

Modelos de Clasificacion

Daniela Cuesta - Paola Peralta Flores

Se realiza los diferentes metodos de clasificacion.

```
library(ggplot2)
library(ggpubr)
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(glmnet) ## regresiones logisitcas
```

Loading required package: Matrix

Loaded glmnet 4.1-7

```
library(caret) ### bayes y knn
```

Loading required package: lattice

```

library(e1071) ## bayes

# quitamos la primera columna
datos <- read.table("./yeast.data",header = F)[,-1]

# Funciones de transformacion
min.max.mean <- function(X) apply(X,2,function(x) (x-mean(x))/(max(x)-min(x)))
min.max.median <- function(X) apply(X,2,function(x) (x-median(x))/(max(x)-min(x)))
min.max <- function(X) apply(X,2,function(x) (x-min(x))/(max(x)-min(x)))
zscore <- function(X) apply(X,2,function(x) (x-mean(x))/sd(x))
l2 <- function(X) apply(X,2,function(x) x/sqrt(sum(x^2)))

#Particion de datos
datos <- as.data.frame(datos)
datos.numericos <- datos[, which(unlist(lapply(datos, is.numeric)))]
clase <- datos$V10 <- as.factor(datos$V10)
colnames(datos.numericos) <- paste0("Var", rep(1:8))

### procedemos a crear una lista con todas las transformaciones
datos.lista <- list(
  raw = bind_cols(datos.numericos,clase=clase),
  zscore = bind_cols(zscore(datos.numericos),
                     clase = clase),
  l2 = bind_cols(l2(datos.numericos), clase = clase),
  media = bind_cols(min.max.mean(datos.numericos), clase =
                    clase),
  mediana = bind_cols(min.max.median(datos.numericos), clase =
                     clase),
  min_max = bind_cols(min.max(datos.numericos),
                     clase = clase))

#Al ser demasiadas variables, podemos realizar un melt
lista_graficos <- vector("list",length=length(datos.lista))
datos.melt <- lapply(datos.lista,reshape2::melt)

```

Using clase as id variables

Using clase as id variables

Using clase as id variables

Using clase as id variables

Using clase as id variables

Using class as id variables

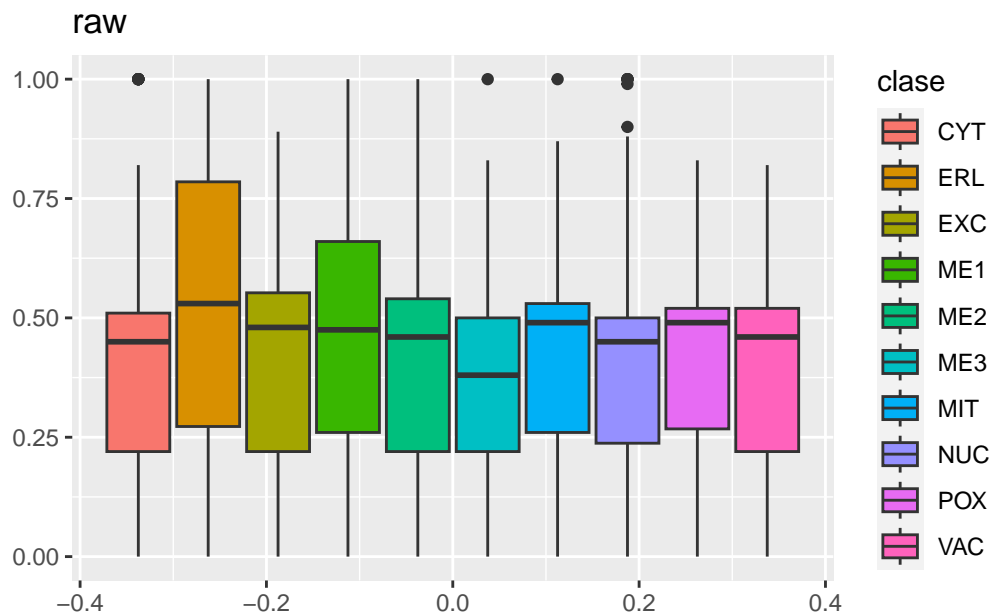
```
#graficos
for(l in 1:length(datos.melt)){

  X <- datos.melt[[l]]
  nombre <- names(datos.melt)[l]
  lista_graficos[[l]] <- ggplot(X,aes(y=value,fill=clase))+geom_boxplot()+ggtitle(nombre)+

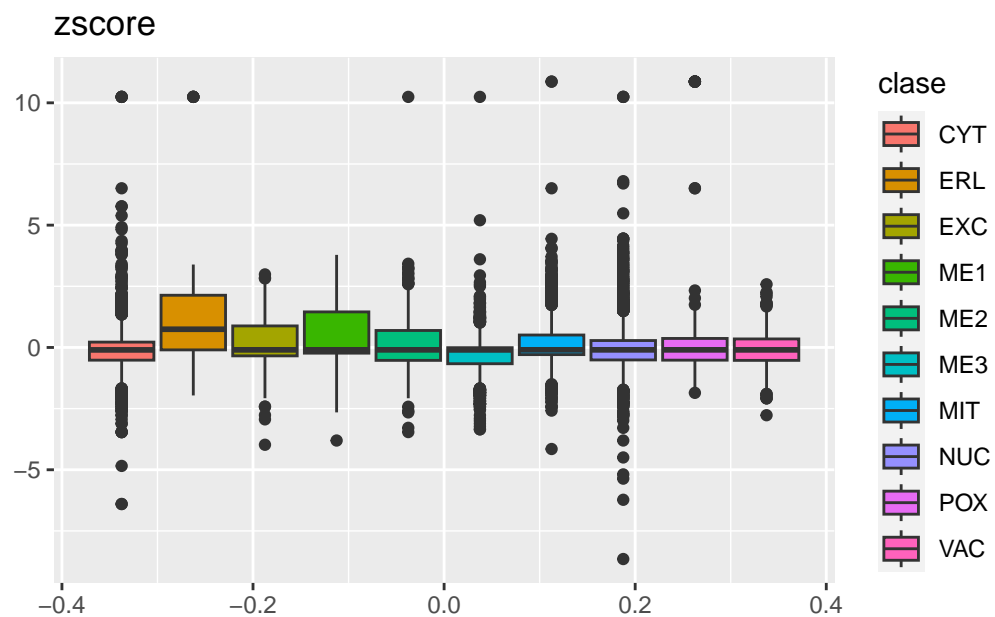
}

names(lista_graficos) <- paste0("plot",1:length(datos.lista))

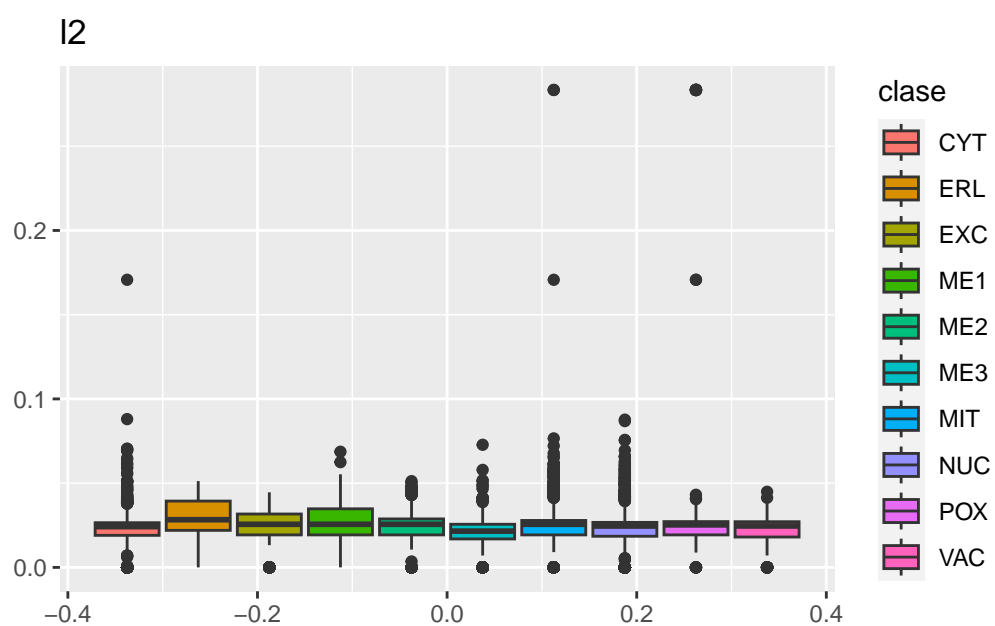
lista_graficos$plot1
```



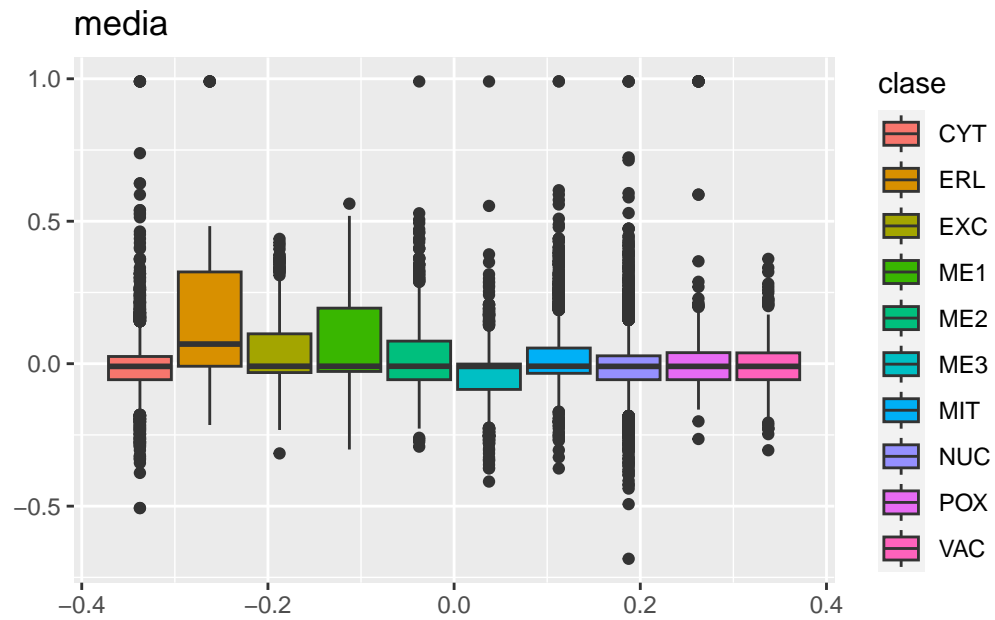
```
lista_graficos$plot2
```



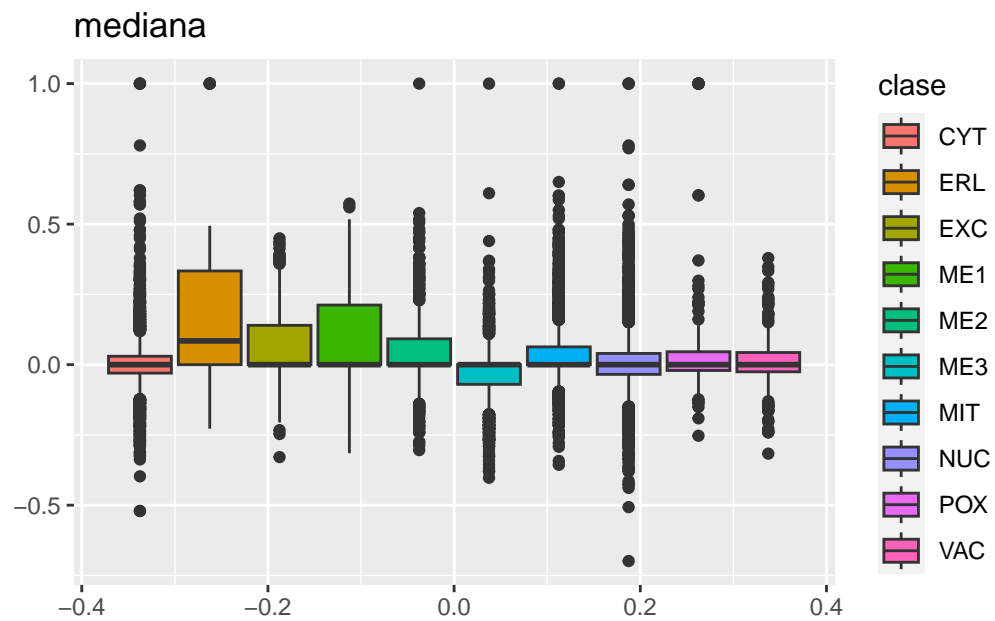
```
lista_graficos$plot3
```



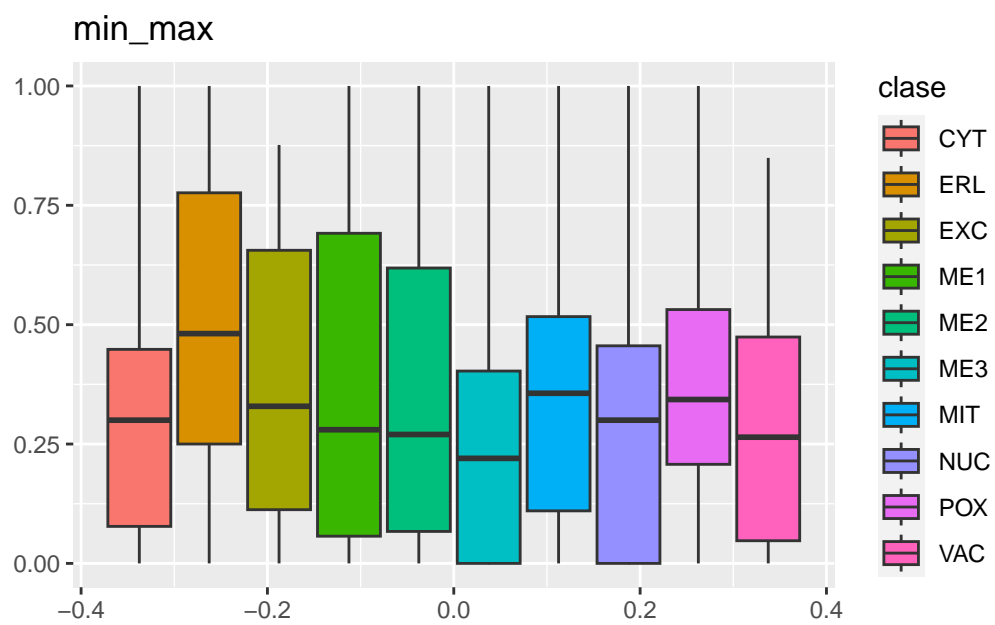
```
lista_graficos$plot4
```



```
lista_graficos$plot5
```



```
lista_graficos$plot6
```



```

#grafico de densidad
for(l in 1:length(datos.melt)){

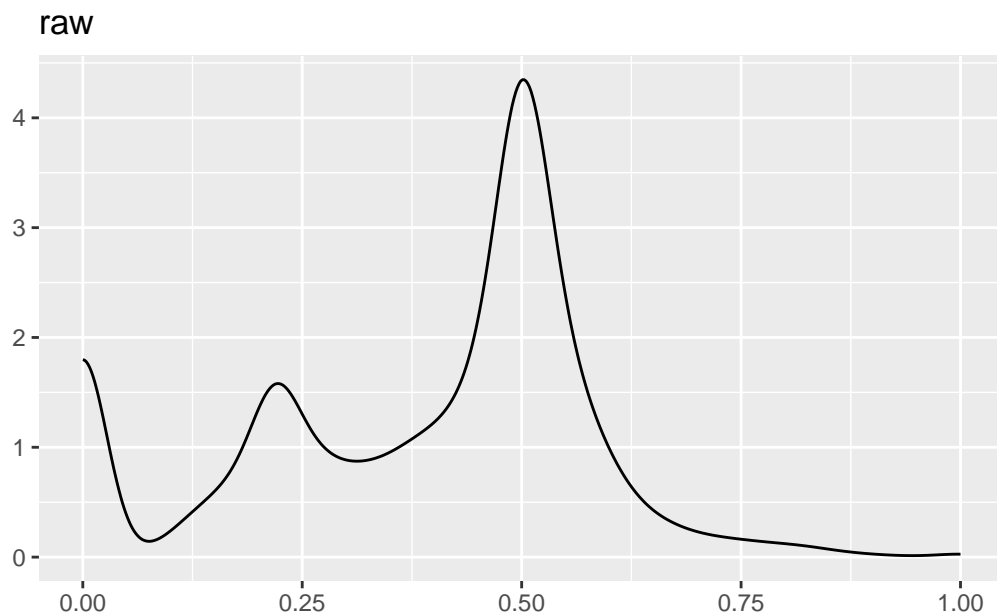
  X <- datos.melt[[l]]
  nombre <- names(datos.melt)[l]
  lista_graficos[[l]] <- ggplot(X,aes(x=value))+geom_density()+ggtitle(nombre)+xlab("")+yl

}

names(lista_graficos) <- paste0("plot",1:length(datos.lista))

lista_graficos$plot1

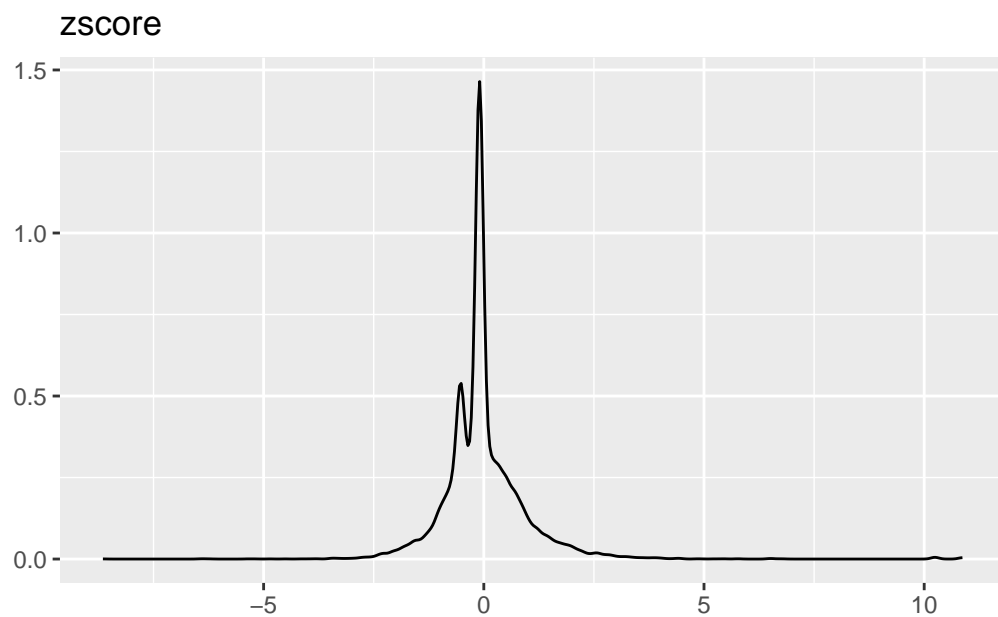
```



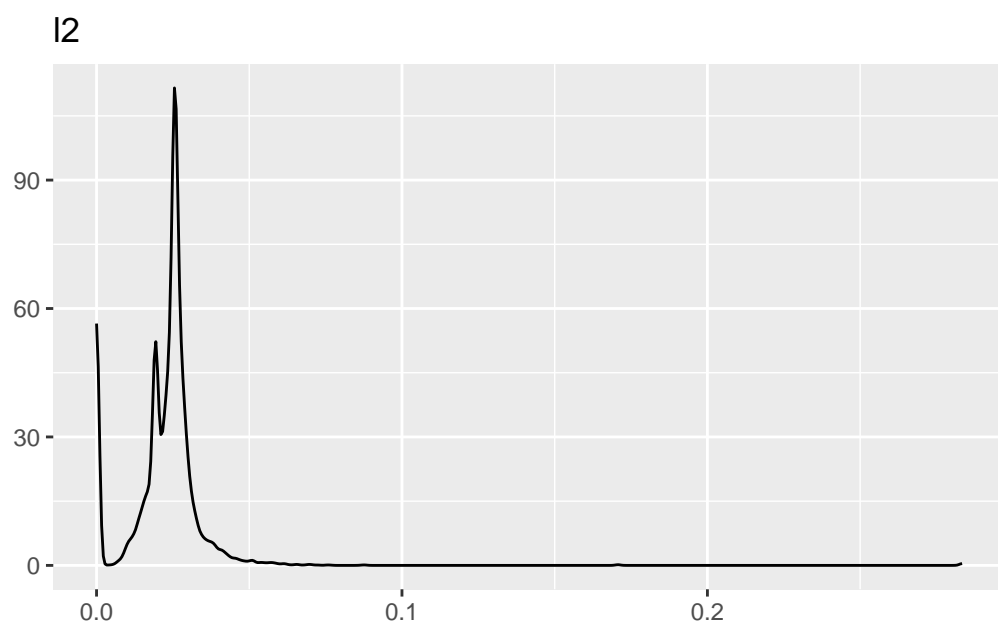
```

lista_graficos$plot2

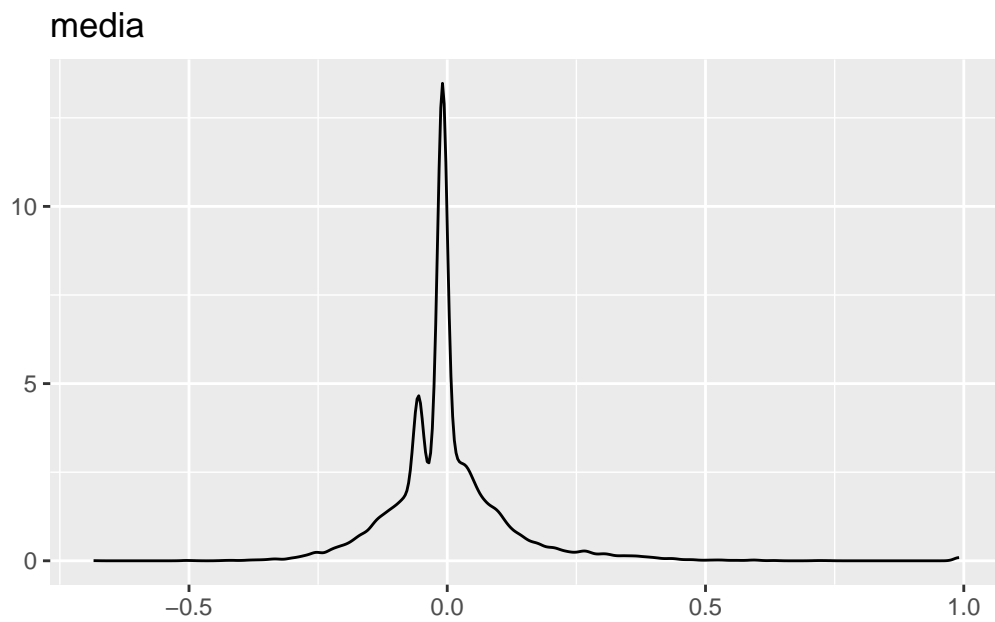
```



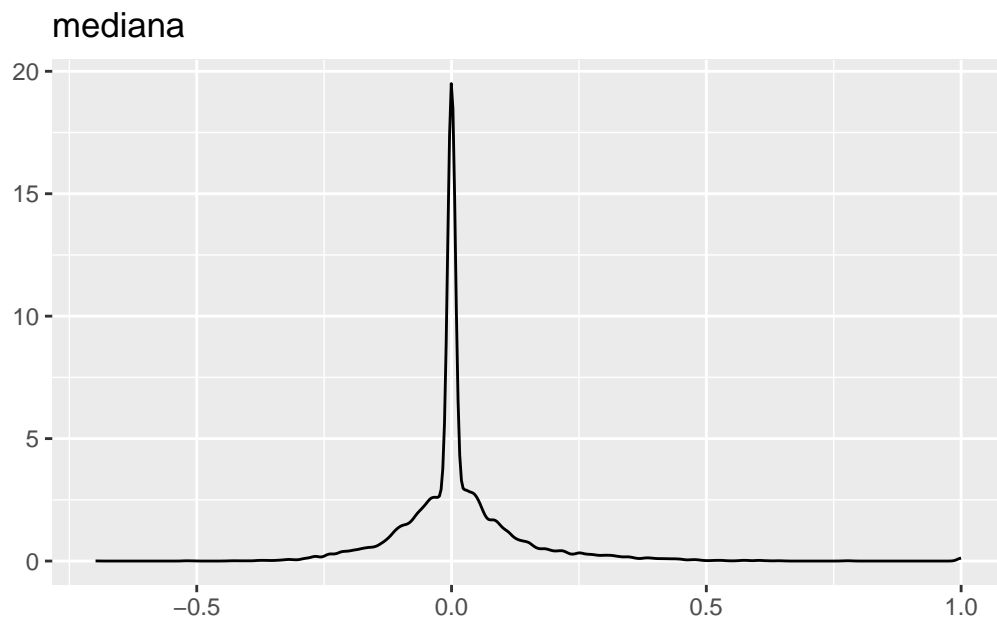
```
lista_graficos$plot3
```



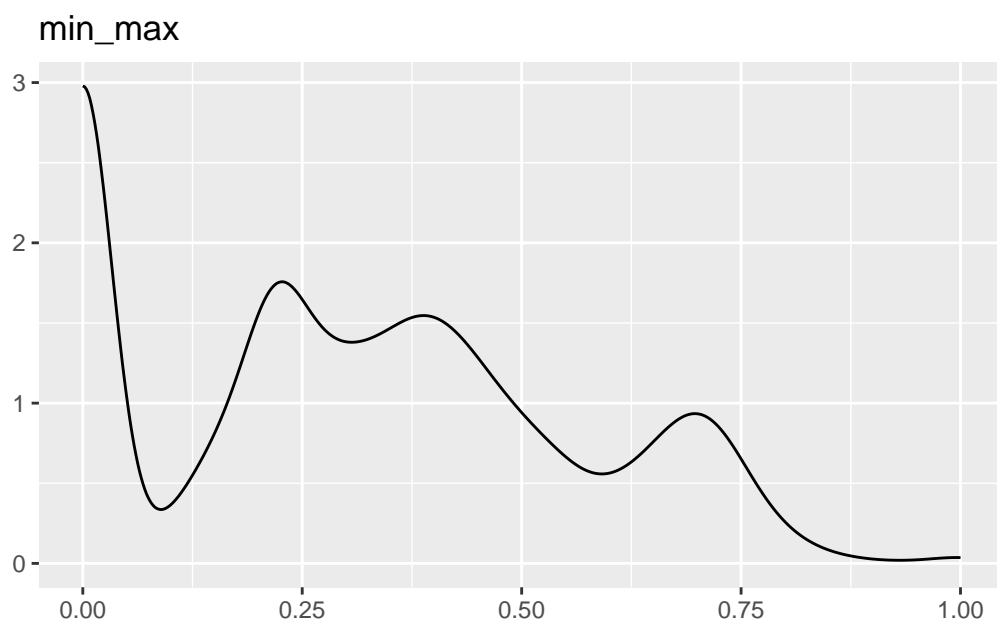

```
lista_graficos$plot4
```



```
lista_graficos$plot5
```



```
lista_graficos$plot6
```



```

#Fijamos la semilla y la muestra
set.seed(123456789)
trControl <- trainControl(method = 'cv', number = 10)
n <- nrow(datos)
idx <- sample(1:n,size=n*0.7,replace=F)
lambda_seq <- seq(0.01, 1, by = 0.01)

### para conjunto de datos podemos realizar el split
entrenamiento <- lapply(datos.lista, function(x) x[idx,])
test <- lapply(datos.lista, function(x) x[-idx,])

#Regresion logistica Lineal
set.seed(123456789)
myfnlog <- function(x) train(clase ~ ., data = x, method = "multinom", trControl = trControl)
logistica.lista <- lapply(entrenamiento,myfnlog)
logisita.pred <- vector("list",length = length(datos.lista))
for(l in 1:length(datos.lista)){
  logisita.pred[[l]] <- predict(logistica.lista[[l]],test[[l]])
}
names(logisita.pred) <- names(datos.lista)
accuracy <- vector("numeric",length = length(datos.lista))
for(l in 1:length(datos.lista)){
  accuracy[l] <- confusionMatrix(test$raw$class,logisita.pred[[l]])$overall[1]
}
names(accuracy) <- names(datos.lista)
accuracy_logis<-accuracy

#Ridge
set.seed(123456789)
myfnridge <- function(x) train(clase ~ ., data = x, method = "glmnet", trControl = trControl)

logistica.lista <- lapply(entrenamiento,myfnridge)

```

Warning in lognet(xd, is.sparse, ix, jx, y, weights, offset, alpha, nobs, : one multinomial or binomial class has fewer than 8 observations; dangerous ground

Warning in lognet(xd, is.sparse, ix, jx, y, weights, offset, alpha, nobs, : one multinomial or binomial class has fewer than 8 observations; dangerous ground

Warning in lognet(xd, is.sparse, ix, jx, y, weights, offset, alpha, nobs, : one multinomial or binomial class has fewer than 8 observations; dangerous ground

Warning in lognet(xd, is.sparse, ix, jx, y, weights, offset, alpha, nobs, : one multinomial or binomial class has fewer than 8 observations; dangerous ground

Warning in lognet(xd, is.sparse, ix, jx, y, weights, offset, alpha, nobs, : one multinomial or binomial class has fewer than 8 observations; dangerous ground

Warning in lognet(xd, is.sparse, ix, jx, y, weights, offset, alpha, nobs, : one multinomial or binomial class has fewer than 8 observations; dangerous ground

Warning in lognet(xd, is.sparse, ix, jx, y, weights, offset, alpha, nobs, : one multinomial or binomial class has fewer than 8 observations; dangerous ground

Warning in lognet(xd, is.sparse, ix, jx, y, weights, offset, alpha, nobs, : one multinomial or binomial class has fewer than 8 observations; dangerous ground

Warning in lognet(xd, is.sparse, ix, jx, y, weights, offset, alpha, nobs, : one multinomial or binomial class has fewer than 8 observations; dangerous ground

```
logisita.pred <- vector("list",length = length(datos.lista))
for(l in 1:length( datos.lista)){
  logisita.pred[[l]] <- predict(logistica.lista[[l]],test[[l]])
}
names(logisita.pred) <- names(datos.lista)
accuracy <- vector("numeric",length = length(datos.lista))
for(l in 1:length(datos.lista)){
  accuracy[l] <- confusionMatrix(test$raw$class,logisita.pred[[l]])$overall[1]
}
names(accuracy) <- names(datos.lista)
accuracy_ride <- accuracy

#Lasso
set.seed(123456789)
myfnlasso <- function(x) train(class ~ ., data = x, method = "glmnet", trControl = trContr

logistica.lista <- lapply(entrenamiento,myfnlasso)
```

Warning in lognet(xd, is.sparse, ix, jx, y, weights, offset, alpha, nobs, : one multinomial or binomial class has fewer than 8 observations; dangerous ground

Warning in lognet(xd, is.sparse, ix, jx, y, weights, offset, alpha, nobs, : one multinomial or binomial class has fewer than 8 observations; dangerous ground

Warning in lognet(xd, is.sparse, ix, jx, y, weights, offset, alpha, nobs, : one multinomial or binomial class has fewer than 8 observations; dangerous ground

Warning in lognet(xd, is.sparse, ix, jx, y, weights, offset, alpha, nobs, : one multinomial or binomial class has fewer than 8 observations; dangerous ground

Warning in lognet(xd, is.sparse, ix, jx, y, weights, offset, alpha, nobs, : one multinomial or binomial class has fewer than 8 observations; dangerous ground

Warning in lognet(xd, is.sparse, ix, jx, y, weights, offset, alpha, nobs, : one multinomial or binomial class has fewer than 8 observations; dangerous ground

Warning in lognet(xd, is.sparse, ix, jx, y, weights, offset, alpha, nobs, : one multinomial or binomial class has fewer than 8 observations; dangerous ground

Warning in lognet(xd, is.sparse, ix, jx, y, weights, offset, alpha, nobs, : one multinomial or binomial class has fewer than 8 observations; dangerous ground

Warning in lognet(xd, is.sparse, ix, jx, y, weights, offset, alpha, nobs, : one multinomial or binomial class has fewer than 8 observations; dangerous ground

```
logisita.pred <- vector("list",length = length(datos.lista))
for(l in 1:length( datos.lista)){
  logisita.pred[[l]] <- predict(logistica.lista[[l]],test[[l]])
}
names(logisita.pred) <- names(datos.lista)
accuracy <- vector("numeric",length = length(datos.lista))
for(l in 1:length(datos.lista)){
  accuracy[l] <- confusionMatrix(test$raw$clase,logisita.pred[[l]])$overall[1]
}
names(accuracy) <- names(datos.lista)
accuracy_lasso <- accuracy
```