0 ...
## $ X2ndFlrSF
##
   $ LowQualFinSF : int
                         0 0 0 0 0 0 0 0 0 0 ...
## $ GrLivArea
                  : int 1710 1262 1786 1717 2198 1362 1694 2090 1774 1077 ...
## $ BsmtFullBath : int 1 0 1 1 1 1 1 1 0 1 ...
## $ BsmtHalfBath : int 0 1 0 0 0 0 0 0 0 ...
##
   $ FullBath
                  : int 2 2 2 1 2 1 2 2 2 1 ...
## $ HalfBath
                  : int 1010110100...
## $ BedroomAbvGr : int 3 3 3 3 4 1 3 3 2 2 ...
## $ KitchenAbvGr : int 1 1 1 1 1 1 1 2 2 ...
   $ KitchenQual : Factor w/ 4 levels "Ex", "Fa", "Gd", ...: 3 4 3 3 3 4 3 4 4 4 ...
## $ TotRmsAbvGrd : int 8 6 6 7 9 5 7 7 8 5 ...
                 : Factor w/ 7 levels "Maj1", "Maj2", ...: 7 7 7 7 7 7 7 3 7 ...
## $ Functional
##
   $ Fireplaces
                   : int 0 1 1 1 1 0 1 2 2 2 ...
   $ FireplaceQu : Factor w/ 6 levels "0", "Ex", "Fa", ...: 1 6 6 4 6 1 4 6 6 6 ...
##
## $ GarageType
                  : Factor w/ 7 levels "0", "2Types", "Attchd", ...: 3 3 3 7 3 3 3 7 3 ...
## $ GarageYrBlt : num 2003 1976 2001 1998 2000 ...
## $ GarageFinish : Factor w/ 4 levels "0", "Fin", "RFn", ...: 3 3 3 4 3 4 3 3 4 3 ...
                  : int 2223322221 ...
## $ GarageCars
## $ GarageArea
                  : int 548 460 608 642 836 480 636 484 468 205 ...
## $ GarageQual
                 : Factor w/ 6 levels "0", "Ex", "Fa", ...: 6 6 6 6 6 6 6 6 3 4 ...
## $ GarageCond
                  : Factor w/ 6 levels "0", "Ex", "Fa", ...: 6 6 6 6 6 6 6 6 6 ...
```

```
## $ PavedDrive : Factor w/ 3 levels "N", "P", "Y": 3 3 3 3 3 3 3 3 3 3 ...
## $ WoodDeckSF : int 0 298 0 0 192 40 255 235 90 0 ...
## $ OpenPorchSF : int 61 0 42 35 84 30 57 204 0 4 ...
## $ EnclosedPorch: int 0 0 0 272 0 0 0 228 205 0 ...
## $ X3SsnPorch : int
                        0 0 0 0 0 320 0 0 0 0 ...
## $ ScreenPorch : int 0 0 0 0 0 0 0 0 0 ...
## $ PoolArea : int 0 0 0 0 0 0 0 0 0 ...
                  : Factor w/ 4 levels "0", "Ex", "Fa", ...: 1 1 1 1 1 1 1 1 1 1 ...
## $ PoolQC
## $ Fence
                  : Factor w/ 5 levels "0", "GdPrv", "GdWo", ...: 1 1 1 1 1 4 1 1 1 1 ...
## $ MiscFeature : Factor w/ 5 levels "0", "Gar2", "Othr", ..: 1 1 1 1 1 4 1 4 1 1 ...
## $ MiscVal
                 : int 0 0 0 0 0 700 0 350 0 0 ...
                  : int 2 5 9 2 12 10 8 11 4 1 ...
## $ MoSold
                  : int 2008 2007 2008 2006 2008 2009 2007 2009 2008 2008 ...
## $ YrSold
                : Factor w/ 9 levels "COD", "Con", "ConLD", ...: 9 9 9 9 9 9 9 9 9 ...
## $ SaleType
## $ SaleCondition: Factor w/ 6 levels "Abnorml", "AdjLand",..: 5 5 5 1 5 5 5 5 1 5 ...
                : int 208500 181500 223500 140000 250000 143000 307000 200000 129900 118000 ...
## $ SalePrice
## $ tipoDeCasa
                 : num 3 2 3 1 3 1 3 2 1 1 ...
```

3. Use como variable respuesta la variable categórica que especifica si la casa es barata, media o cara

```
train <- completeFun(train, "tipoDeCasa") #variable respuesta, variable categorica

#datos con factor
frstselect <- train[,c(2:5,8,9,11:43,46,49:54,56:62,64:72,76:80,82)]

#datos cuantitativos
scndselect <- subset (train, select = c(2,4,5,18,19,20,21,27,35,37,38,39,44,45,46,47,48,49,50,51,52,53,scndselect[is.na(scndselect)] <- 0</pre>
```

4. Genere varios (más de 2) modelos de SVM con diferentes kernels y distintos valores en los parámetros c, gamma (circular) y d (en caso de que utilice el polinomial). Puede tunear el modelo de forma automática siempre que explique los resultados.

```
LotFrontage
                              0.16 0.23 0.24 0.09 0.15 0.03 0.07 0.09 0.06 0.01 0.05-0.22 0.02 0.2 -0.01-0.05 0.03
      BsmtUnfSF 0.16 0.41 0.31 0.1 0.3 0.15 0.18 -0.0 1-0.51 0.43 0.02 -0.14 0 0.23 -0.24 0.14 -0.1
                                                                                                     8.0
     TotalBsmtSF 0.23 0.41 0.8 0.33 0.5 0.38 0.27 0.25 0.5 0.3 -0.03 0.24 0.21 0.41 0.11 -0.16 0
         X1stFIrSF 0.24 0.31 0.8
                                          0.31 0.44 0.26 0.22 0.29 0.42 0.24-0.0<mark>2-0.25-0.25</mark>0.53 0.11-0.14 0
                                                                                                    0.6
     MasVnrArea 0.09 0.1 0.33 0.31 0.37 0.3 0.16 0.08 0.24 0.08 <u>-0.070.03 0.13 0.34 -0.070.12</u>0.02
      OverallQual 0.15 0.3 0.5 0.44 0.37 0.57 0.54 0.08 0.2 0.1 -0.050.04 0.27 0.56-0.060.09 0.04
                                                                                                     0.4
          YearBuilt 0.030.150.380.26 0.3 0.57 0.590.010.230.18-0.170.03 0 0.18-0.050.370.04
                                                                                                    0.2
YearRemodAdd 0.07 0.18 0.27 0.22 0.16 0.54 0.59
                                                           0 0.11 0.11-0.070.05 0.13 0.27-0.070.08-0.01
            LotArea 0.09-0.010.25 0.29 0.08 0.08 0.01 0
                                                             0.2 0.15 0 -0.140.03 0.25 0.12 0 0.05
     BsmtFinSF1 0.06-0.51 0.5 0.42 0.24 0.2 0.23 0.11 0.2
                                                                 0.65-0.06-0.06-0.170.16-0.05-0.030.07
   BsmtFullBath 0.01-0.43 0.3 0.24 0.08 0.1 0.180.11 0.15 0.65 -0.040.01-0.180.02 0.16-0.050.15
                                                                                                     -0.2
 LowQualFinSF 0.05 0.02-0.030.020.070.050.170.07 0 -0.060.04
                                                                        0.04 0.05 0.12 0.02 0 0
    MSSubClass -0.220.140.240.250.030.040.030.05-0.140.060.010.04
                                                                            0.31 0.09-0.07-0.06 0
                                                                                                     -0.4
        X2ndFlrSF 0.02 0 -0.240.250.130.27 0 0.130.03-0.170.180.05 0.31 0.68-0.1 0.03-0.03
        GrLivArea 0.2 0.23 0.41 0.53 0.34 0.56 0.18 0.27 0.25 0.16 0.02 0.12 0.09 0.68
                                                                                                     -0.6
                                                                                     0 -0.08-0.03
     BsmtFinSF2 -0.01-0.210.110.11-0.070.060.050.070.12-0.050.160.02-0.07-0.1 0
                                                                                                     -0.8
     OverallCond -0.0$0.140.160.140.120.090.370.08 0 -0.030.05 0 -0.060.03-0.080.04
                                                                                            0.12
  BsmtHalfBath -0.03-0.1 0 0 0.02-0.040.040.010.050.07-0.15 0 0 -0.030.030.070.12
```

```
GarageYrBlt
                            WoodDeckSF 0.12 0.190.170.280.210.21-0.030.030.05 0.1 0.040.15-0.080.03-0.130.07-0.090.01
                                                                                             0.8
       Fireplaces 0.180.19 0.230.450.280.250.01-0.020.160.19 0.1 0.3 0.180.06-0.020.08-0.12 0
          FullBath 0.130.170.23
                                     0.570.450.390.04-0.010.260.120.360.53-0.020.06-0.120.030.14-0.01
                                                                                             0.6
     tipoDeCasa 0.260.280.450.57 0.610.580.06-0.010.330.310.160.44 0.1 0.07-0.140.05-0.17 0
    GarageCars 0.6 0.210.280.450.61 0.880.04-0.030.2 0.2 0.080.330.040.05-0.150.01-0.050.04
                                                                                             0.4
    GarageArea 0.570.210.250.390.580.88 0.04-0.020.220.150.07 0.3 0.040.04-0.120.05-0.060.03
    X3SsnPorch 0.03-0.03.010.040.060.040.04 0.02 0 0 -0.02 0 -0.03.03-0.040.040.03 0
                                                                                             0.2
             YrSold -0.010.03-0.020.040.040.030.020.02 -0.050.040.030.030.01-0.150.040.080.03 0
 OpenPorchSF 0.050.050.160.260.33 0.2 0.22 0 -0.05 0.19 0.1 0.220.060.08-0.090.06-0.070.02
                                                                                              0
          HalfBath 0.11 0.1 0.190.120.31 0.2 0.15 0 -0.010.19
                                                             0.220.330.07-0.040.090.01-0.07 0
BedroomAbvGr -0.010.04 0.1 0.360.160.080.07-0.020.030.1 0.22 0.690.040.050.040.06 0.2 0.01
                                                                                             -0.2
TotRmsAbvGrd 0.090.15 0.3 0.530.440.33 0.3 0 -0.030.220.330.69 0.030.050.010.070.280.03
                                                                                             -0.4
   ScreenPorch 0.06-0.080.18-0.020.1 0.040.04-0.030.010.060.070.040.03
           MoSold 0.020.030.060.060.070.050.040.03<mark>-0.15</mark>0.08-0.010.050.050.03
                                                                                             -0.6
EnclosedPorch -0.080.130.020.120.140.150.120.040.040.090.090.040.01-0.080.03
         PoolArea 0.010.070.080.030.050.010.05-0.040.060.010.060.070.06-0.040.06
                                                                                             -0.8
  KitchenAbvGr -0.160.090.120.14-0.170.050.060.030.03-0.070.070.2 0.28-0.050.030.04-0.01
           MiscVal -0.040.01 0 -0.01 0 -0.040.03 0 0 -0.02 0 0.010.030.03-0.010.020.030.06
```

```
3edroomAbvGr
                                                                        EnclosedPorcl
                                    HalfBath
ScreenPorch
                                                                    SarageCars
                                                                GarageYrB
             LotArea 0.110.08 0 0.030.11-0.020.080.170.040.070.14-0.020.270.170.02-0.010.16 0
 LowQualFinSF 0.1 0.07-0.040.03 0 0.01-0.010.12 0 -0.160.110.07-0.040.08 0 -0.020.020.02
                                                                                                         0.8
     LotFrontage 0.140.09-0.030.01 0.1 0.040.06 0.21-0.060.02 0.15 0.040.03 0.19 0.03-0.040.030.02
        X2ndFirSF 0.5 0.05 0.61 0.02 0.4 0.070.21 0.61 0.02 0.06 0.16 0.070.18 0.12-0.020.02 0.07 0.04
                                                                                                         0.6
    MSSubClass -0.030.01 0.18-0.030.14 0.28 0 0.05-0.040.080.030.040.040.090.040.020.01-0.02
      OverallQual 0.1 0.05 0.25 0.05 0.53-0.180.29 0.38-0.030.29 0.58-0.110.37 0.54 0.04-0.020.22 0.08
                                                                                                         0.4
      BsmtUnfSF 0.16-0.03-0.05-0.030.28 0.03 0.12 0.24-0.020.04 0.21-0.010.04 0.18 0.02-0.03 0.010.04
                                                                                                         0.2
     OverallCond 0.01 0 -0.060.04-0.190.090.030.050.07 0 -0.180.07-0.020.150.020.05 0 0
         GrLivArea 0.53 0.14 0.41 0.08 0.61 0.12 0.33 0.82 0 0.16 0.44 0.01 0.44 0.44 0.03-0.030.23 0.07
                                                                                                          0
YearRemodAdd -0.04 0 0.17-0.050.43-0.150.22 0.17-0.010.14 0.41-0.19 0.1 0.36 0.05 0.04 0.2 0.03
     MasVnrArea 0.1 0.02 0.18 0.07 0.23-0.030.13 0.23-0.030.13 0.34-0.11 0.22 0.35 0.03 0 0.14 0.01
                                                                                                         -0.2
    TotalBsmtSF 0.05 0.11-0.080.08 0.29-0.060.23 0.24-0.020.17 0.41-0.110.31 0.47 0.04-0.010.22 0.03
           YearBuilt -0.07 0 0.24-0.050.46-0.170.19 0.07-0.030.27 0.53-0.390.14 0.47 0.03-0.010.22 0.02
                                                                                                         -0.4
     BsmtFinSF1 -0.110.12-0.010.070.03-0.080.1 0 0.010.11 0.2 -0.1 0.240.27 0.03 0.02 0.19-0.01
         X1stFIrSF 0.13 0.12-0.15 0.08 0.35 0.08 0.19 0.37-0.020.16 0.41-0.070.39 0.47 0.06-0.010.22 0.05
                                                                                                         -0.6
   BsmtHalfBath 0.040.02-0.010.04-0.060.040.02-0.020.010.02-0.020.010.03-0.020.04-0.050.040.04
                                                                                                         -0.8
     BsmtFinSF2 -0.010.05-0.030.09-0.07-0.040.01-0.03 0 0.04-0.040.04 0.05-0.01-0.030.03 0.07-0.01
   BsmtFullBath -0.150.06-0.040.03-0.070.040.06-0.070.020.050.12-0.050.130.17 0 0.070.17-0.02
```

tipocor<- cor(scndselect[,-1],scndselect\$tipoDeCasa) tipocor</pre>

```
##
                          [,1]
                  0.124781325
## LotFrontage
## LotArea
                  0.219555016
## OverallQual
                  0.738892019
## OverallCond
                  -0.095483076
                  0.565329689
## YearBuilt
## YearRemodAdd
                  0.544774124
## MasVnrArea
                  0.344254632
## BsmtFinSF1
                  0.276950433
## BsmtFinSF2
                  0.012463641
## BsmtUnfSF
                  0.208466092
## TotalBsmtSF
                  0.513208477
## X1stFlrSF
                  0.496997346
## X2ndFlrSF
                  0.297500739
## LowQualFinSF
                 -0.068654379
## GrLivArea
                  0.622251690
## BsmtFullBath
                  0.189408729
## BsmtHalfBath
                 -0.035329884
## FullBath
                  0.574967912
## HalfBath
                  0.307309514
## BedroomAbvGr
                  0.157624456
## KitchenAbvGr -0.167361574
```

```
## TotRmsAbvGrd 0.441616356
## Fireplaces 0.454122445
## GarageYrBlt 0.262710019
## GarageCars 0.610605411
## GarageArea 0.575708954
## WoodDeckSF
                0.281222474
## OpenPorchSF 0.326152417
## EnclosedPorch -0.138763225
## X3SsnPorch 0.059689959
## ScreenPorch 0.099859557
## PoolArea
              0.051525222
## MiscVal
                -0.001375927
## MoSold
                 0.070906122
## YrSold
                -0.012373046
## tipoDeCasa
                1.000000000
porciento <- 70/100
#todo tipo de variables
trainRowsNumber<-sample(1:nrow(frstselect),porciento*nrow(frstselect))</pre>
train<-frstselect[trainRowsNumber,]</pre>
test<-frstselect[-trainRowsNumber,]</pre>
#variables cuantitativas
trainRowsNum<-sample(1:nrow(scndselect),porciento*nrow(scndselect))</pre>
train1<-scndselect[trainRowsNum,]</pre>
test1<-scndselect[-trainRowsNum,]</pre>
#Modelos
modeloSVM_L1<-svm(tipoDeCasa~., data=train,type="C-classification", cost=2^5, kernel="linear")</pre>
modeloSVM_L2<-svm(tipoDeCasa~., data=train,type="C-classification", cost=0.5, kernel="linear")
modeloSVM_L3<-svm(tipoDeCasa~., data=train,type="C-classification", cost=2^-5, kernel="linear")</pre>
modeloSVM_R1<-svm(tipoDeCasa~., data=train,type="C-classification", gamma=0.005,kernel="radial")
modeloSVM_R2<-svm(tipoDeCasa~., data=train,type="C-classification", gamma=0.05,kernel="radial")
modeloSVM_R3<-svm(tipoDeCasa~., data=train,type="C-classification", gamma=2^-5,kernel="radial")</pre>
modeloSVM_P1<-svm(tipoDeCasa~., data=train,type="C-classification", gamma=1, kernel="polynomial", coef0
modeloSVM_P2<-svm(tipoDeCasa~., data=train,type="C-classification", gamma=5, kernel="polynomial", coef0
modeloSVM_P3<-svm(tipoDeCasa~., data=train,type="C-classification", gamma=2^-5, kernel="polynomial", co
summary(modeloSVM_L1)
##
## Call:
## svm(formula = tipoDeCasa ~ ., data = train, type = "C-classification",
       cost = 2^5, kernel = "linear")
##
##
##
## Parameters:
     SVM-Type: C-classification
##
```

```
## SVM-Kernel: linear
##
         cost: 32
##
## Number of Support Vectors: 337
##
##
   (91 153 93)
##
##
## Number of Classes: 3
##
## Levels:
## 1 2 3
summary(modeloSVM_R1)
##
## Call:
## svm(formula = tipoDeCasa ~ ., data = train, type = "C-classification",
       gamma = 0.005, kernel = "radial")
##
##
## Parameters:
     SVM-Type: C-classification
##
## SVM-Kernel: radial
##
         cost: 1
##
## Number of Support Vectors: 641
## ( 164 305 172 )
##
##
## Number of Classes: 3
## Levels:
## 1 2 3
summary(modeloSVM_P1)
##
## Call:
## svm(formula = tipoDeCasa ~ ., data = train, type = "C-classification",
       gamma = 1, kernel = "polynomial", coef0 = 1, degree = 8)
##
##
##
## Parameters:
     SVM-Type: C-classification
##
##
  SVM-Kernel: polynomial
##
        cost: 1
##
       degree: 8
##
       coef.0: 1
##
## Number of Support Vectors: 826
##
```

```
## ( 175 325 326 )
##
##
## Number of Classes: 3
##
## Levels:
## 1 2 3
```

5. Use los modelos para predecir el valor de la variable respuesta

```
# Linear
process_timeL1 <- proc.time()</pre>
prediccionL1<-predict(modeloSVM L1, newdata=test[,1:67])</pre>
process_timeL1 <- proc.time() - process_timeL1</pre>
process timeL2 <- proc.time()</pre>
prediccionL2<-predict(modeloSVM_L2, newdata=test[,1:67])</pre>
process_timeL2 <- proc.time() - process_timeL2</pre>
process timeL3 <- proc.time()</pre>
prediccionL3<-predict(modeloSVM_L3,newdata=test[,1:67])</pre>
process_timeL3 <- proc.time() - process_timeL3</pre>
process_timeL_avarage <- (process_timeL1[3] + process_timeL2[3] + process_timeL3[3])/3</pre>
# Radial
process_timeR1 <- proc.time()</pre>
prediccionR1<-predict(modeloSVM_R1, newdata=test[,1:67])#[,1:37]</pre>
process_timeR1 <- proc.time() - process_timeR1</pre>
process_timeR2 <- proc.time()</pre>
prediccionR2<-predict(modeloSVM_R2, newdata=test[,1:67])#[,1:37]</pre>
process_timeR2 <- proc.time() - process_timeR2</pre>
process timeR3 <- proc.time()</pre>
prediccionR3<-predict(modeloSVM R3,newdata=test[,1:67])#[,1:37]</pre>
process_timeR3 <- proc.time() - process_timeR3</pre>
process_timeR_avarage <- (process_timeR1[3] + process_timeR2[3] + process_timeR3[3])/3</pre>
# Polinomial
process_timeP1 <- proc.time()</pre>
prediccionP1<-predict(modeloSVM P1, newdata=test[,1:67])</pre>
process_timeP1 <- proc.time() - process_timeP1</pre>
process_timeP2 <- proc.time()</pre>
prediccionP2<-predict(modeloSVM_P2, newdata=test[,1:67])</pre>
process_timeP2 <- proc.time() - process_timeP2</pre>
process_timeP3 <- proc.time()</pre>
prediccionP3<-predict(modeloSVM_P3,newdata=test[,1:67])</pre>
process_timeP3 <- proc.time() - process_timeP3</pre>
process_timeP_avarage <- (process_timeP1[3] + process_timeP2[3] + process_timeP3[3])/3</pre>
#Cambio de tipo de data a factors
test$tipoDeCasa<- as.factor(test$tipoDeCasa)</pre>
test1$tipoDeCasa<- as.factor(test$tipoDeCasa)</pre>
```

6. Haga las matrices de confusión respectivas.

```
cmL1<-confusionMatrix(test$tipoDeCasa,prediccionL1)</pre>
cmL2<-confusionMatrix(test$tipoDeCasa,prediccionL2)</pre>
cmL3<-confusionMatrix(test$tipoDeCasa,prediccionL3)</pre>
#linear
cmL1
## Confusion Matrix and Statistics
##
##
            Reference
## Prediction 1 2
           1 128 29
##
                       1
##
           2 24 104 22
##
           3
              1 33 89
##
## Overall Statistics
##
##
                  Accuracy: 0.7448
                    95% CI : (0.7009, 0.7853)
##
##
      No Information Rate: 0.3852
##
      P-Value [Acc > NIR] : <2e-16
##
##
                     Kappa: 0.6143
##
##
  Mcnemar's Test P-Value: 0.4451
## Statistics by Class:
##
##
                        Class: 1 Class: 2 Class: 3
## Sensitivity
                          0.8366 0.6265
                                          0.7946
## Specificity
                          0.8921
                                   0.8264
                                            0.8934
## Pos Pred Value
                          0.8101 0.6933
                                           0.7236
## Neg Pred Value
                          0.9084 0.7794
                                           0.9253
## Prevalence
                          0.3550 0.3852
                                           0.2599
## Detection Rate
                          0.2970
                                0.2413
                                           0.2065
## Detection Prevalence
                          0.3666
                                   0.3480
                                            0.2854
## Balanced Accuracy
                          0.8643
                                   0.7265
                                            0.8440
#linear
cmL2
## Confusion Matrix and Statistics
##
##
            Reference
## Prediction 1 2
                        3
           1 140 18
##
                        0
##
            2 22 106 22
```

##

##

3

Overall Statistics

1 24 98

Accuracy: 0.7981

```
95% CI: (0.7571, 0.835)
##
##
       No Information Rate: 0.3782
       P-Value [Acc > NIR] : <2e-16
##
##
##
                     Kappa: 0.6953
##
  Mcnemar's Test P-Value: 0.6853
##
## Statistics by Class:
##
##
                        Class: 1 Class: 2 Class: 3
## Sensitivity
                                  0.7162
                          0.8589
                                            0.8167
## Specificity
                                            0.9196
                          0.9328
                                   0.8445
## Pos Pred Value
                                   0.7067
                                            0.7967
                          0.8861
## Neg Pred Value
                          0.9158
                                   0.8505
                                            0.9286
## Prevalence
                          0.3782
                                   0.3434
                                            0.2784
## Detection Rate
                          0.3248
                                   0.2459
                                            0.2274
## Detection Prevalence
                          0.3666
                                   0.3480
                                            0.2854
## Balanced Accuracy
                          0.8959
                                   0.7804
                                            0.8681
#linear
cmL3
## Confusion Matrix and Statistics
##
             Reference
## Prediction 1 2
            1 147 11
##
##
            2 19 115 16
##
                0 23 100
##
## Overall Statistics
##
##
                  Accuracy : 0.8399
##
                    95% CI: (0.8018, 0.8733)
##
       No Information Rate: 0.3852
##
       P-Value [Acc > NIR] : < 2.2e-16
##
##
                     Kappa: 0.7581
##
   Mcnemar's Test P-Value : NA
##
## Statistics by Class:
##
##
                        Class: 1 Class: 2 Class: 3
## Sensitivity
                          0.8855
                                 0.7718
                                            0.8621
## Specificity
                          0.9585
                                   0.8759
                                            0.9270
## Pos Pred Value
                          0.9304
                                   0.7667
                                            0.8130
## Neg Pred Value
                          0.9304
                                   0.8790
                                            0.9481
## Prevalence
                                   0.3457
                          0.3852
                                            0.2691
## Detection Rate
                          0.3411
                                   0.2668
                                            0.2320
## Detection Prevalence
                          0.3666
                                   0.3480
                                            0.2854
```

0.8945

0.8238

0.9220

Balanced Accuracy

```
#radial
cmR1<-confusionMatrix(test$tipoDeCasa,prediccionR1)
cmR2<-confusionMatrix(test$tipoDeCasa,prediccionR2)
cmR3<-confusionMatrix(test$tipoDeCasa,prediccionR3)

#radial
cmR1
## Confusion Matrix and Statistics</pre>
```

```
##
##
            Reference
## Prediction 1 2
                   9
##
           1 149
##
           2 23 117 10
##
           3
              1 32 90
##
## Overall Statistics
##
##
                 Accuracy: 0.826
                   95% CI: (0.7868, 0.8606)
##
##
      No Information Rate: 0.4014
##
      P-Value [Acc > NIR] : < 2.2e-16
##
##
                    Kappa : 0.736
##
##
  Mcnemar's Test P-Value: 0.0003231
## Statistics by Class:
##
##
                       Class: 1 Class: 2 Class: 3
## Sensitivity
                         0.8613 0.7405
                                         0.9000
## Specificity
                         0.9651
                                 0.8791
                                          0.9003
## Pos Pred Value
                         0.9430 0.7800
                                          0.7317
## Neg Pred Value
                         0.9121 0.8541
                                          0.9675
## Prevalence
                         0.4014 0.3666
                                          0.2320
## Detection Rate
                         0.3457
                                 0.2715
                                         0.2088
## Detection Prevalence
                         0.3666
                                 0.3480
                                          0.2854
## Balanced Accuracy
                         0.9132
                                  0.8098
                                           0.9002
```

#radial cmR2

```
## Confusion Matrix and Statistics
##
##
            Reference
## Prediction 1 2
                       3
           1 143 15
##
                       0
##
           2 23 119
##
           3
              2 37 84
##
## Overall Statistics
##
##
                 Accuracy: 0.8028
```

```
95% CI: (0.762, 0.8393)
##
##
       No Information Rate: 0.3968
       P-Value [Acc > NIR] : < 2.2e-16
##
##
##
                     Kappa: 0.7003
##
  Mcnemar's Test P-Value: 5.455e-05
##
## Statistics by Class:
##
##
                        Class: 1 Class: 2 Class: 3
## Sensitivity
                                  0.6959
                                            0.9130
                          0.8512
## Specificity
                                            0.8850
                          0.9430
                                   0.8808
## Pos Pred Value
                          0.9051
                                   0.7933
                                            0.6829
## Neg Pred Value
                          0.9084
                                   0.8149
                                            0.9740
## Prevalence
                          0.3898
                                   0.3968
                                            0.2135
## Detection Rate
                                            0.1949
                          0.3318
                                   0.2761
## Detection Prevalence
                          0.3666
                                   0.3480
                                            0.2854
## Balanced Accuracy
                          0.8971
                                   0.7883
                                            0.8990
#radial
cmR3
## Confusion Matrix and Statistics
##
             Reference
## Prediction 1 2
                        3
            1 145 13
##
                        0
##
            2 19 122
##
               0
                  31
                      92
##
## Overall Statistics
##
##
                  Accuracy : 0.8329
##
                    95% CI: (0.7943, 0.8669)
##
       No Information Rate: 0.3852
##
       P-Value [Acc > NIR] : < 2.2e-16
##
##
                     Kappa: 0.7467
##
   Mcnemar's Test P-Value : NA
##
## Statistics by Class:
##
##
                        Class: 1 Class: 2 Class: 3
## Sensitivity
                                  0.7349
                                            0.9109
                          0.8841
## Specificity
                                            0.9061
                          0.9513
                                  0.8943
## Pos Pred Value
                                   0.8133
                                            0.7480
                          0.9177
## Neg Pred Value
                          0.9304
                                   0.8434
                                            0.9708
## Prevalence
                                   0.3852
                          0.3805
                                            0.2343
## Detection Rate
                          0.3364
                                   0.2831
                                            0.2135
## Detection Prevalence
                          0.3666
                                   0.3480
                                            0.2854
```

0.8146

0.9177

Balanced Accuracy

0.9085

```
# Polinomial
cmP1<-confusionMatrix(test$tipoDeCasa,prediccionP1)</pre>
cmP2<-confusionMatrix(test$tipoDeCasa,prediccionP2)</pre>
cmP3<-confusionMatrix(test$tipoDeCasa,prediccionP3)</pre>
# Polinomial
cmP1
## Confusion Matrix and Statistics
##
##
            Reference
## Prediction 1 2
##
           1 145 13
                       0
##
           2 23 120
                       7
##
           3
              1 46 76
## Overall Statistics
##
##
                 Accuracy : 0.7912
                   95% CI : (0.7497, 0.8286)
##
##
      No Information Rate: 0.4153
##
      P-Value [Acc > NIR] : < 2.2e-16
##
##
                    Kappa: 0.682
##
##
  Mcnemar's Test P-Value: 4.154e-07
## Statistics by Class:
##
##
                       Class: 1 Class: 2 Class: 3
## Sensitivity
                         0.8580 0.6704
                                          0.9157
## Specificity
                         0.9504 0.8810
                                           0.8649
## Pos Pred Value
                         0.9177 0.8000
                                          0.6179
                         0.9121 0.7900
## Neg Pred Value
                                          0.9773
## Prevalence
                         0.3921 0.4153
                                          0.1926
## Detection Rate
                         0.3364 0.2784
                                          0.1763
## Detection Prevalence
                         0.3666 0.3480
                                           0.2854
                         0.9042
## Balanced Accuracy
                                  0.7757
                                           0.8903
# Polinomial
cmP2
## Confusion Matrix and Statistics
##
##
            Reference
## Prediction 1 2
                       3
           1 143 15
##
##
           2 23 108 19
##
           3
              0 23 100
##
## Overall Statistics
```

Accuracy : 0.8144

##

```
95% CI : (0.7744, 0.85)
##
##
      No Information Rate: 0.3852
      P-Value [Acc > NIR] : < 2.2e-16
##
##
##
                     Kappa: 0.7197
##
  Mcnemar's Test P-Value : NA
##
## Statistics by Class:
##
##
                        Class: 1 Class: 2 Class: 3
                          0.8614 0.7397
## Sensitivity
                                            0.8403
## Specificity
                                            0.9263
                          0.9434
                                  0.8526
## Pos Pred Value
                          0.9051
                                   0.7200
                                            0.8130
## Neg Pred Value
                          0.9158
                                   0.8648
                                            0.9383
## Prevalence
                          0.3852
                                   0.3387
                                            0.2761
## Detection Rate
                          0.3318
                                            0.2320
                                   0.2506
## Detection Prevalence
                          0.3666
                                   0.3480
                                            0.2854
## Balanced Accuracy
                          0.9024
                                   0.7962
                                            0.8833
# Polinomial
cmP3
## Confusion Matrix and Statistics
##
##
             Reference
## Prediction
                1
                    2
                        3
            1 140 18
            2 21 111
##
                      18
##
                0 20 103
##
## Overall Statistics
##
##
                  Accuracy: 0.8213
##
                    95% CI: (0.7819, 0.8564)
##
      No Information Rate: 0.3735
##
      P-Value [Acc > NIR] : < 2.2e-16
##
##
                     Kappa: 0.7304
##
##
   Mcnemar's Test P-Value : NA
##
## Statistics by Class:
##
                        Class: 1 Class: 2 Class: 3
##
## Sensitivity
                          0.8696
                                 0.7450
                                            0.8512
## Specificity
                          0.9333
                                   0.8617
                                            0.9355
## Pos Pred Value
                                   0.7400
                                            0.8374
                          0.8861
## Neg Pred Value
                          0.9231
                                   0.8648
                                            0.9416
## Prevalence
                          0.3735
                                 0.3457
                                            0.2807
## Detection Rate
                          0.3248 0.2575
                                            0.2390
## Detection Prevalence
                          0.3666
                                   0.3480
                                            0.2854
```

0.8934

0.8033

0.9014

Balanced Accuracy

- 7. Analice si los modelos están sobreajustados o desajustados. ¿Qué puede hacer para manejar el sobreajuste o desajuste?
- 8. Compare los resultados obtenidos con los diferentes modelos que hizo en cuanto a efectividad, tiempo de procesamiento y equivocaciones (donde el algoritmo se equivocó más, donde se equivocó menos y la importancia que tienen los errores).
- 9. Compare la eficiencia del mejor modelo de SVM con los resultados obtenidos en los algoritmos de las hojas de trabajo anteriores que usen la misma variable respuesta (árbol de decisión y random forest, naive bayes). ¿Cuál es mejor para predecir? ¿Cuál se demoró más en procesar?
- 10. Genere un buen modelo de regresión, use para esto la variable del precio de la casa directamente.
- 11. Compare los resultados del modelo de regresión generado con los de hojas anteriores que utilicen la misma variable, como la de regresión lineal.
- 12. Genere un informe de los resultados y las explicaciones.