



UNIVERSITAT
POLITÈCNICA
DE VALÈNCIA



Máster en Inteligencia Artificial, Reconocimiento de Formas e Imagen Digital
Universitat Politècnica de València

Medidas de calidad

Biometría

Autor: Juan Antonio López Ramírez

Curso 2019-2020

Introducción

El objetivo de esta práctica es realizar un script para obtener los siguientes resultados:

- Curva ROC
- $FP(FN=X)$ y umbral
- $FN(FP=X)$ y umbral
- $FP=FN$ y umbral
- Área bajo la curva ROC
- D-Prime

Partimos de los cuatro ficheros proporcionados en poliformaT que recopilan los scores de clientes e impostores para dos sistemas A y B.

Implementación y resultados

Ejecución del script

Para poder hacer uso del programa que hemos implementado, hay que ejecutar en la terminal la siguiente instrucción:

```
1 python3 medidas_calidad.py c i x1 x2 ... xn
```

Donde c e i son los archivos de scores de clientes e impostores, respectivamente, del sistema determinado (A o B); y $x1, x2 \dots xn$ son los distintos valores de x para los que se desea calcular $FP(FN=x)$, el umbral $FN=x$, $FN(FP=x)$ y el umbral $FP=x$.

El resultado de ejecutar este script se puede apreciar en la figura 1.

```
jalopec@JALOPEZ:/mnt/d/Cosas de Juan/Master/Semestre B/Biometria/Practicas/Medidas de calidad$ python3 medidas_calidad.py
scoresA_clientes scoresA_impostores 0.25 0.5 0.75
Para x=0.250000:
  El umbral FN=x es 0.059195 y FP(FN=x) es 0.138462
  El umbral FP=x es 0.044937 y FN(FP=x) es 0.163636
Para x=0.500000:
  El umbral FN=x es 0.105233 y FP(FN=x) es 0.023718
  El umbral FP=x es 0.024678 y FN(FP=x) es 0.067832
Para x=0.750000:
  El umbral FN=x es 0.289883 y FP(FN=x) es 0.000000
  El umbral FP=x es 0.009815 y FN(FP=x) es 0.016084
El umbral FN=FP es 0.050333
La curva ROC se ha generado en una imagen .png y el área bajo esta es de 0.891341
El factor dprime obtenido es 0.759954
jalopec@JALOPEZ:/mnt/d/Cosas de Juan/Master/Semestre B/Biometria/Practicas/Medidas de calidad$ python3 medidas_calidad.py
scoresB_clientes scoresB_impostores 0.25 0.5 0.75
Para x=0.250000:
  El umbral FN=x es 0.058333 y FP(FN=x) es 0.162179
  El umbral FP=x es 0.032287 y FN(FP=x) es 0.162238
Para x=0.500000:
  El umbral FN=x es 0.169448 y FP(FN=x) es 0.030769
  El umbral FP=x es 0.006427 y FN(FP=x) es 0.059441
Para x=0.750000:
  El umbral FN=x es 0.532139 y FP(FN=x) es 0.000000
  El umbral FP=x es 0.000000 y FN(FP=x) es 0.015385
El umbral FN=FP es 0.044359
La curva ROC se ha generado en una imagen .png y el área bajo esta es de 0.886677
El factor dprime obtenido es 0.873482
```

Figura 1: Resultado de ejecutar el script realizado para los dos sistemas A y B

A continuación explicaremos como hemos obtenidos esos resultados.

Librerías utilizadas

El script se ha hecho en Python utilizando las siguientes librerías:

- **Sys**, para acceder a variables y argumentos utilizados por el intérprete que el usuario le pasa como parámetros.
- El módulo **Pyplot** de la librería **Matplotlib**, para dibujar las gráficas que representan las curvas ROC.
- **Numpy**, que nos asiste a la hora de trabajar con funciones matemáticas de alto nivel y poder así operar con vectores o matrices, en caso de ser necesario.

Procesado de datos

Primero se han procesado los datos de los ficheros iniciales de tal forma que se crean tres listas: una donde se guardan los scores de los clientes, otra para los de los impostores y otra con todas las muestras donde cada una está etiquetada con un identificador (0 si es cliente y 1 si es impostor), constituyendo una lista de tuplas.

Cálculo de la curva ROC y AROC

A partir de la lista de todos los scores con su correspondiente etiqueta, recorreremos cada una de las muestras, calculamos el ratio de falsos positivos (FP) y de falsos negativos (FN) tomando como umbral la muestra actual.

Luego, guardamos todos los ratios de FP en una lista y los de 1- FN en otra.

Posteriormente, para calcular el área bajo la curva ROC, seguimos la fórmula siguiente:

$$AROC = \frac{1}{mn} \sum_{i=1}^m \sum_{j=1}^n 1_{p_i > p_j} \quad (1)$$

Donde m y n son el número de clientes e impostores, respectivamente. De esta forma, cada score de los clientes se compara con cada uno de los impostores y, en caso de ser mayor el del cliente, se suma 1 al total. Finalmente, se normaliza dividiendo por el producto de m y n .

Así, lo ideal es que todos los clientes tengan mejor score que los impostores (y, por tanto, ocupen las primeras posiciones), y la ecuación tiene en cuenta la posición que debería tener un cliente entre la posición que realmente ocupa.

Una vez hecho esto, dibujamos la curva ROC y su área en una gráfica.

Hacemos esto para ambos sistemas A y B (véase las figuras 2 y 3).

Cálculo de FP(FN=X), umbral FN=X, FN(FP=X), umbral FP=X y umbral FP=FN

En esta parte, primero creamos una lista donde almacenamos los FN, partiendo de la lista de 1 - FN calculada previamente.

Para cada x que se ha proporcionado, calculamos FP(FN= x) y FN(FP= x).

Para calcular el primero se busca x en la lista de FN (en caso de no estar, nos quedamos con el valor más cercano), guardándonos su índice en dicha lista.

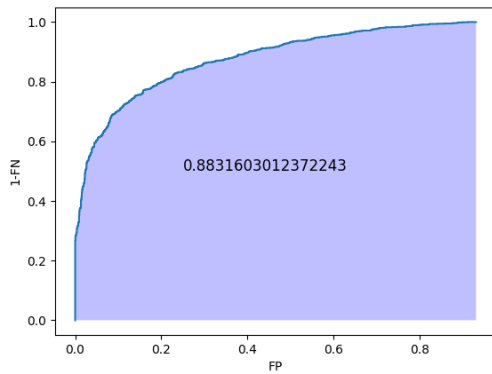


Figura 2: Curva ROC y área para el sistema A

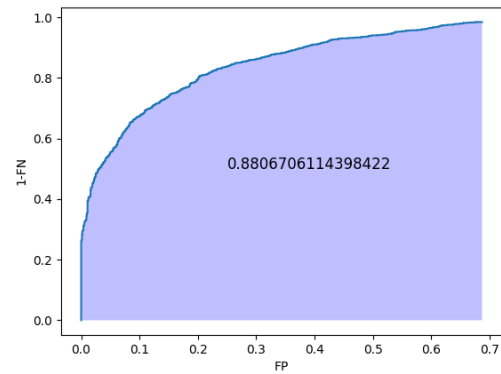


Figura 3: Curva ROC y área para el sistema B

Con este índice, indexamos en FP y obtenemos $FP(FN=x)$. Para calcular su umbral, indexamos el índice en la lista total de scores.

Para $FN(FP=x)$ y su umbral, se realiza el mismo procedimiento, pero buscando primero en la lista de FP y luego indexando en la de FN.

Finalmente, calculamos el umbral $FP=FN$. Para ello, calculamos la distancia entre cada elemento de FP con su correspondiente en la misma posición en FN, y añadimos dicha distancia a una lista.

Dentro de esta lista, obtenemos el índice de la distancia mínima de todas ellas, la indexamos en la lista total de scores, obteniendo el umbral $FN=FP$. Para el sistema A, este umbral $FN=FP$ es 0.05, mientras que para el sistema B es 0.04 (véase la figura 1 para apreciar estos resultados).

En nuestro caso, se realizaron estos experimentos dando a x los valores 0.25, 0.5 y 0.75. Los resultados obtenidos se pueden apreciar en las siguientes tablas:

x	Umbral $FN=x$	$FP(FN=x)$	Umbral $FP=x$	$FN(FP=x)$
.25	.06	.14	.04	.16
.5	.11	.02	.02	.07
.75	.29	0	.01	.02

Tabla 1: Resultados (aproximados a la centésima) del sistema A para cada valor de x .

x	Umbral $FN=x$	$FP(FN=x)$	Umbral $FP=x$	$FN(FP=x)$
.25	.06	.16	.03	.16
.5	.17	.03	.01	.06
.75	.53	0	0	.01

Tabla 2: Resultados (aproximados a la centésima) del sistema B para cada valor de x .

Cálculo del factor d-prime

Para este apartado, hemos partido de la fórmula proporcionada en las transparencias de la asignatura, que es:

$$d' = \frac{\mu_{pos} - \mu_{neg}}{\sqrt{\sigma_{pos}^2 + \sigma_{neg}^2}} \quad (2)$$

Donde μ_{pos} es el promedio de todos los scores de clientes y μ_{neg} el de los impostores. σ es la desviación típica, que la hemos calculado como:

$$\sigma_c = \sqrt{\frac{\sum_i^{N_c} (x_{ic} - \mu_c)^2}{N_c}} \quad (3)$$

Siendo c la etiqueta que discrimina entre clientes e impostores y x_{ic} las muestras de los scores pertenecientes a la clase c .

De esta forma, hemos obtenido un factor d-prime de 0.76 para los scores de A y 0.87 para los de B.