Biometrics (CSE 40537 / 60537)
University of Notre Dame, Spring 2022
Assignment 1: Comparison of Biometric Systems
Due date: Jan 28, 11:59 PM ET
Total: 10 points

1. Introduction

The purpose of this assignment is to train and evaluate the students' capabilities to compare the output of different Biometric systems, regardless of the trait they rely on. To do so, three files containing Biometric system outputs are being provided, one for each hypothetical system, namely *s1.csv*, *s2.csv*, and *s3.csv* (https://bit.ly/35dEy69). The *Python* function to load the content of these files into memory was provided in the first coding class (https://bit.ly/3nKjzhp).

1.1. Input file format

Each input file is a text file containing a header line (starting with the "#" character, which should be ignored), and 10,000 following lines, one for each response of the respective Biometric system. Lines contain comma-separated values (hence the ".csv" extension), being one value related to the actual *label* (a.k.a. ground truth) of the output provided by the system, and the other value related to the *score* computed by the system. We call each of these lines a "*(label, score)*" system *observation*. Figure 1 summarizes the content of *s1.csv*, for illustration.

```
# Output of System 1. All scores express similarities.
# label [0: impostor, 1: genuine], similarity score
0,67.1673
1,122.1142
1,123.3850
1,93.5485
0,21.3947
(...)
```

Figure 1. First 7 lines of file *s1.csv*

Labels are either *0* (for *impostor* trait presentations, such as face presentation followed by wrong identity claim), or *1* (for *genuine* trait presentations, such as face presentation followed by correct identity claim). Scores are real numbers comprising either *similarities* or *distances* between the presented trait and the claimed identity template, computed by the respective hypothetical system. Each input file independently defines if all of its scores are either similarities or distances.

1.2. Assignment directions

To complete this assignment, you will need a computer with *Python 3* (https://www.python.org/downloads/) and the *matplotlib* library (https://matplotlib.org/). A quick install of *matplotlib* can be done through *PyPI* (https://pypi.org/).

A good starting point to solve this assignment is the *Python* program implemented as *metrics.py* (https://bit.ly/3nKjzhp), which was presented during the first coding class. All the three input files and the metrics program are available to the students through the following link: (https://bit.ly/35dEy69). After downloading and unzipping their content, please follow the instructions and answer the questions presented in Sec. 2

There is no formal template for providing your answers. You may use the editor you like. The following options should work fine:
- A single PDF file or Word document containing all your answers and generated figures.
- An e-mail with your answers and the generated figures attached to the message.

Please share your answers through Slack (https://nd-biometrics-spr22.slack.com, *@dmoreira*) or send them to daniel.moreira@nd.edu by January 28, 2022, 11:59 PM ET.

2. Questions
Considering the content of the three input files *s1.csv*, *s2.csv*, and *s3.csv*, one for each hypothetical Biometric system, please answer the following questions. You may leverage and adapt the functions and metrics available in *metrics.py*.

2.1. For each one of the three Biometric systems, what score threshold (a.k.a. operating point) should you use? Please explain your answer and describe how you have obtained each one of the respective system thresholds. (0.2 point)

2.2. For each system, plot and provide a graph with the distributions of their respective scores. (0.1 point)

2.3. According to the d' (d-prime) values that one might compute for each system, which of the three should you use if you had to select only one for identification? Please justify your answer. (0.2 point)

2.4. Plot and provide a single graph with the ROC curves and AUCs of all the three systems together. A reference to help you: https://bit.ly/3rrnSPR. (0.1 point)

2.5. According to the ROC curves and AUC values, which one of the three systems should you use if you had to select only one for identification? Please justify your answer. (0.2 point)

2.6. Please provide your modified version of *metrics.py*, adapted to deal with scores that are distances rather than similarities. (0.2 point)