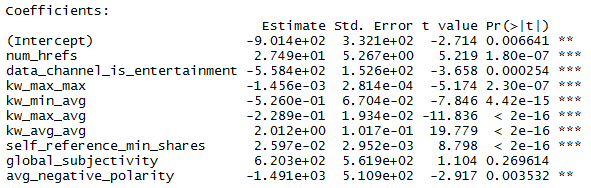
Andrés Felipe Paredes Jiménez

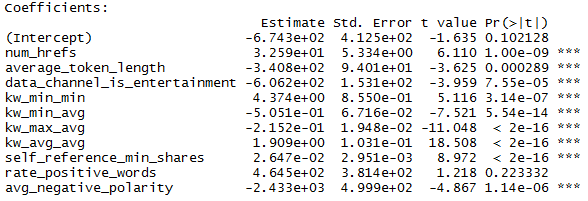
John Sebastián Marín Benítez

# Taller en clase: Selección Automática de modelos

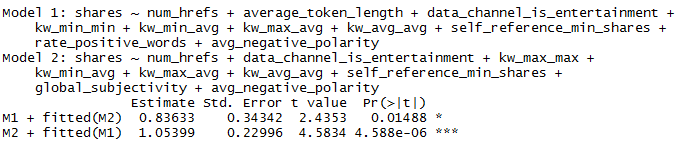
Creamos el modelo para explicar la variable dependiente shares con todas las demás variables de la base de datos.

Con base en este modelo, buscamos automáticamente el mejor modelo con base en el criterio BIC y usando el método Forward. El modelo obtenido es el siguiente:

Obtenemos el mejor modelo con base en el criterio BIC y usando el método Backward. El modelo es el siguiente:

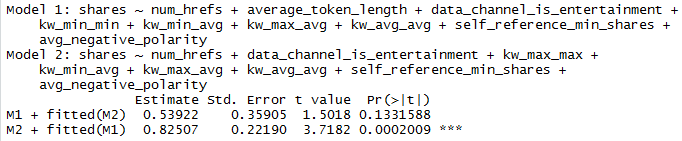


Por último realizamos una prueba J para modelos no anidados y obtenemos el siguiente resultado:



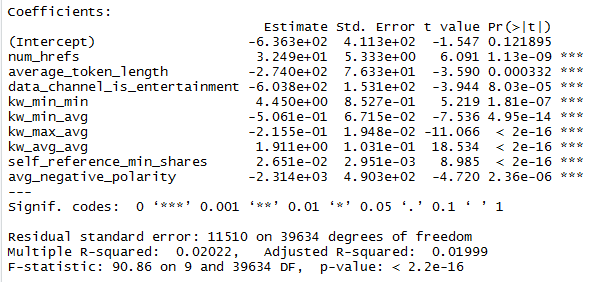
No podemos concluir que uno sea mejor que el otro.

Procedemos a eliminar variables no significativas de ambos modelos para ver si hay alguna diferencia. En este caso eliminamos la variable global\_subjectivity del modelo forward y la variable rate\_positive\_words del modelo backward y volvemos a comparar con la prueba J. Obtenemos los siguientes resultados:



Podemos rechazar con 99% de confianza que el modelo 2 es mejor que el modelo 1.

Con base en este análisis podemos concluir que el mejor modelo para este caso particular es el obtenido por el método Backward.

Num\_href, kw\_min\_min, kw\_avg\_avg y self\_reference\_min\_shares afectan positivamente a la variable shares.

Average\_token\_lenght, data\_channel\_is\_entertainment, kw\_min\_avg, kw\_max\_avg y avg\_negative\_polarity afectan negativamente a la variable shares.

Este modelo explica el 2.02% de la variable dependiente shares.

Código:

data<-read.csv("DatosTallerSelAuto.csv",sep=",")

data<-data[,-c(1:2)]

colnames(data)

model <- lm(shares ~ ., data = data)

summary(model)

library(olsrr)

library(leaps)

fwd.model2.sbc<- regsubsets(x = data[,1:58], y = data[,59],

nvmax=10000, method = "forward")

summary(fwd.model2.sbc)

plot(fwd.model2.sbc, scale ="bic", main ="Empleando el criterio de BIC forward")

bwd.model2.sbc<- regsubsets(x = data[,1:58], y = data[,59],

nvmax=10000, method = "backward")

plot(bwd.model2.sbc, scale ="bic", main ="Empleando el criterio de BIC backward")

bwdmodel <- lm(shares ~ num\_hrefs + average\_token\_length + data\_channel\_is\_entertainment + kw\_min\_min +

kw\_min\_avg + kw\_max\_avg + kw\_avg\_avg + self\_reference\_min\_shares +

avg\_negative\_polarity, data = data)

summary(bwdmodel)

fwdmodel <- lm(shares ~ num\_hrefs + data\_channel\_is\_entertainment + kw\_max\_max + kw\_min\_avg + kw\_max\_avg +

kw\_avg\_avg + self\_reference\_min\_shares +avg\_negative\_polarity, data = data)

summary(fwdmodel)

library(AER)

Ja <- jtest(bwdmodel, fwdmodel)

Ja