BayesIngenuo

Cristopher Barrios, Carlos Daniel Estrada

2023-03-17

librerias

```
library(rpart)
library(rpart.plot)
library(dplyr)
```

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(fpc)
library(cluster)
library("ggpubr")
```

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

```
library(mclust)
```

```
## Package 'mclust' version 6.0.0
## Type 'citation("mclust")' for citing this R package in publications.
```

```
library(caret)
```

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

```
library(tree)
library(randomForest)
```

```
## randomForest 4.7-1.1

## Type rfNews() to see new features/changes/bug fixes.
```

```
##
## Attaching package: 'randomForest'

## The following object is masked from 'package:ggplot2':
##
##     margin

## The following object is masked from 'package:dplyr':
##
##     combine
```

```
library(plyr)
```

```
## ------------------------------------------------------------------------------

## You have loaded plyr after dplyr - this is likely to cause problems.
## If you need functions from both plyr and dplyr, please load plyr first, then dplyr:
## library(plyr); library(dplyr)

## ------------------------------------------------------------------------------

##
## Attaching package: 'plyr'

## The following object is masked from 'package:ggpubr':
##
##     mutate

## The following objects are masked from 'package:dplyr':
##
##     arrange, count, desc, failwith, id, mutate, rename, summarise,
##     summarize
```

```
library("stats")
library("datasets")
library("prediction")
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.2
## --

## v tibble  3.1.8     v purrr   1.0.1
## v tidyr   1.3.0     v stringr 1.5.0
## v readr   2.1.3     v forcats 1.0.0
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x plyr::arrange()         masks dplyr::arrange()
## x randomForest::combine() masks dplyr::combine()
## x purrr::compact()        masks plyr::compact()
## x plyr::count()           masks dplyr::count()
## x plyr::failwith()        masks dplyr::failwith()
```

```
## x dplyr::filter()         masks stats::filter()
## x plyr::id()              masks dplyr::id()
## x dplyr::lag()            masks stats::lag()
## x purrr::lift()           masks caret::lift()
## x purrr::map()            masks mclust::map()
## x randomForest::margin()  masks ggplot2::margin()
## x plyr::mutate()          masks ggpubr::mutate(), dplyr::mutate()
## x plyr::rename()          masks dplyr::rename()
## x plyr::summarise()       masks dplyr::summarise()
## x plyr::summarize()       masks dplyr::summarize()
```

```
library(e1071)
```

1. Use los mismos conjuntos de entrenamiento y prueba que utilizó en las dos hojas anteriores.

```
datos = read.csv("./train.csv")

test<- read.csv("./test.csv", stringsAsFactors = FALSE)
```

```
set_entrenamiento <- sample_frac(datos, .7)
set_prueba <-setdiff(datos, set_entrenamiento)


drop <- c("LotFrontage", "Alley", "MasVnrType", "MasVnrArea", "BsmtQual", "BsmtCond", "BsmtExposure", "
set_entrenamiento <- set_entrenamiento[, !(names(set_entrenamiento) %in% drop)]
set_prueba <- set_prueba[, !(names(set_prueba) %in% drop)]
```

2. Elabore un modelo de regresión usando bayes ingenuo (naive bayes), el conjunto de entrenamiento y la variable respuesta SalesPrice. Prediga con el modelo y explique los resultados a los que llega. Asegúrese que los conjuntos de entrenamiento y prueba sean los mismos de las hojas anteriores para que los modelos sean comparables.

```
#percentiles
percentil <- quantile(datos$SalePrice)


estado<-c('Estado')
datos$Estado<-estado
datos <- within(datos, Estado[SalePrice<=129975] <- 'Economica')
datos$Estado[(datos$SalePrice>129975 &datos$SalePrice<=163000)] <- 'Intermedia'
datos$Estado[datos$SalePrice>163000] <- 'Cara'


#Bayes
porcentaje<-0.7

set.seed(1234)
corte <- sample(nrow(datos),nrow(datos)*porcentaje)
```

```r
#Entrenamiento
train<-datos[corte,]
#Prueba
test<-datos[-corte,]
```

3. Haga un modelo de clasificación, use la variable categórica que hizo con el precio de las casas (barata, media y cara) como variable respuesta.

```r
#modelo
modelo<-naiveBayes(train$Estado~., data=train)

#Casting
test$GrLivArea<-as.numeric(test$GrLivArea)
test$YearBuilt<-as.numeric(test$YearBuilt)
test$BsmtUnfSF<-as.numeric(test$BsmtUnfSF)
test$TotalBsmtSF<-as.numeric(test$TotalBsmtSF)
test$GarageArea<-as.numeric(test$GarageArea)
test$YearRemodAdd<-as.numeric(test$YearRemodAdd)
test$SalePrice<-as.numeric(test$SalePrice)
test$LotArea<-as.numeric(test$LotArea)

#prediccion
predBayes<-predict(modelo, newdata = test[,c("GrLivArea","YearBuilt","BsmtUnfSF","TotalBsmtSF","GarageA

#Convertimos
predBayes<-as.factor(predBayes)
```

4. Utilice los modelos con el conjunto de prueba y determine la eficiencia del algoritmo para predecir y clasificar.

```r
prediction <- predict(modelo, test)
prediction
```

```
##   [1] Cara       Intermedia Cara       Cara       Cara       Cara
##   [7] Cara       Intermedia Economica  Economica  Economica  Cara
##  [13] Economica  Economica  Cara       Cara       Cara       Economica
##  [19] Cara       Economica  Cara       Cara       Economica  Economica
##  [25] Cara       Economica  Cara       Economica  Cara       Economica
##  [31] Economica  Economica  Economica  Cara       Economica  Economica
##  [37] Economica  Economica  Cara       Intermedia Cara       Economica
##  [43] Cara       Cara       Economica  Cara       Cara       Cara
##  [49] Economica  Cara       Cara       Economica  Cara       Economica
##  [55] Cara       Economica  Cara       Economica  Cara       Cara
##  [61] Cara       Cara       Cara       Cara       Economica  Economica
##  [67] Cara       Cara       Cara       Economica  Cara       Cara
##  [73] Cara       Intermedia Economica  Cara       Cara       Cara
##  [79] Cara       Economica  Cara       Economica  Cara       Economica
##  [85] Economica  Cara       Cara       Economica  Cara       Economica
##  [91] Economica  Cara       Cara       Cara       Intermedia Cara
```

```
##  [97] Cara       Cara       Intermedia Intermedia Intermedia Intermedia
## [103] Cara       Cara       Economica  Economica  Cara       Cara
## [109] Cara       Cara       Cara       Economica  Economica  Economica
## [115] Economica  Cara       Intermedia Economica  Cara       Cara
## [121] Economica  Intermedia Cara       Economica  Cara       Intermedia
## [127] Economica  Economica  Intermedia Cara       Economica  Cara
## [133] Cara       Economica  Cara       Cara       Cara       Economica
## [139] Economica  Economica  Economica  Economica  Cara       Cara
## [145] Economica  Cara       Economica  Intermedia Economica  Economica
## [151] Economica  Cara       Economica  Economica  Intermedia Economica
## [157] Cara       Economica  Cara       Economica  Intermedia Economica
## [163] Economica  Cara       Intermedia Cara       Economica  Intermedia
## [169] Cara       Cara       Cara       Intermedia Economica  Cara
## [175] Cara       Cara       Cara       Intermedia Cara       Cara
## [181] Cara       Economica  Cara       Economica  Cara       Cara
## [187] Economica  Intermedia Cara       Intermedia Economica  Cara
## [193] Economica  Economica  Economica  Cara       Cara       Economica
## [199] Cara       Economica  Cara       Cara       Economica  Cara
## [205] Intermedia Cara       Cara       Economica  Economica  Cara
## [211] Economica  Cara       Economica  Cara       Intermedia Economica
## [217] Economica  Cara       Economica  Cara       Intermedia Cara
## [223] Cara       Economica  Cara       Cara       Cara       Cara
## [229] Economica  Economica  Cara       Cara       Cara       Economica
## [235] Economica  Economica  Cara       Economica  Economica  Cara
## [241] Economica  Cara       Economica  Economica  Cara       Economica
## [247] Economica  Intermedia Economica  Cara       Intermedia Cara
## [253] Cara       Intermedia Cara       Intermedia Cara       Economica
## [259] Economica  Economica  Cara       Economica  Economica  Cara
## [265] Intermedia Economica  Intermedia Intermedia Economica  Cara
## [271] Economica  Economica  Cara       Cara       Intermedia Economica
## [277] Economica  Cara       Economica  Cara       Economica  Intermedia
## [283] Cara       Cara       Cara       Economica  Cara       Intermedia
## [289] Cara       Cara       Economica  Economica  Economica  Intermedia
## [295] Economica  Economica  Economica  Cara       Cara       Cara
## [301] Cara       Economica  Intermedia Economica  Cara       Intermedia
## [307] Intermedia Cara       Cara       Intermedia Economica  Cara
## [313] Cara       Cara       Economica  Economica  Cara       Economica
## [319] Intermedia Intermedia Cara       Economica  Cara       Economica
## [325] Intermedia Economica  Cara       Cara       Economica  Economica
## [331] Economica  Cara       Economica  Economica  Cara       Economica
## [337] Cara       Economica  Economica  Economica  Economica  Economica
## [343] Cara       Economica  Intermedia Cara       Economica  Cara
## [349] Cara       Cara       Economica  Cara       Cara       Economica
## [355] Intermedia Cara       Cara       Cara       Economica  Cara
## [361] Intermedia Cara       Economica  Economica  Economica  Intermedia
## [367] Cara       Economica  Economica  Cara       Economica  Intermedia
## [373] Economica  Economica  Cara       Economica  Economica  Cara
## [379] Cara       Economica  Cara       Cara       Economica  Intermedia
## [385] Intermedia Cara       Economica  Intermedia Cara       Cara
## [391] Cara       Cara       Cara       Economica  Cara       Economica
## [397] Cara       Cara       Economica  Economica  Cara       Economica
## [403] Intermedia Intermedia Economica  Cara       Cara       Cara
## [409] Economica  Cara       Intermedia Cara       Cara       Economica
## [415] Economica  Cara       Economica  Cara       Cara       Economica
```

```
## [421] Cara       Intermedia Economica  Cara       Cara       Economica
## [427] Cara       Intermedia Economica  Economica  Cara       Economica
## [433] Economica  Cara       Cara       Cara       Cara       Economica
## [439] Intermedia
## Levels: Cara Economica Intermedia
```

5. Analice los resultados del modelo de regresión. ¿Qué tan bien le fue prediciendo?

6. Compare los resultados con el modelo de regresión lineal y el árbol de regresión que hizo en las hojas pasadas. ¿Cuál funcionó mejor?

7. Haga un análisis de la eficiencia del modelo de clasificación usando una matriz de confusión. Tenga en cuenta la efectividad, donde el algoritmo se equivocó más, donde se equivocó menos y la importancia que tienen los errores.

```
#confusion
cm<-caret::confusionMatrix(as.factor(predBayes),as.factor(test$Estado))
cm
```

```
## Confusion Matrix and Statistics
##
##             Reference
## Prediction   Cara Economica Intermedia
##   Cara        204         1          4
##   Economica     2       100         11
##   Intermedia   16         5         96
##
## Overall Statistics
##
##                Accuracy : 0.9112
##                  95% CI : (0.8806, 0.9361)
##     No Information Rate : 0.5057
##     P-Value [Acc > NIR] : <2e-16
##
##                   Kappa : 0.8589
##
##  Mcnemar's Test P-Value : 0.0205
##
## Statistics by Class:
##
##                      Class: Cara Class: Economica Class: Intermedia
## Sensitivity               0.9189           0.9434            0.8649
## Specificity               0.9770           0.9610            0.9360
## Pos Pred Value            0.9761           0.8850            0.8205
## Neg Pred Value            0.9217           0.9816            0.9534
## Prevalence                0.5057           0.2415            0.2528
## Detection Rate            0.4647           0.2278            0.2187
## Detection Prevalence      0.4761           0.2574            0.2665
## Balanced Accuracy         0.9479           0.9522            0.9004
```

8. Analice el modelo. ¿Cree que pueda estar sobre ajustado?

Si un modelo presenta un alto nivel de precisión y porcentajes de comportamiento similares, es posible suponer que haya ocurrido un sobreajuste. Sin embargo, para confirmarlo, es necesario compararlo con otro conjunto de datos mediante la validación cruzada. De esta forma, podremos determinar si realmente ha habido sobreajuste o no.

9. Haga un modelo usando validación cruzada, compare los resultados de este con los del modelo anterior. ¿Cuál funcionó mejor?

10. Compare la eficiencia del algoritmo con el resultado obtenido con el árbol de decisión (el de clasificación) y el modelo de random forest que hizo en la hoja pasada. ¿Cuál es mejor para predecir? ¿Cuál se demoró más en procesar?