

UNIVERSIDAD TÉCNICA DEL NORTE

FICA-CIERCOM

SISTEMAS EMBEBIDOS

CALIDAD DEL AGUA

Benavides Wilmer, Farinango Luis, Velasco Angel

13 de enero de 2020

1. Introducción

2. Marco Teórico

2.1. Matriz de Confusión

2.2. Curva ROC

Una curva ROC proporciona una representación de la sensibilidad y especificidad para cada valor umbral ,que es invariante mediante transformaciones monótonas a los datos de la variable de decisión y que permite comparar dos o más clasificadores en función de su función discriminante. Un análisis ROC de sus siglas en ingles Receiver operating characteristics, es una metodología desarrollada para analizar un sistema de decisión. Este análisis se basa en las nociones de sensibilidad y especificidad.

Una curva ROC proporciona una representación de la sensibilidad generalmente ubicada en el eje de las “y” ,es decir se muestra los verdaderos positivos.

$$\frac{(VP)}{(VP) + (FN)} \quad (1)$$

Donde VP es el número de verdaderos positivos,FN es el número de falsos negativos.

La especificidad se ubica en el eje de las “x” y representa los falsos positivos .

$$\frac{(FP)}{(FP) + (VN)} \quad (2)$$

Donde FP es el número de falsos positivos,VN es el número de verdaderos negativos.

La curva basa sus dos valores sen función de cada valor umbral ,que es invariante mediante transformaciones monótonas a los datos de la variable de decisión y que permite comparar dos o más clasificadores en función de su función discriminante. Además, en la curva ROC cada uno de los puntos representa un par de sensibilidad/especificidad correspondiente a un umbral de decisión particular.

2.3. Medición AUC

3. Desarrollo

3.1. Matriz de Confusión

3.2. Curva ROC

La curva ROC en el entorno de R necesita de un análisis previo que debe dar como resultado u n conjunto de datos binarios ,dado que la función solo acepta este tipo de datos. Para el ejemplo se ha utilizado un análisis de RandomForest. Se ha estudiado la varaiaible temperatura,para lograr la obtención de la curva ROC se ha dado un umbral de un temperatura de 15 grados celsius.Si se sobrepasa dicha temperatura entonces la base de datos cambiara a al número 1. Partiendo de este concepto se aplica el algoritmo de clasificación antes mencionado.

```

1 library(knitr)
2 library(ROCR)
3 library(e1071)
4 library(caTools)
5 library(nnet)
6 library(scales)
7 set.seed(222)
8 #Cargara la base de datos al entorno de R
9 datos=read.csv("datos.csv",header=TRUE,sep=",")
10 datos$Temperatura=ifelse(datos$Temperatura>15,1,0)#
11 datos$Temperatura=as.factor(datos$Temperatura)
12 str(datos)
13 indice=sample(1:nrow(datos),size = 0.8*nrow(datos))
14 training=datos[indice,]
15 test=datos[-indice,]
16 library(randomForest)
17 #MODELO random forest
18 rf_model=randomForest(Temperatura~.,data =training)
19 rf_prediction=predict(rf_model,test,type='vote')
20 #modelo de regresion lineal
21 #lr_model=glm(PH~.,data=training,family = "binomial")
22 #lr_pred=prediction(lr_model,tes,type="response")
23 #curvas roc
24 library(pROC)
25 ROC_rf = roc(test$Temperatura,rf_prediction[,2])
26 #ROC_lr <- roc(test$PH, lr_prediction)
27 plot(ROC_rf)

```

Figura 1: Código en R para el desarrollo de la Curva ROC

3.3. Medición AUC

4. Análisis de Resultados

Para el caso de la curva ROC se obtuvo la siguiente gráfica,cabe recalcar que se analizo la temperetura inicialmente.

En este caso se observa que el algoritmo de clasificacion que se utilizo como ejemplo no es muy efectivo ya que para que un algoritmo sea efectivo la curva se debe acercar lo mayor posible a la izquierda

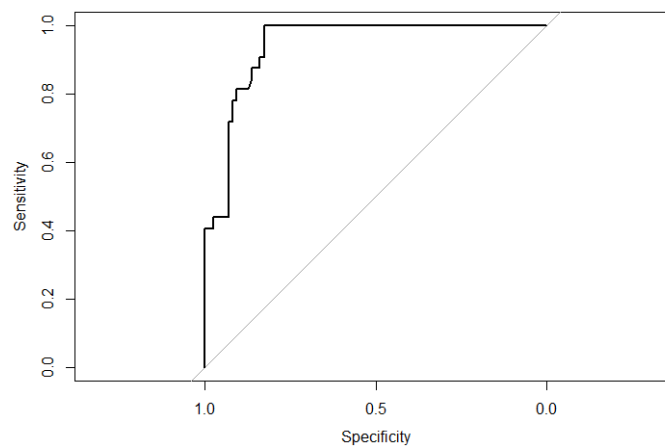


Figura 2: Curva ROC obtenida de la variable temperatura

5. Conclusiones y Recomendaciones

Conclusiones

- La curva ROC es un método que puede determinar la efectividad de un algoritmo de clasificación ,entre mas cerca de la izquierda esta ubicada dicha curva más efectivo es el algoritmo de clasificación.

Recomendaciones

*

6. Anexos