**2. Questions**

2.1. As explained in class, the provided third-party face recognition library is able to extract a 512-dimensional float feature vector from a given face image, as well as calculate the angular distance between two feature vectors, using ArcFace (https://bit.ly/3J8Tgtc). The expected behavior for the software is to generate small distances for two face images that depict the same individual (genuine pair), and large distances for two images that depict different individuals (impostor pair).

Leveraging the content of **only** the "dataset" folder within the provided data, the third-party face recognition library, and metrics learned in class, please determine **what is a good angular distance threshold to separate genuine from impostor pairs**. While providing your answer for the distance threshold, please explain in detail how you computed it. (1 point)

My distance threshold is 1.6755. I calculated this by first creating 360 genuine and imposter pairs each. I did 36 combinations of the different photos for each subject. I chose 36 because each subject had at least 9 photos and there are 36 combinations of 9 photos. For each of the 10 subjects, this resulted in 360 genuine pairs. I then created 360 imposter pairs by doing similar combinations except using different subjects. From here I used the metrics function to find the optimal EER which was 1.6755.

2.2. **What is the AUC** of the face recognition system you are using? In addition, please provide a graph with the system's **ROC curve**. (1 point)

The AUC was 0.96 as seen in the graph below.
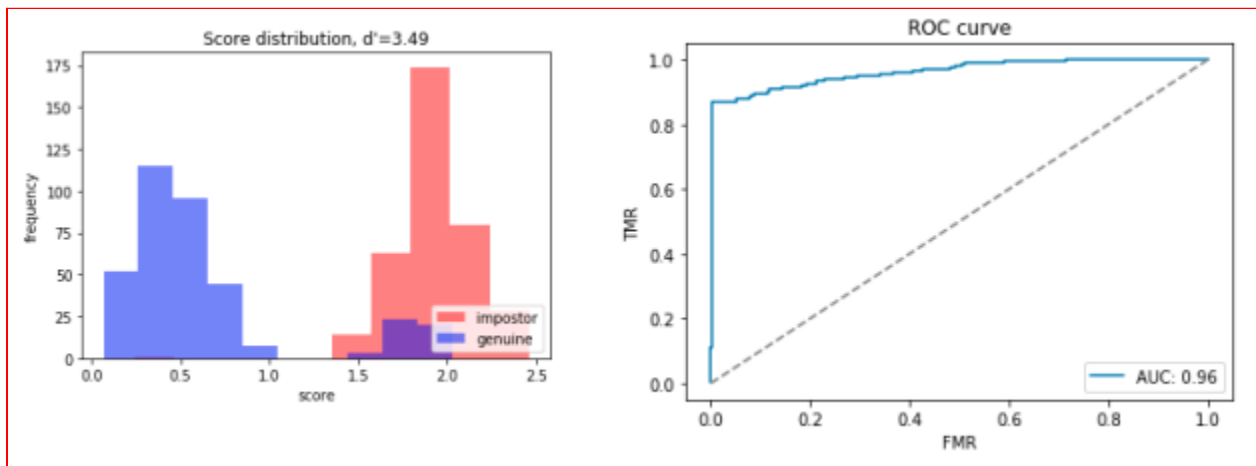
2.3. By leveraging the face recognition system and the distance threshold previously computed, and by either capturing your face with your webcam or providing an image with your face, find within the "dataset" folder what is the individual that is the most similar to you. **Please provide the subject ID and the angular distance between your face and theirs**. In your opinion, do

you have anything in common with this subject (e.g., gender, ethnicity, age, etc.)? If yes, what is it? (2 points)

I had the closest distance with subject4. The closest distance computed with my picture was his picture with centerlight that had a distance of 1.549. We are both males and look to be around the same age. We are not of the same ethnicity but we have similar hair lengths and our eyebrows are similar looking.

2.4. By leveraging the face recognition system and the distance threshold previously computed, **please provide the subject ID** (or "**UNKNOWN**", if the individual does not have a face within the "datasets" folder), as well as **the respective angular distances** that supported your decision, for each one of the 15 images provided within the "queries" folder. (6 points)

Based on the calculated threshold, all of these were genuine pairs, but the ones indicated with the FALSE were actually imposters. The threshold was too high, but this was because I compared each query to every possible image. Looking at the score distribution graph, a threshold of around 1.2 would make more sense and make the system more secure. If that was the threshold then all of the queries would have been identified correctly. When generating the genuine pairs there were a few pairs that had a high distance, probably due to lighting and glasses. This skewed the calculated threshold, but looking at the graph and seeing these results a threshold of 1.2 would make it more secure and result in better comparisons.

The subjects bolded are genuine matches using the threshold of 1.2 that is based on the graph. The matches with FALSE are imposters and do not have a match in the dataset.

4168: distance of 0.13350804 with subject01
4387: distance of 0.0 with subject02
4507: distance of 0.0 with subject07
4535: distance of 0.0 with subject03
5314: distance of 0.0 with subject08
6012: distance of 1.4790814 with subject07 FALSE
6510: distance of 1.511008 with subject10 FALSE
6653: distance of 0.063178495 with subject05
6706: distance of 1.6050752 with subject09 FALSE
7076: distance of 0.0 with subject06
7549: distance of 0.0 with subject09
7633: distance of 0.0 with subject04
7745: distance of 1.4956629 with subject02 FALSE
9395: distance of 0.034079973 with subject10
9708: distance of 1.3111169 with subject05 FALSE