

# Tarea ML

Carlos Alberto Gómez Correa

1)

Importemos y veamos los primeros datos de la base COVID que contiene 347,502 observaciones y 11 variables. Se realizará una operacionalización de la variable categórica **TIPO.PACIENTE**, como variable de predicción, para poder ajustar un modelo de regresión logística. De esta manera, los datos quedan como sigue:

```
## 'data.frame': 347502 obs. of 12 variables:
## $ X           : int 0 1 2 3 4 5 6 7 8 9 ...
## $ SEXO        : chr "MUJER" "MUJER" "HOMBRE" "HOMBRE" ...
## $ OBESIDAD    : chr "NO" "NO" "NO" "NO" ...
## $ DIABETES    : chr "NO" "NO" "NO" "NO" ...
## $ EPOC         : chr "NO" "NO" "NO" "NO" ...
## $ ASMA         : chr "NO" "NO" "NO" "NO" ...
## $ CARDIOVASCULAR : chr "NO" "NO" "NO" "NO" ...
## $ RENAL.CRONICA : chr "NO" "NO" "NO" "NO" ...
## $ INMUNOSUPRESION: chr "NO" "NO" "NO" "NO" ...
## $ HIPERTENSION  : chr "NO" "NO" "NO" "NO" ...
## $ TIPO.PACIENTE : chr "AMBULATORIO" "AMBULATORIO" "AMBULATORIO" "AMBULATORIO" ...
## $ EDAD         : int 51 37 18 40 78 67 8 38 31 32 ...
```

	X	SEXO	OBESIDAD	DIABETES	EPOC	ASMA	CARDIOVASCULAR	RENAL.CRONICA	INMUNOSUPRESION	HIPERTENSION	TIPO.PACIENTE	EDAD
## 1	0	MUJER	NO	NO	NO	NO	NO	NO	NO	NO	AMBULATORIO	51
## 2	1	MUJER	NO	NO	NO	NO	NO	NO	NO	NO	AMBULATORIO	37
## 3	2	HOMBRE	NO	NO	NO	NO	NO	NO	NO	NO	AMBULATORIO	18
## 4	3	HOMBRE	NO	NO	NO	NO	NO	NO	NO	NO	AMBULATORIO	40
## 5	4	MUJER	NO	SI	NO	NO	NO	NO	NO	SI	AMBULATORIO	78
## 6	5	HOMBRE	NO	NO	NO	NO	NO	NO	NO	NO	AMBULATORIO	67

Efectuemos un ajuste de un modelo de regresión logística a los datos con todas las covariables. Para esto utilizaremos la función *glm* con la que se ajustan modelos lineales generalizados, colocando el parámetro *family* en binomial, es decir, especificando que los residuales tienen una distribución binomial.

```

library(MASS)

## Warning: package 'MASS' was built under R version 4.0.3

##
## Attaching package: 'MASS'

## The following object is masked from 'package:dplyr':
##   select

library(stats)
#Ajustando un modelo de regresión logistica con todas las covariables
ModeloCovidCompleto<-glm(factor(TIPO.PACIENTE)~SEXO + OBESIDAD + DIABETES + EPOC +
  ASMA + CARDIOVASCULAR + RENAL.CRONICA + INMUNOSUPRESION +
  HIPERTENSION + EDAD, data=covid,family = binomial)

summary(ModeloCovidCompleto)

##
## Call:
## glm(formula = factor(TIPO.PACIENTE) ~ SEXO + OBESIDAD + DIABETES +
##       EPOC + ASMA + CARDIOVASCULAR + RENAL.CRONICA + INMUNOSUPRESION +
##       HIPERTENSION + EDAD, family = binomial, data = covid)
##
## Deviance Residuals:
##    Min      1Q  Median      3Q     Max 
## -2.8856 -0.7274 -0.5036  0.8197  2.8408 
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)    
## (Intercept) -3.5171080  0.0158060 -222.518 < 2e-16 ***
## SEXOMUJER   -0.5001602  0.0086104  -58.088 < 2e-16 ***
## OBESIDADSII 0.2772065  0.0103591   26.760 < 2e-16 ***
## DIABETESSII  0.6714155  0.0109999   61.038 < 2e-16 ***
## EPOCSI       0.3041699  0.0306488    9.924 < 2e-16 ***
## ASMASI       -0.1643232  0.0275873   -5.956 2.58e-09 ***
## CARDIOVASCULARSI 0.0909165  0.0269128    3.378 0.00073 ***
## RENAL.CRONICASI  0.9690105  0.0288477   33.591 < 2e-16 ***
## INMUNOSUPRESIONSI 0.8287691  0.0354537   23.376 < 2e-16 ***
## HIPERTENSIONSII 0.2367040  0.0107656   21.987 < 2e-16 ***
## EDAD         0.0522757  0.0003071  170.246 < 2e-16 ***
## ---        
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
## Null deviance: 414267  on 347501  degrees of freedom
## Residual deviance: 343452  on 347491  degrees of freedom
## AIC: 343474
##
## Number of Fisher Scoring iterations: 4

```

Al parecer, todas las covariables son significativas a un nivel de confianza del 95%. Efectuemos una selección del modelo mediante los modos de búsqueda *backward*, *forward* y *stepwise* con el criterio de Akaike como métrica.

### Backward

```
backmode<-stepAIC(ModeloCovidCompleto, direction = "backward", trace = F)
```

```
backmode$anova
```

```
## Stepwise Model Path
## Analysis of Deviance Table
##
## Initial Model:
## factor(TIPO.PACIENTE) ~ SEXO + OBESIDAD + DIABETES + EPOC + ASMA +
##     CARDIOVASCULAR + RENAL.CRONICA + INMUNOSUPRESION + HIPERTENSION +
##     EDAD
##
## Final Model:
## factor(TIPO.PACIENTE) ~ SEXO + OBESIDAD + DIABETES + EPOC + ASMA +
##     CARDIOVASCULAR + RENAL.CRONICA + INMUNOSUPRESION + HIPERTENSION +
##     EDAD
##
## Step Df Deviance Resid. Df Resid. Dev      AIC
## 1           347491    343451.7 343473.7
```

```
AIC(ModeloCovidCompleto)
```

```
## [1] 343473.7
```

```
AIC(backmode)
```

```
## [1] 343473.7
```

## Forward

```
## Stepwise Model Path
## Analysis of Deviance Table
##
## Initial Model:
## factor(TIPO.PACIENTE) ~ 1
##
## Final Model:
## factor(TIPO.PACIENTE) ~ EDAD + DIABETES + SEXO + RENAL.CRONICA +
## OBESIDAD + INMUNOSUPRESION + HIPERTENSION + EPOC + ASMA +
## CARDIOVASCULAR
##
##
##          Step Df     Deviance Resid. Df Resid. Dev      AIC
## 1                   347501  414266.6 414268.6
## 2   + EDAD    1  57928.42380  347500  356338.2 356342.2
## 3   + DIABETES  1  5931.22959  347499  350406.9 350412.9
## 4   + SEXO    1  3228.56895  347498  347178.4 347186.4
## 5   + RENAL.CRONICA  1  1596.77015  347497  345581.6 345591.6
## 6   + OBESIDAD  1  911.51896  347496  344670.1 344682.1
## 7 + INMUNOSUPRESION  1  570.23533  347495  344099.9 344113.9
## 8   + HIPERTENSION  1  502.11161  347494  343597.7 343613.7
## 9   + EPOC     1   99.14419  347493  343498.6 343516.6
## 10  + ASMA     1   35.48408  347492  343463.1 343483.1
## 11 + CARDIOVASCULAR  1   11.42505  347491  343451.7 343473.7
##
## [1] 343473.7
```

## Stepwise

```
library(RcmdrMisc)

## Warning: package 'RcmdrMisc' was built under R version 4.0.3

## Loading required package: car

## Warning: package 'car' was built under R version 4.0.3

## Loading required package: carData

## 
## Attaching package: 'car'

## The following object is masked from 'package:dplyr':
## 
##     recode

## Loading required package: sandwich

library(Rcmdr)

## Warning: package 'Rcmdr' was built under R version 4.0.3

## Loading required package: splines

## Loading required package: effects

## Registered S3 methods overwritten by 'lme4':
##   method           from
##   cooks.distance.influence.merMod car
##   influence.merMod            car
##   dfbeta.influence.merMod      car
##   dfbetas.influence.merMod     car

## lattice theme set by effectsTheme()
## See ?effectsTheme for details.

## La interfaz R-Commander sólo funciona en sesiones interactivas

## 
## Attaching package: 'Rcmdr'

## The following object is masked from 'package:base':
## 
##     errorCondition
```

```

stepwiseModel<-stepwise(ModeloCovidCompleto,direction = "forward/backward",criterion = "AIC", trace=F)

##
## Direction: forward/backward
## Criterion: AIC

stepwiseModel$anova

## Stepwise Model Path
## Analysis of Deviance Table
##
## Initial Model:
## factor(TIPO.PACIENTE) ~ 1
##
## Final Model:
## factor(TIPO.PACIENTE) ~ EDAD + DIABETES + SEXO + RENAL.CRONICA +
##     OBESIDAD + INMUNOSUPRESION + HIPERTENSION + EPOC + ASMA +
##     CARDIOVASCULAR
##
##
##          Step Df    Deviance Resid. Df Resid. Dev      AIC
## 1              347501  414266.6 414268.6
## 2 + EDAD     1 57928.42380  347500  356338.2 356342.2
## 3 + DIABETES 1 5931.22959  347499  350406.9 350412.9
## 4 + SEXO     1 3228.56895  347498  347178.4 347186.4
## 5 + RENAL.CRONICA 1 1596.77015  347497  345581.6 345591.6
## 6 + OBESIDAD 1 911.51896  347496  344670.1 344682.1
## 7 + INMUNOSUPRESION 1 570.23533  347495  344099.9 344113.9
## 8 + HIPERTENSION 1 502.11161  347494  343597.7 343613.7
## 9 + EPOC      1 99.14419   347493  343498.6 343516.6
## 10 + ASMA     1 35.48408   347492  343463.1 343483.1
## 11 + CARDIOVASCULAR 1 11.42505  347491  343451.7 343473.7

AIC(stepwiseModel)

## [1] 343473.7

```

Nótese que obtuvimos el mismo modelo en todos los casos.

## Criterio de BIC

Efectuemos el método *stepwise* pero ahora con el Criterio de Información Bayesiano.

```
stepwiseModelBIC<-stepwise(ModeloCovidCompleto,direction = "forward/backward",criterion = "BIC", trace=1)

##
## Direction: forward/backward
## Criterion: BIC

BIC(ModeloCovidCompleto)

## [1] 343592

stepwiseModelBIC$anova

## Stepwise Model Path
## Analysis of Deviance Table
##
## Initial Model:
## factor(TIPO.PACIENTE) ~ 1
##
## Final Model:
## factor(TIPO.PACIENTE) ~ EDAD + DIABETES + SEXO + RENAL.CRONICA +
##      OBESIDAD + INMUNOSUPRESION + HIPERTENSION + EPOC + ASMA
##
##
##          Step Df     Deviance Resid. Df Resid. Dev      AIC
## 1                   347501   414266.6 414279.4
## 2       + EDAD  1  57928.42380  347500   356338.2 356363.7
## 3       + DIABETES  1  5931.22959  347499   350406.9 350445.2
## 4       + SEXO  1  3228.56895  347498   347178.4 347229.4
## 5   + RENAL.CRONICA  1  1596.77015  347497   345581.6 345645.4
## 6       + OBESIDAD  1   911.51896  347496   344670.1 344746.6
## 7   + INMUNOSUPRESION  1   570.23533  347495   344099.9 344189.2
## 8   + HIPERTENSION  1   502.11161  347494   343597.7 343699.8
## 9       + EPOC  1    99.14419  347493   343498.6 343613.4
## 10      + ASMA  1    35.48408  347492   343463.1 343590.7
```

```
BIC(stepwiseModelBIC)
```

```
## [1] 343590.7
```

Mediante este (BIC/forward-backward), se obtuvo un modelo con una covariable menos que el modelo completo. La variable que se desechó fue *CARDIOVASCULAR*.

Ahora, usemos (BIC/backward-forward):

```
stepwiseModelBIC<-stepwise(ModeloCovidCompleto,direction = "backward/forward",criterion = "BIC", trace=1)
```

```

## 
## Direction: backward/forward
## Criterion: BIC

BIC(ModeloCovidCompleto)

## [1] 343592

stepwiseModelBIC$anova

## Stepwise Model Path
## Analysis of Deviance Table
##
## Initial Model:
## factor(TIPO.PACIENTE) ~ SEXO + OBESIDAD + DIABETES + EPOC + ASMA +
##      CARDIOVASCULAR + RENAL.CRONICA + INMUNOSUPRESION + HIPERTENSION +
##      EDAD
##
## Final Model:
## factor(TIPO.PACIENTE) ~ SEXO + OBESIDAD + DIABETES + EPOC + ASMA +
##      RENAL.CRONICA + INMUNOSUPRESION + HIPERTENSION + EDAD
##
##
##          Step Df Deviance Resid. Df Resid. Dev      AIC
## 1                  347491   343451.7 343592.0
## 2 - CARDIOVASCULAR  1 11.42505   347492   343463.1 343590.7

```

```
BIC(stepwiseModelBIC)
```

```
## [1] 343590.7
```

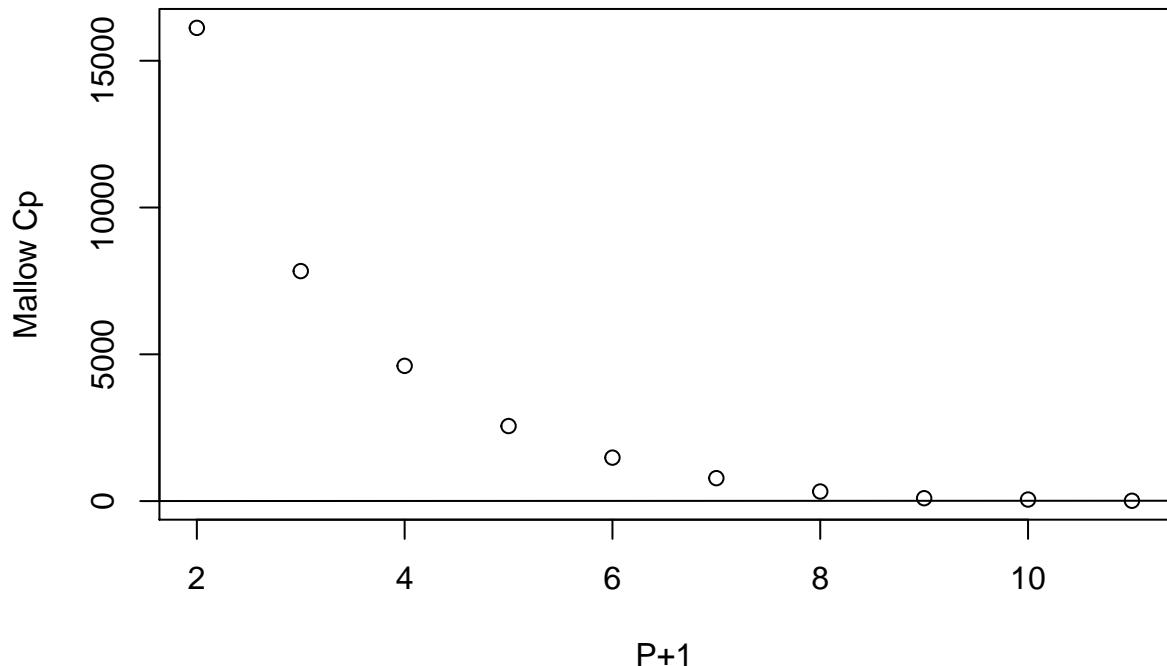
Notemos que con ambos enfoques, se obtiene un modelo con 9 covariables retirando la covariable *CARDIO-VASCULAR* y con un BIC de 343590.7

Realicemos todos los modelos posibles para obtener el criterio de Mallow y poder compararlo entre ellos.

```
library(leaps)
Todos<-regsubsets(factor(TIPO.PACIENTE)~SEXO + OBESIDAD + DIABETES + EPOC +
ASMA + CARDIOVASCULAR + RENAL.CRONICA + INMUNOSUPRESION +
HIPERTENSION + EDAD, data= covid, nbest=1,nvmax = 11)

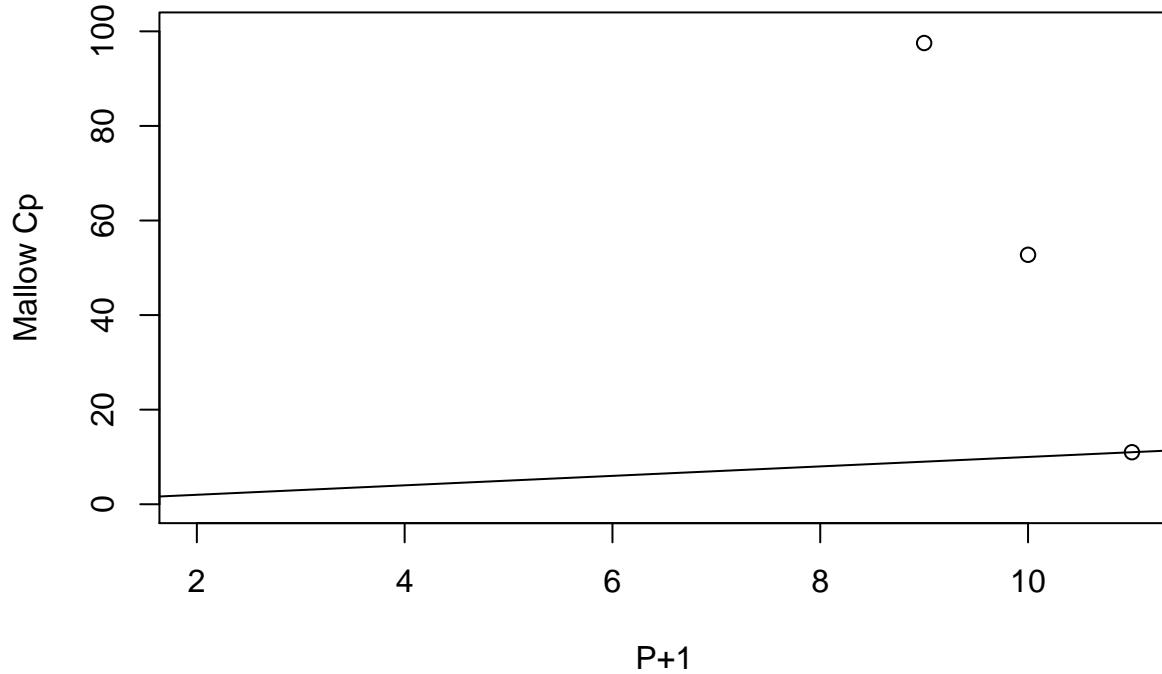
resumenMallow<-summary(Todos)
```

Confeccionemos la gráfica de Mallow  $C_p$  contra el número de variables predictoras  $p$ . Note que se busca el modelo que cumpla  $C_p \approx p$ , es decir, buscamos los modelos que estén sobre o por debajo de la recta  $x = y$ .



Los mejores modelos son los que tienen más de 4 variables. Identifiquemos cuáles son estos modelos:

```
plot(2:11,resumenMallow$cp, xlab="P+1", ylab="Mallow Cp",ylim=c(0,100))
abline(a=0,b=1)
```



Por la gráfica anterior, se determina que el mejor modelo es aquél con 10 variables.

Supongamos que debemos eliminar dos variables. Determinemos cuales serían mediante `varImp()`, posteriormente, ordenemos las variables por su nivel de importancia.

```
##                               Overall
## EDAD                  170.246337
## DIABETESSI            61.038446
## SEXOMUJER             58.088238
## RENAL.CRONICASI       33.590590
## OBESIDADSI            26.759752
## INMUNOSUPRESIONSI    23.376111
## HIPERTENSIONSI        21.987160
## EPOCSI                 9.924355
## ASMASI                 5.956470
## CARDIOVASCULARSI      3.378184
```

Por lo tanto, las dos variables que se eliminarían serían *ASMA* y *CARVIOVASCULAR*, pues son las que obtienen valores menores respecto a la estadística *t* de Wald que permite establecer qué variables son de relevancia para explicar la probabilidad de que ocurra el suceso  $Y = 1$ . Es decir, se contrastan las hipótesis  $H_0 : \beta_i = 0$  v.s.  $H_a : \beta_i \neq 0$ . Note que estos valores coinciden, por lo tanto, si se hubiese obtenido una jerarquización de las variables predictoras mediante los *p-value's* resultantes de la función `summary` del modelo ajustado con todas las covariables.

Veamos que tan buenas son las predicciones con el modelo de 10 variables y el de 8 variables.

Efecutemos un split del dataset utilizando 80% para los datos de entrenamiento y 20% para los datos de prueba.

```
set.seed(22)
n<-nrow(covid)

# Permutación de los renglones
RenglonesPermutados<-sample(n)

#Ordenando aleatoriamente los datos
covid_shuffled<-covid[RenglonesPermutados,]

# Punto de split
split <- round(n*.8)

#Training
covidTrain<-covid_shuffled[1:split,]

#Test, 0,1
covidTest<-covid_shuffled[(split+1):n]
```

Efecutemos el ajuste de regresión logística para el set de entrenamiento y hagamos predicciones para el set de prueba:

```
# Modelo con 10 covariables
ModeloCovidML10Cov<-glm(factor(TIPO.PACIENTE)~SEXO + OBESIDAD + DIABETES + EPOC +
ASMA + CARDIOVASCULAR + RENAL.CRONICA + INMUNOSUPRESION +
HIPERTENSION + EDAD, data=covidTrain, family = binomial)

ModeloCovidML8Cov<-glm(factor(TIPO.PACIENTE)~SEXO + OBESIDAD + DIABETES + EPOC + RENAL.CRONICA + INMUNOSUPRESION +
HIPERTENSION + EDAD, data=covidTrain, family = binomial)

p1<-predict(ModeloCovidML10Cov,covidTest,type="response")
p2<-predict(ModeloCovidML8Cov,covidTest,type="response")
```

Una vez ajustados estos modelos, evaluemos algunas métricas a través de la matriz de confusión.

De esta manera, para el modelo con 8 variables tenemos las siguientes métricas de la matriz de confusión:

```
## Confusion Matrix and Statistics
##
##             Reference
## Prediction      0      1
##       0 45590 12040
##       1  4280  7590
##
##               Accuracy : 0.7652
```

```

##                               95% CI : (0.762, 0.7683)
##      No Information Rate : 0.7176
##      P-Value [Acc > NIR] : < 2.2e-16
##
##                           Kappa : 0.3418
##
##  Mcnemar's Test P-Value : < 2.2e-16
##
##                           Sensitivity : 0.3867
##                           Specificity : 0.9142
##      Pos Pred Value : 0.6394
##      Neg Pred Value : 0.7911
##                           Prevalence : 0.2824
##                           Detection Rate : 0.1092
##      Detection Prevalence : 0.1708
##                           Balanced Accuracy : 0.6504
##
##      'Positive' Class : 1
##

```

De manera análoga, estos son los resultados para el modelo con 10 covariables:

ConfMatModelo10

```

## Confusion Matrix and Statistics
##
##      Reference
## Prediction    0     1
##      0 45603 12055
##      1  4267  7575
##
##                           Accuracy : 0.7652
##                           95% CI : (0.762, 0.7683)
##      No Information Rate : 0.7176
##      P-Value [Acc > NIR] : < 2.2e-16
##
##                           Kappa : 0.3414
##
##  Mcnemar's Test P-Value : < 2.2e-16
##
##                           Sensitivity : 0.3859
##                           Specificity : 0.9144
##      Pos Pred Value : 0.6397
##      Neg Pred Value : 0.7909
##                           Prevalence : 0.2824
##                           Detection Rate : 0.1090
##      Detection Prevalence : 0.1704
##                           Balanced Accuracy : 0.6502
##
##      'Positive' Class : 1
##

```

En ambos modelos se tienen resultados similares en la precisión (aproximadamente 76%), sensibilidad (aproximadamente 38%), y especificidad (90%), cada uno. Esto nos indica que si bien la precisión esta en un

nivel decente, muchos individuos se están clasificando como pacientes ambulatorios que en los datos reales sí están hospitalizados. Lo anterior lo podemos medir, utilizando los datos del primero modelo, mediante el siguiente cálculo:

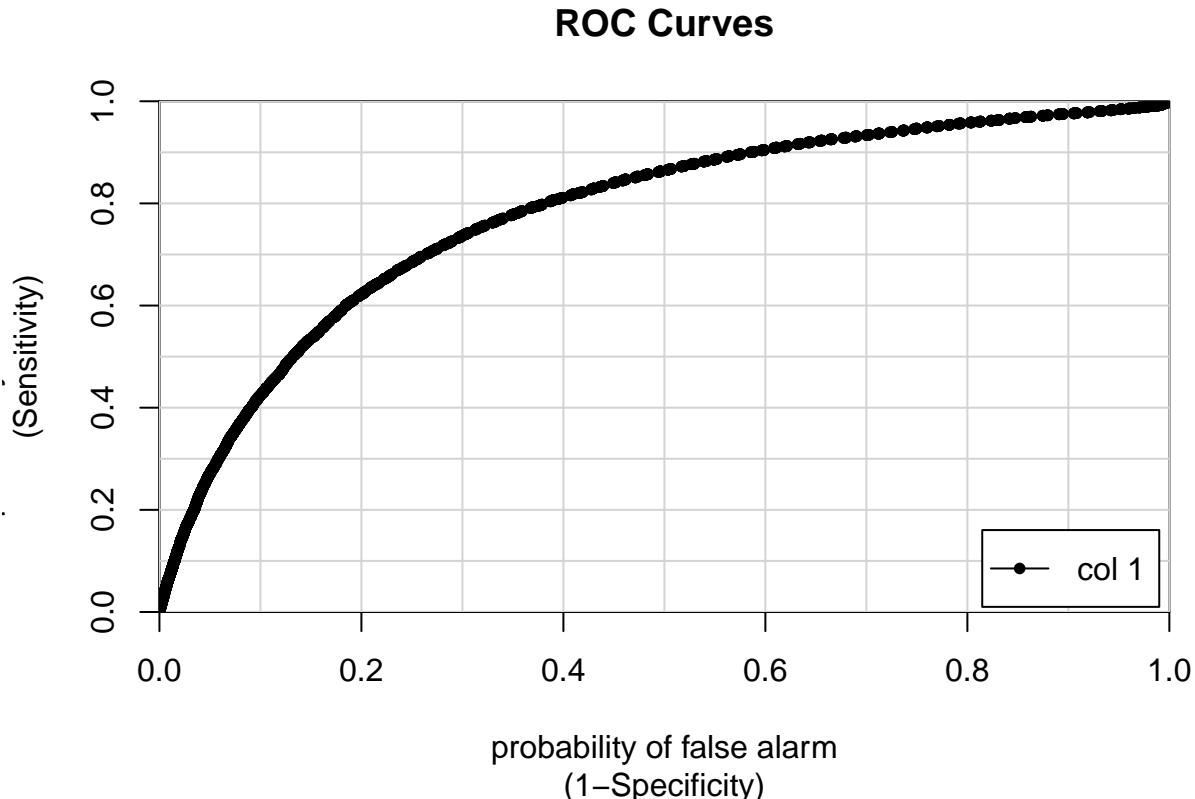
$$\frac{FN}{FN + TP} = 0.6133469 \quad \dots(1)$$

Efectuemos la evaluación de diferentes thresholds mediante la gráfica de la curva ROC, para ambos modelos.

```
library(caTools)

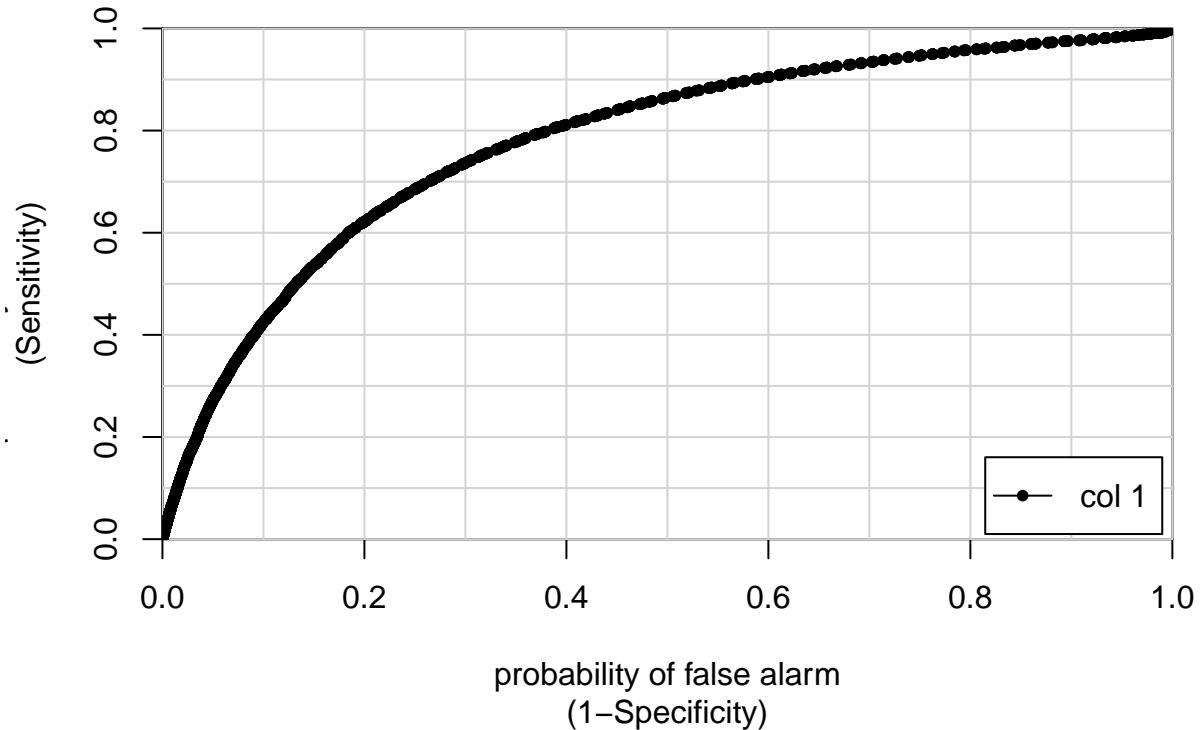
## Warning: package 'caTools' was built under R version 4.0.3

library(caret)
ROC1<-colAUC(p1, covidTest[["TIPO.PACIENTE"]],plotROC = T)
```



```
ROC2<- colAUC(p2, covidTest[["TIPO.PACIENTE"]],plotROC = T)
```

## ROC Curves



ROC1

```
## [1]
## 0 vs. 1 0.778586
```

ROC2

```
## [1]
## 0 vs. 1 0.7786809
```

Note que los valore del AUC son un poco mayores al 77%, lo que denota que el modelo ejecuta de manera decente la clasificación.

Ajustemos los parámetros del modelo respecto al AUC en lugar de la precisión para mejorar el resultado, mediante Cross-Validation.

```
# Tunning control
ControlTunning<-trainControl(method = "cv", number=5,
                                summaryFunction = twoClassSummary,
                                classProbs = T,verboseIter = T,)

# Tunneado Modelo
#levels(covid_modelo_tunneado$TIPO.PACIENTE)=c("1", "0")
```

```

# Levels categóricos
levels(covid$TIPO.PACIENTE)<-c("AMBULATORIO","HOSPITALIZADO")

ModeloCovidTuneado<-train(TIPO.PACIENTE~SEXO + OBESIDAD + DIABETES + EPOC +
ASMA + CARDIOVASCULAR + RENAL.CRONICA + INMUNOSUPRESION +
HIPERTENSION + EDAD, data=covid, method="glm", trControl=ControlTunning, metric="ROC")

## + Fold1: parameter=none
## - Fold1: parameter=none
## + Fold2: parameter=none
## - Fold2: parameter=none
## + Fold3: parameter=none
## - Fold3: parameter=none
## + Fold4: parameter=none
## - Fold4: parameter=none
## + Fold5: parameter=none
## - Fold5: parameter=none
## Aggregating results
## Fitting final model on full training set

ModeloCovidTuneado

## Generalized Linear Model
##
## 347502 samples
##      10 predictor
##      2 classes: 'AMBULATORIO', 'HOSPITALIZADO'
##
## No pre-processing
## Resampling: Cross-Validated (5 fold)
## Summary of sample sizes: 278002, 278001, 278001, 278002, 278002
## Resampling results:
##
##    ROC      Sens      Spec
##    0.7805027  0.9169996  0.382134

ModeloCovidTuneado$finalModel

##
## Call:  NULL
##
## Coefficients:
## (Intercept)      SEXOMUJER      OBESIDADS1      DIABETESS1
##           -3.51711       -0.50016       0.27721       0.67142
##          EPOCSI        ASMASI  CARDIOVASCULARSI  RENAL.CRONICASI
##            0.30417       -0.16432       0.09092       0.96901
##  INMUNOSUPRESIONSI  HIPERTENSIONSI      EDAD
##            0.82877        0.23670       0.05228
##
## Degrees of Freedom: 347501 Total (i.e. Null);  347491 Residual
## Null Deviance:      414300
## Residual Deviance: 343500   AIC: 343500

```

```
ModeloCovidTuneado$metric
```

```
## [1] "ROC"
```

Notemos como a través de la validación cruzada conseguimos ha mejorado ligeramente la sensibilidad, la especificidad y el área bajo la curva ROC. Ahora, efectuemos con el modelo lineal generalizado obtenido, predicciones sobre el set de prueba y verifiquemos su eficacia.

```
calc_acc = function(actual, predicted) {
  mean(actual == predicted)
}

levels(covidTest$TIPO.PACIENTE)<-c("AMBULATORIO", "HOSPITALIZADO")

prediccion_tuneado<-predict(ModeloCovidTuneado,newdata = covidTest, type="raw")

calc_acc(actual = covidTest$TIPO.PACIENTE,
         predicted = prediccion_tuneado)

## [1] 0.7650504

confusionMatrix(covidTest$TIPO.PACIENTE,prediccion_tuneado, positive = "HOSPITALIZADO")

## Confusion Matrix and Statistics
##
##                 Reference
## Prediction      AMBULATORIO HOSPITALIZADO
##   AMBULATORIO      45643        4227
##   HOSPITALIZADO      12102        7528
##
##                 Accuracy : 0.7651
##                   95% CI : (0.7619, 0.7682)
##   No Information Rate : 0.8309
##   P-Value [Acc > NIR] : 1
##
##                 Kappa : 0.3401
##
##   Mcnemar's Test P-Value : <2e-16
##
##                 Sensitivity : 0.6404
##                 Specificity : 0.7904
##   Pos Pred Value : 0.3835
##   Neg Pred Value : 0.9152
##                 Prevalence : 0.1691
##   Detection Rate : 0.1083
##   Detection Prevalence : 0.2824
##   Balanced Accuracy : 0.7154
##
```

```
##      'Positive' Class : HOSPITALIZADO  
##
```

Realicemos el siguiente cálculo:

$$\frac{FN}{FN + TP} = 0.35959 \quad \dots(2)$$

Con estos resultados tenemos que, para el set de prueba, la proporción de predicciones correctas (precisión) se mantuvo sobre el 76%, sin embargo, hubo un mejoramiento en el balance respecto a la sensibilidad y la especificidad, pues la primera aumentó a 64% y la segunda disminuyó al 79%. De esta manera, por (1) y (2) concluimos que con este entrenamiento del modelo el porcentaje de pacientes hospitalizados que la regresión clasificó como ambulatorios pasó de 61% a 35%.