Taller 2 Regresión lineal Multiple

Andrés Felipe Palomino - David Stiven Rojas

2023-04-21

1 Introducción

La base de datos "*yarn*" obtenida de la librería (PLS) contiene información sobre espectros NIR y mediciones de densidad de hilos de PET, consta de 28 individuos (hilos de PET), 268 variables predictoras (NIRS) y una variable de respuesta (densidad). Se ajustará un modelo lineal múltiple para estimar la densidad del hilo PET, mediante mediciones NIR

```
#Importación de librerías necesarias
library(car)
library(glmnet)
library(MASS)
library(xtable)
library(lmtest)
library(readxl)
library(lmridge)
library(pls)
library(olsrr)
```

1.1 Base de datos

En la siguiente tabla se encuentra un encabezado de la base de datos que se trabajara, esta consta de 30 covariables predictoras, las cuales estarán desde NIR1 hasta NIR30. De primera mano se observa que los valores de los NIR disminuyen a medida que la covariable aumenta

```
X <- data.frame(matrix(c(yarn$NIR[,1:30],yarn$density),nrow =28, ncol= 31))
colnames(X) <- c(paste("NIR",1:30,sep=""),"density")

xtable(head(X[,1:11]))
```

% latex table generated in R 4.3.0 by xtable 1.8-4 package % Sat Apr 29 18:27:51 2023

	NIR1	NIR2	NIR3	NIR4	NIR5	NIR6	NIR7	NIR8	NIR9	NIR10	NIR11
1	3.07	3.09	3.11	3.10	3.00	2.83	2.62	2.40	2.19	2.01	1.84
2	3.07	3.09	3.10	3.07	2.98	2.84	2.68	2.51	2.35	2.22	2.12
3	3.08	3.10	3.09	3.03	2.88	2.69	2.48	2.27	2.08	1.92	1.77
4	3.08	3.10	3.10	3.07	2.99	2.87	2.74	2.61	2.50	2.42	2.38
5	3.10	3.10	3.08	3.02	2.89	2.72	2.54	2.38	2.24	2.13	2.05
6	3.08	3.08	3.05	2.93	2.73	2.51	2.29	2.10	1.93	1.79	1.67

```
xtable(head(X[,12:21]))
```

% latex table generated in R 4.3.0 by xtable 1.8-4 package % Sat Apr 29 18:27:51 2023

	NIR12	NIR13	NIR14	NIR15	NIR16	NIR17	NIR18	NIR19	NIR20	NIR21
1	1.69	1.58	1.50	1.44	1.34	1.22	1.14	1.12	1.13	1.16
2	2.04	1.98	1.96	1.94	1.89	1.82	1.75	1.71	1.68	1.65
3	1.65	1.55	1.49	1.44	1.35	1.26	1.20	1.18	1.19	1.21
4	2.35	2.35	2.37	2.40	2.40	2.38	2.33	2.28	2.21	2.11
5	1.99	1.95	1.94	1.93	1.90	1.85	1.80	1.76	1.73	1.68
6	1.56	1.48	1.43	1.39	1.32	1.25	1.20	1.19	1.19	1.19

```
xtable(head(X[,22:31]))
```

% latex table generated in R 4.3.0 by xtable 1.8-4 package % Sat Apr 29 18:27:51 2023

	NIR22	NIR23	NIR24	NIR25	NIR26	NIR27	NIR28	NIR29	NIR30	density
1	1.16	1.15	1.15	1.13	1.07	1.02	1.01	1.03	1.08	100.00
2	1.58	1.51	1.45	1.38	1.29	1.20	1.15	1.13	1.14	80.22
3	1.20	1.18	1.17	1.15	1.10	1.07	1.06	1.08	1.12	79.49
4	1.98	1.85	1.75	1.63	1.51	1.40	1.30	1.23	1.20	60.80
5	1.60	1.52	1.46	1.39	1.31	1.24	1.19	1.16	1.17	59.97
6	1.18	1.15	1.14	1.12	1.09	1.06	1.06	1.07	1.11	60.48

1.2 Funciones creadas

Antes de empezar con el proceso de seleccionar las variables para ajustar el modelo se crean funciones para optimizar el proceso de validación de supuestos.

```
##Validacion grafica para homocedasticidad y normalidad y pruebas formales
validaciongrafica<- function(model,cor=F){

  par(mfrow=c(1,2))
  plot(fitted.values(model),studres(model),panel.first=grid(),pch=19,ylab='Residuos Estudentizados',xla
  lines(lowess(studres(model)~fitted.values(model)), col = "red1")
  abline(h=c(-2,0,2),lty=2)
  qqPlot(model,pch=19,ylab='Residuos Estudentizados',xlab='Cuantiles Teóricos',col=carPalette()[1],col.
  print('Shapiro Test')
  print(shapiro.test(studres(model)))
  print('Breusch Pagan Test')
  print(bptest(model))
  if(cor==T){
    par(mfrow=c(1,2))
    plot(studres(model),type="b",xlab="Tiempo",ylab="Residuos Estudentizados",main="A",pch=19,panel.fir
    plot(studres(model)[-length(fitted.values(model))],studres(model)[-1],pch=19,panel.first = grid(),c
    abline(lm(studres(model)[-1]~studres(model)[-length(fitted.values(model))]))
    print('Durbin Watson Test')
    print(durbinWatsonTest(model,method='resample',reps=10000))
  }
  par(mfrow=c(1,1))
}
```

```
## Calculo de lambda optimo para boxcox
lambda<- function(model,a,b){
  par(mfrow=c(1,1))
  box.cox<-boxcox(model,lambda=seq(a,b,length.out = 1000),
                  ylab='log-verosimilitud')
  bc<-round(box.cox$x[box.cox$y ==max(box.cox$y)],2)
  print(bc)
}
```

2 Selección de variables

En el proceso de selección de variables se procede a realizar la Regresion de LASSO para identificar las posibles variables que tengan un aporte poco relevante, Por ultimo se ajustara el modelo cuyas variables tengan buenos indicadores y se pueda realizar corrección de supuestos

2.1 Regresión de LASSO

Este es un método de regularización que se implementa cuando se tiene muchas covariables disponibles y se cree que pocas tienen un aporte relevante.

Se asume el modelo de regresión usual, donde :

$$\text{E(y|x)}=x^T\beta,\text{ y V(y|x)}=\sigma^2$$

Donde se asume que algunos β son cero. El objetivo del estimador es seleccionar los coeficientes que tienen valores diferentes de cero. El cual se obtiene minimizando la siguiente expresión:

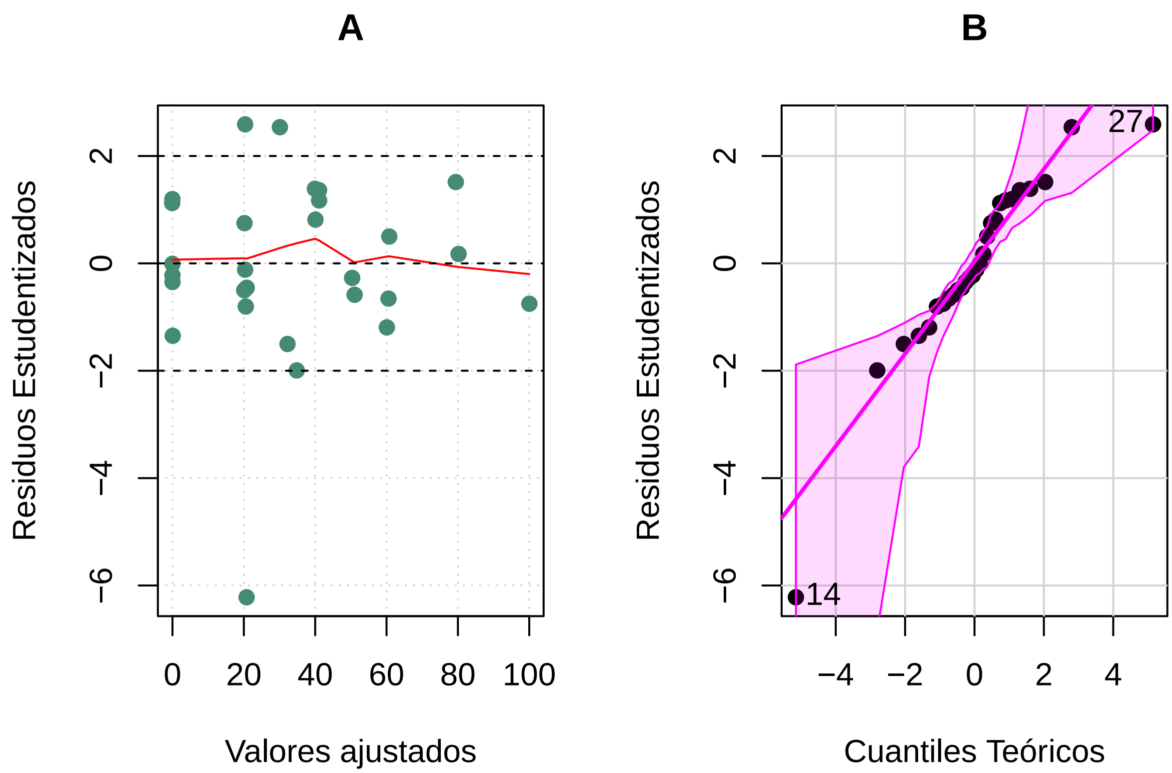
$$S_{lasso}(\beta) = \sum_{i=1}^{n}(y_i - x^T\beta)^2 + \lambda\sum_{j=1}^{p-1}|\beta_j|$$

Esta es la suma de cuadrados del estimador por MCO más una penalización (λ), a la suma del valor absoluto de los coeficientes. A medida que λ aumenta la penalización tendrá mas peso sobre la estimación de los coeficientes, es decir que si la penalización es muy grande, todas las estimaciones seran cero. No hay solución analitica para $\hat{\beta}_{lasso}$ por lo que se usan algoritmos para la estimación, como lo es la funcion de glmnet de la libreria glmnet.

2.1.1 Modelo a realizar regresión LASSO

Como se establecio anteriormente, se asume un modelo de regresión usual, el cual debe cumplir los siguientes supuestos: E(y|x)=$x^T\beta$, y V(y|x)=σ^2, por ende es necesario proponer un modelo con p<n, en el cual se eliminaran las variables con menor correlación con la variable y. Dicho modelo se expresa acontinuación y se evaluan los supuestos:

```
model <- lm(density ~ .-NIR1-NIR8-NIR9-NIR10-NIR11-NIR7, data=X)
validaciongrafica(model)
```



[1] "Shapiro Test"

```
Shapiro-Wilk normality test
```

data: studres(model) W = 0.86458, p-value = 0.001868

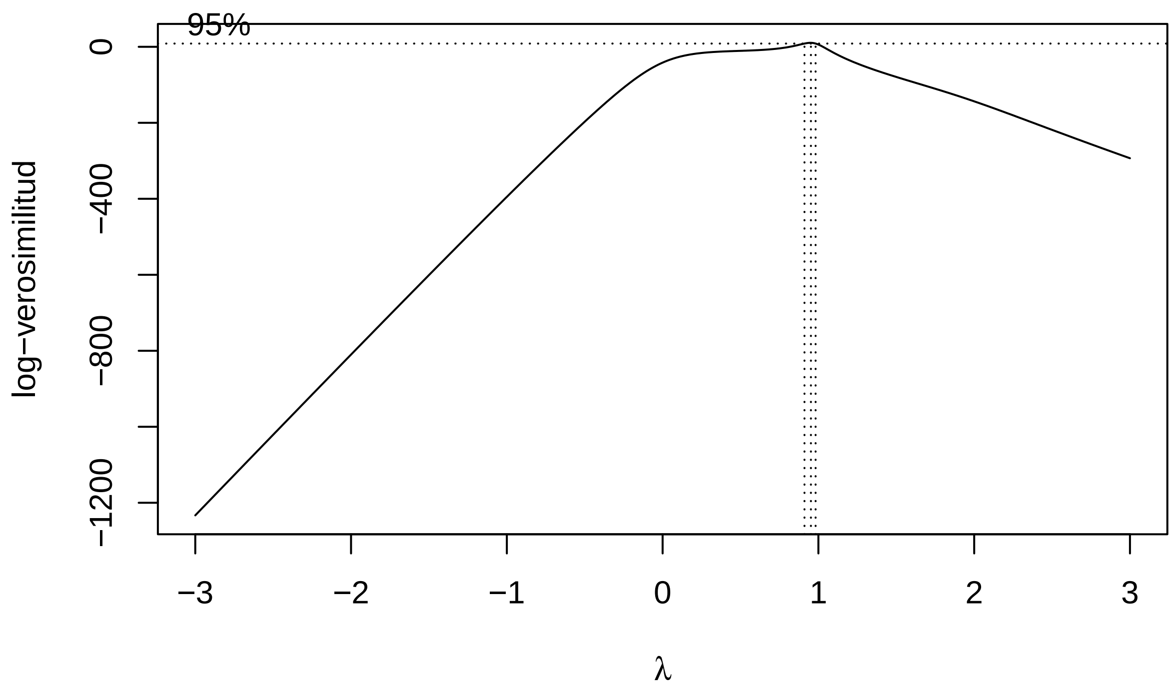
[1] "Breusch Pagan Test"

```
studentized Breusch-Pagan test
```

data: model BP = 27.288, df = 24, p-value = 0.2912

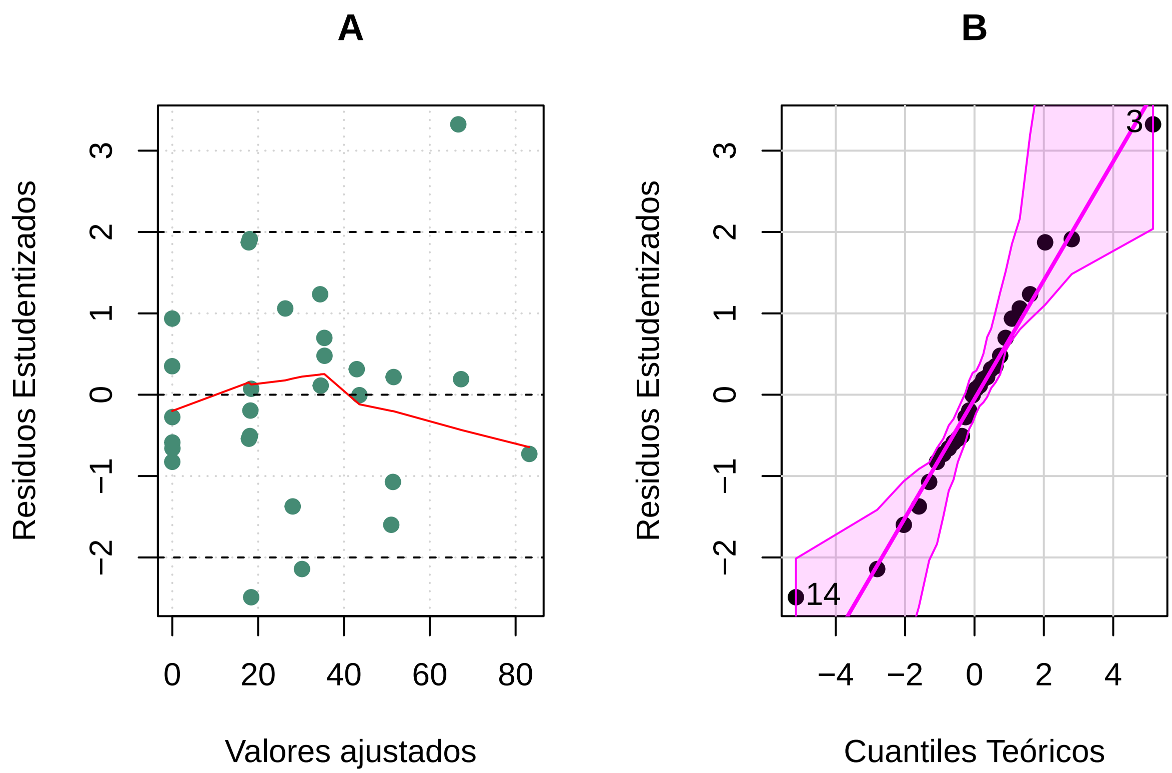
Como no se cumple el supuesto de normalidad se procede a corregir mediante el metodo de BoxCox y se verifica el cumplimiento de los mismos.

```
model <- lm(density+0.0000001 ~ .-NIR1-NIR8-NIR9-NIR10-NIR11-NIR7, data=X)
lambda(model,-3,3)
```



[1] 0.95

```
model.box <- lm(I(density^0.96) ~.-NIR1-NIR8-NIR9-NIR10-NIR11-NIR7,data=X)
validaciongrafica(model.box)
```



[1] "Shapiro Test"

```
Shapiro-Wilk normality test
```

data: studres(model) W = 0.97774, p-value = 0.7934

[1] "Breusch Pagan Test"

```
studentized Breusch-Pagan test
```

data: model BP = 23.94, df = 24, p-value = 0.4651

3

```
X.<-model.matrix(model.box)[,-1]
lasso.mod <- glmnet(X., X$density, alpha = 1,nlambda = 100)
lasso.mod$beta
```

24 x 81 sparse Matrix of class "dgCMatrix"

NIR2 . . . .
NIR3 . . . .
NIR4 . . 3.794159 12.13638 19.73747 26.66329 32.97384 38.72378 NIR5 . . . .
NIR6 . . . .

NIR12 . . . .
NIR13 . . . .
NIR14 . . . .
NIR15 . . . .
NIR16 . . . .
NIR17 . . . .
NIR18 . . . .
NIR19 . . . .
NIR20 . . . .
NIR21 . . . .
NIR22 . . . .
NIR23 . . . .
NIR24 . . . .
NIR25 . . . .
NIR26 . . . .
NIR27 . . . .
NIR28 . . . .
NIR29 . -12.70035 -25.722352 -39.45434 -51.96639 -63.36691 -73.75464 -83.21956 NIR30 . . . .

NIR2 . . . . . .
NIR3 . . . . . .
NIR4 43.96291 48.73662 53.08623 57.04944 60.66057 61.276309 NIR5 . . . . . 2.383012 NIR6 . . . . . .
NIR12 . . . . . .
NIR13 . . . . . .
NIR14 . . . . . .
NIR15 . . . . . .
NIR16 . . . . . .
NIR17 . . . . . .
NIR18 . . . . . .
NIR19 . . . . . .
NIR20 . . . . . .
NIR21 . . . . . .
NIR22 . . . . . .
NIR23 . . . . . .
NIR24 . . . . . .
NIR25 . . . . . .
NIR26 . . . . . .
NIR27 . . . . . .
NIR28 . . . . . .
NIR29 -91.84363 -99.70157 -106.86143 -113.38522 -119.32947 -125.150920 NIR30 . . . . . .

NIR2 . . . . . .
NIR3 . . . . . .
NIR4 47.80431 35.51972 24.32678 14.1290 4.83321 .
NIR5 16.84005 30.02088 42.03049 52.9725 62.94586 68.901969 NIR6 . . . . . .
NIR12 . . . . . .
NIR13 . . . . . .
NIR14 . . . . . .
NIR15 . . . . . .
NIR16 . . . . . .
NIR17 . . . . . .
NIR18 . . . . . .
NIR19 . . . . . .
NIR20 . . . . . .
NIR21 . . . . . .

NIR22 . . . . . .
NIR23 . . . . . .
NIR24 . . . . . .
NIR25 . . . . . .
NIR26 . . . . . .
NIR27 . . . . . .
NIR28 . . . . . -1.143204 NIR29 -132.45140 -139.10460 -145.16671 -150.6902 -155.72346 -157.985303 NIR30 . . . . . .

NIR2 . . . . . .
NIR3 . . . . . .
NIR4 . . . . . .
NIR5 70.531043 72.015581 73.368727 74.601191 75.725100 76.74866 NIR6 . . . . . .
NIR12 . . . . . .
NIR13 . . . . . .
NIR14 . . . . . .
NIR15 . . . . . .
NIR16 . . . . . .
NIR17 . . . . . .
NIR18 . . . . . .
NIR19 . . . . . .
NIR20 . . . . . .
NIR21 . . . . . .
NIR22 . . . . . .
NIR23 . . . . . .
NIR24 . . . . . .
NIR25 . . . . . .
NIR26 . . . . . .
NIR27 . . . . . .
NIR28 -2.921442 -4.596703 -6.142362 -7.565341 -8.866074 -10.05397 NIR29 -158.269644 -158.440388 -158.565340 -158.655379 -158.731275 -158.79575 NIR30 . . . . . .

NIR2 . . . . . .
NIR3 . . . . . .
NIR4 . . . . . .
NIR5 77.68023 78.53038 79.30481 80.01038 80.6532833 70.56035 NIR6 . . . . 0.0204128 10.24270 NIR12 . . . . . .
NIR13 . . . . . .
NIR14 . . . . . .
NIR15 . . . . . .
NIR16 . . . . . .
NIR17 . . . . . .
NIR18 . . . . . .
NIR19 . . . . . .
NIR20 . . . . . .
NIR21 . . . . . .
NIR22 . . . . . .
NIR23 . . . . . .
NIR24 . . . . . .
NIR25 . . . . . .
NIR26 . . . . . .
NIR27 . . . . . .
NIR28 -11.14471 -12.14122 -13.05091 -13.88142 -14.6511081 -18.37418 NIR29 -158.84040 -158.87761 -158.90864 -158.93425 -158.9547926 -156.73818 NIR30 . . . . . .

NIR2 . . . . . .

NIR3 . . . . . .
NIR4 . . . . . .
NIR5 58.85184 47.89464 37.80761 28.58535 20.17108 12.47659 NIR6 21.95293 32.90195 42.97805 52.18939 60.59337 68.27726 NIR12 . . . . . .
NIR13 . . . . . .
NIR14 . . . . . .
NIR15 . . . . . .
NIR16 . . . . . .
NIR17 . . . . . .
NIR18 . . . . . .
NIR19 . . . . . .
NIR20 . . . . . .
NIR21 . . . . . .
NIR22 . . . . . .
NIR23 . . . . . .
NIR24 . . . . . .
NIR25 . . . . . .
NIR26 . . . . . .
NIR27 . . . . . .
NIR28 -24.50111 -31.04741 -37.35758 -43.22911 -48.62544 -53.57896 NIR29 -150.93409 -144.16955 -137.47679 -131.19056 -125.39106 -120.05476 NIR30 . . . . . .

NIR2 . . . . . .
NIR3 . . . . . .
NIR4 . . . . . .
NIR5 5.49677236 . . . . .
NIR6 75.24894784 81.206627 81.55788 81.997161 82.402227 82.773686 NIR12 . . . . . .
NIR13 . . . . . .
NIR14 . . . . . .
NIR15 . . . . . .
NIR16 . . . . . .
NIR17 . . . . . .
NIR18 -0.01083383 -1.114076 -1.29050 -1.783698 -2.245068 -2.670771 NIR19 . . . . . .
NIR20 . . . . . .
NIR21 . . . . . .
NIR22 . . . . . .
NIR23 . . . . . .
NIR24 . . . . . .
NIR25 . . . . . .
NIR26 . . . . . .
NIR27 . . . . . .
NIR28 -58.04613412 -59.239421 -59.38639 -59.090167 -58.898262 -58.803828 NIR29 -115.20362777 -109.944696 -109.44715 -108.078007 -106.647542 -105.189505 NIR30 . . . . . .

NIR2 -2.057910 -4.238154 -6.219898 -8.029589 -9.674628 -11.172907 NIR3 . . . . . .
NIR4 . . . . . .
NIR5 . . . . . .
NIR6 83.999042 85.105944 86.109820 87.026117 87.859065 88.617692 NIR12 . . . . . .
NIR13 . . . . . .
NIR14 . . . . . .
NIR15 . . . . . .
NIR16 . . . . . .
NIR17 . . . . . .
NIR18 -3.599762 -4.133950 -4.612239 -5.048334 -5.444251 -5.804728 NIR19 . . . . . .
NIR20 . . . . . .

NIR21 . . . . . .
NIR22 . . . . . .
NIR23 . . . . . .
NIR24 . . . . . .
NIR25 . . . . . .
NIR26 . . . . . .
NIR27 . . . . . .
NIR28 -62.446155 -66.845376 -70.849577 -74.534730 -77.862557 -80.890168 NIR29 -95.502717 -86.486263 -78.317767 -70.815154 -64.033565 -57.863003 NIR30 . . . . . .

NIR2 -12.539999 -13.785112 -14.917201 -15.949461 -16.888772 -17.746589 NIR3 . . . . . .
NIR4 . . . . . .
NIR5 . . . . . .
NIR6 89.309702 89.939815 90.512714 91.035070 91.510308 91.944262 NIR12 . . . . . .
NIR13 . . . . . .
NIR14 . . . . . .
NIR15 . . . . . .
NIR16 . . . . . .
NIR17 . . . . . .
NIR18 -6.133354 -6.432135 -6.703404 -6.950788 -7.175475 -7.380793 NIR19 . . . . . .
NIR20 . . . . . .
NIR21 . . . . . .
NIR22 . . . . . .
NIR23 . . . . . .
NIR24 . . . . . .
NIR25 . . . . . .
NIR26 . . . . . .
NIR27 . . . . . .
NIR28 -83.666234 -86.193836 -88.478686 -90.566572 -92.460661 -94.202167 NIR29 -52.212069 -47.069192 -42.416792 -38.167106 -34.311616 -30.770853 NIR30 . . . . . .

NIR2 -18.52559217 -19.2443922 -19.9079914 -20.5168279 -21.078136 -21.592869 NIR3 . . . . . .
NIR4 . . . . . .
NIR5 . . . . . .
NIR6 92.33949368 92.7097147 93.0572965 93.3806451 93.682682 93.962531 NIR12 . . . . . .
NIR13 . . . . . .
NIR14 . . . . . .
NIR15 . . . . . .
NIR16 . . . . . .
NIR17 . . . . . .
NIR18 -7.56615357 -7.7374991 -7.8959371 -8.0414884 -8.176823 -8.301187 NIR19 . . . . . .
NIR20 . . . . . .
NIR21 . . . . . .
NIR22 . . . . . .
NIR23 . . . . . .
NIR24 . . . . . .
NIR25 . . . . . .
NIR26 . . . . . .
NIR27 -0.03041173 -0.1814426 -0.4469192 -0.7849073 -1.175805 -1.594745 NIR28 -95.71283100 -96.9067639 -97.7900178 -98.4218404 -98.868276 -99.166330 NIR29 -27.61048351 -24.7525255 -22.1967142 -19.9373827 -17.898761 -16.078501 NIR30 . . . . . .

NIR2 -22.069050 -22.500598 -22.896204748 -23.26949545 -2.360596e+01 NIR3 . . . . .
NIR4 . . . . .
NIR5 . . . . .
NIR6 94.224311 94.462780 94.683057646 94.90572794 9.510615e+01 NIR12 . . . . .

NIR13 . . . . .
NIR14 . . . . .
NIR15 . . . . .
NIR16 . . . . .
NIR17 . . . . .
NIR18 -8.417521 -8.521819 -8.618090131 -8.71184840 -8.787072e+00 NIR19 . . . . -1.148856e-04 NIR20 . . . . .
NIR21 . . . . .
NIR22 . . . . .
NIR23 . . . . .
NIR24 . . . . .
NIR25 . . . . .
NIR26 . . -0.002890285 -0.07992496 -1.894303e-01 NIR27 -2.039702 -2.473801 -2.900459262 -3.28638148 -3.598078e+00 NIR28 -99.355566 -99.453826 -99.486948183 -99.46891795 -9.941658e+01 NIR29 -14.404859 -12.946306 -11.629490931 -10.31627932 -9.217543e+00 NIR30 . . . . .

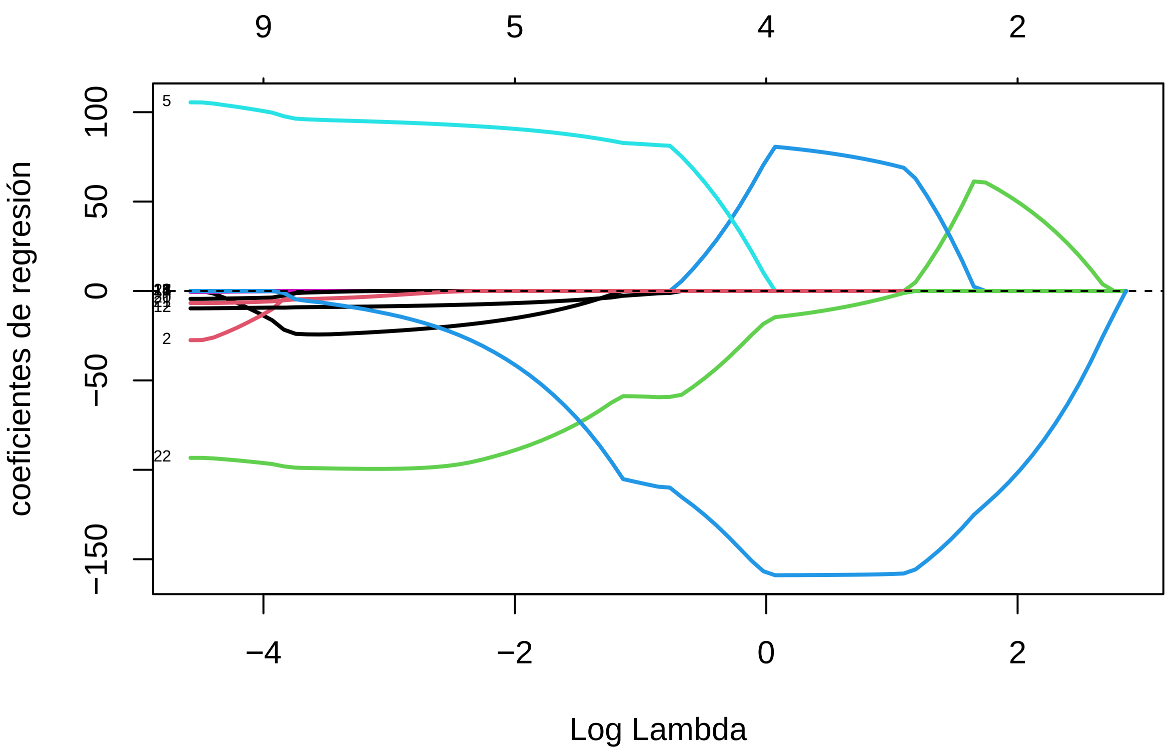
NIR2 -2.392113e+01 -2.421279e+01 -2.431158e+01 -2.426103e+01 -2.395712e+01 NIR3 . . -2.320394e-01 -5.644514e-01 -1.211884e+00 NIR4 . . . . .
NIR5 . . . . .
NIR6 9.529872e+01 9.548073e+01 9.574040e+01 9.598956e+01 9.636566e+01 NIR12 . . . . .
NIR13 . . . . .
NIR14 . . . . .
NIR15 . . . . .
NIR16 . . . . .
NIR17 . . . . .
NIR18 -8.857794e+00 -8.923013e+00 -8.990648e+00 -9.037487e+00 -9.110196e+00 NIR19 -2.513917e-04 -3.704223e-04 -4.811274e-04 -5.381095e-04 -6.215297e-04 NIR20 . . . . -2.409418e-04 NIR21 . -1.207928e-04 -1.106884e-03 -1.826650e-03 -2.333439e-03 NIR22 -1.143476e-03 -2.698802e-03 -3.222333e-03 -3.599785e-03 -4.025860e-03 NIR23 -1.905222e-04 -6.475062e-04 -1.085326e-03 -1.402280e-03 -1.757179e-03 NIR24 . . . . .
NIR25 . . . . .
NIR26 -3.314835e-01 -4.920745e-01 -6.978628e-01 -8.604441e-01 -1.142778e+00 NIR27 -3.866792e+00 -4.105796e+00 -4.328931e+00 -4.505309e+00 -4.695680e+00 NIR28 -9.933604e+01 -9.923442e+01 -9.910492e+01 -9.899506e+01 -9.881803e+01 NIR29 -8.188847e+00 -7.230282e+00 -6.270652e+00 -5.586568e+00 -4.628778e+00 NIR30 . . . . .

NIR2 -2.166155e+01 -16.41090910 -12.9064005 -9.7283402 -6.7176940 NIR3 -4.216273e+00 -10.12971709 -13.8758068 -17.2734601 -20.4907937 NIR4 . . . . .
NIR5 . . . . .
NIR6 9.770773e+01 99.72205425 100.8879087 101.9515364 102.9628077 NIR12 . . . . .
NIR13 . . . . .
NIR14 . . . . .
NIR15 . . . . .
NIR16 . . . . .
NIR17 . . . . .
NIR18 -9.316036e+00 -9.26679283 -9.3509642 -9.4365081 -9.5143902 NIR19 -5.047180e-04 . . . .
NIR20 -2.016432e-04 . . . .
NIR21 -2.385099e-03 . . . .
NIR22 -4.099652e-03 -0.10431203 -0.1768564 -0.2518206 -0.3302481 NIR23 -1.408331e-02 -0.09810884 -0.1457378 -0.2024125 -0.2700970 NIR24 . . . . .
NIR25 . . . . .
NIR26 -2.484058e+00 -3.66649572 -3.8395317 -3.9712904 -4.0926453 NIR27 -5.194754e+00 -5.76308641 -6.0528141 -6.2796102 -6.4676321 NIR28 -9.809083e+01 -96.78756675 -96.0300486 -95.3509070 -94.7110713 NIR29 -1.179934e+00 . . . .
NIR30 . . . . .

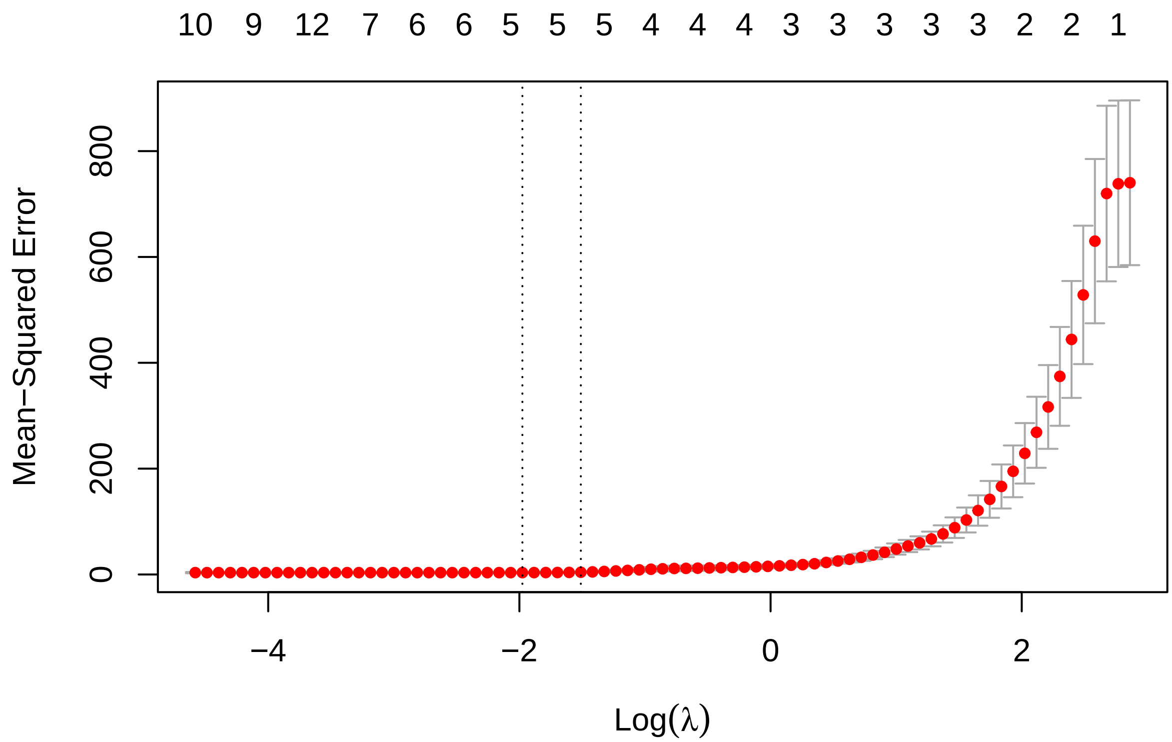
NIR2 -4.0582402 -1.4775166 . .

NIR3 -23.3393072 -25.9857596 -27.4644442 -2.750780e+01 NIR4 . -0.1725018 -0.3950726 -4.163678e-01 NIR5 . . . .
NIR6 103.8640898 104.8125286 105.4457504 1.054958e+02 NIR12 . . . .
NIR13 . . . .
NIR14 . . . .
NIR15 . . . .
NIR16 . . . .
NIR17 . . . -2.650788e-04 NIR18 -9.5805934 -9.6443366 -9.7010468 -9.706853e+00 NIR19 . . . .
NIR20 . . . .
NIR21 . . . .
NIR22 -0.4038774 -0.4765875 -0.5179921 -5.205979e-01 NIR23 -0.3413312 -0.4177564 -0.4645142 -4.677446e-01 NIR24 . . . .
NIR25 . . . .
NIR26 -4.1995439 -4.3079294 -4.4064034 -4.404672e+00 NIR27 -6.6145515 -6.7355174 -6.7208056 -6.727933e+00 NIR28 -94.1475808 -93.6219796 -93.3358281 -9.331713e+01 NIR29 . . . .
NIR30 . . . .

```
plot(lasso.mod,xvar='lambda',label=T,lwd=2,ylab='coeficientes de regresión')
abline(h=0,lty=2)
```




```
lasso.cv <-cv.glmnet(X., X$density, nfolds = 4, alpha = 1,nlambda = 100)
plot(lasso.cv)
```




```
est = glmnet(X., X$density, alpha = 1,lambda = lasso.cv$lambda.1se)
est$beta
```

24 x 1 sparse Matrix of class “dgCMatrix” s0 NIR2 -9.925178 NIR3 .
NIR4 .
NIR5 .
NIR6 87.822803 NIR12 .
NIR13 .
NIR14 .
NIR15 .
NIR16 .
NIR17 .
NIR18 -5.289614 NIR19 .
NIR20 .
NIR21 .
NIR22 .
NIR23 .
NIR24 .
NIR25 .
NIR26 .
NIR27 .
NIR28 -91.135757 NIR29 -43.247958 NIR30 .