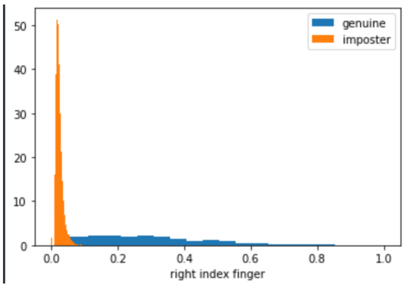
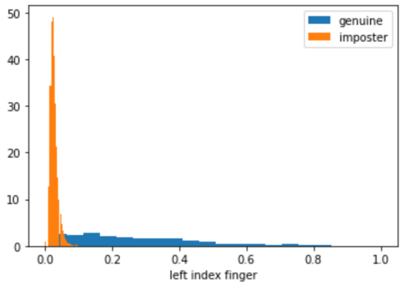
Assignment1

Question1: Score distributions

1.1 Plot the genuine and impostor score distributions in a single plot.

Draw the distribution using `matplotlib.pyplot.hist`





1.2 Do you need to normalize the distributions? Why (not)?

No. In the genuine and imposter graph, the main focus is how the threshold would influence the proportion of false rejection and false acceptance. We are interested in the absolute value of threshold.

1.3 Describe qualitatively this combined plot (hint: limit the score range for better understanding)

The imposter distribution is located in the left range and limited mainly between 0 and 0.1, while the genuine distribution is located in the right and has a wider range.

As threshold decrease toward 0, less genuine would be recognized as imposter. Though some imposters might be wrongly accepted.

If we increase the threshold a little bit, false reject will increase and false accept will decrease.

If we continue to increase the threshold, the number of imposters will drop rapidly and it would mainly be genuine. There wouldn't be any imposter accepted as genuine. But some genuine might be refused.

The gap between the genuine and imposter is relatively small, which means the system has a good ability for verification.

Question2: ROC Curves

2.1 Calculate FPR, TPR from the matching scores

By using sklearn tool, it is easy to compute FPR and TPR for different thresholds.

The code is shown below

```
from sklearn import metrics
FPR_r, TPR_r, thresholds_r = metrics.roc_curve(ri_genuine_id, ri_scores, pos_label=1)
FPR_l, TPR_l, thresholds_l = metrics.roc_curve(li_genuine_id, li_scores, pos_label=1)
```

The full result is printed in the jupyter notebook.

For example, for the right hand, if the threshold is 0.066, the TPR and FPR is 0.9 and 0.007 respectively .

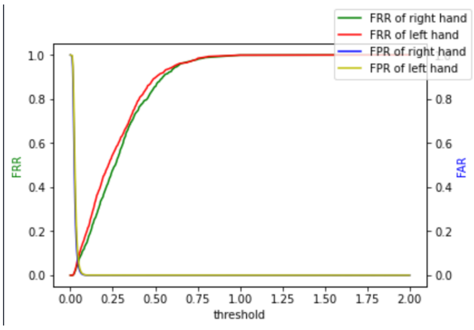
2.2 Plot FAR and FRR as a function of matching scores.

FAR means false acceptance rate, which is equal to false positive divide by condition negative.

FRR means false reject rate, which is equal to false negative divide by condition positive.

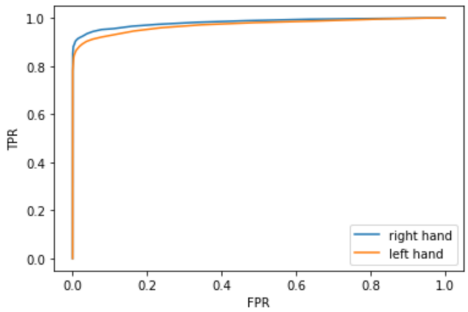
Since FAR = FPR and FRR = FNR = 1- TPR, we can use the result from the above code of ROC.

The graph of FAR and FRR as a function of decision threshold is shown below:



As is shown in the graph, for both left and hand, if we increase the threshold, the FAR will drop quickly, which means there will few imposters be recognised as genuine. Compared with FAR, the increase of FRR is slower, which means that it is very possible that genuine are treated as imposters. So the system is very safe but not very convenient.

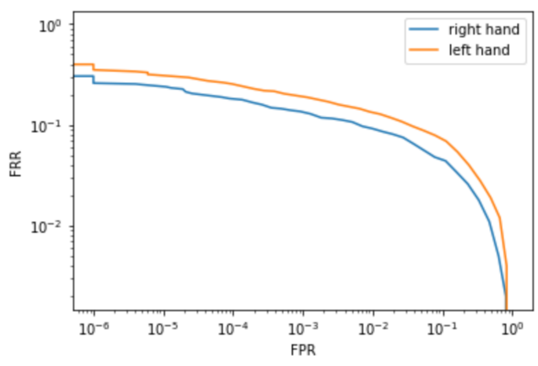
2.3 Plot the ROC curve. Plot for linear and logarithmic scale if needed. What do you observe?.



This algorithm has a very sharp slope. The left upper point it very close to (0,1) which means it can have a very high accuracy by increasing threshold.

The curve of right hand is above left hand, which means the classifier of right and has overall better performance. The right hand can accept more genuine while keeps the same false accept rate.

2.4 Plot the Detection Error Trade-off (DET) curve. How does it compare to ROC?



Since FAR = FPR , FRR = 1- TPR. So DET is the upside down version of ROC with different scaling of axis. Compared with ROC, it is easier to see the comparison between errors. Since the slope of ROC is very steep, using DET can better compare different classifier.

Question3: F1 and accuracy as metrics

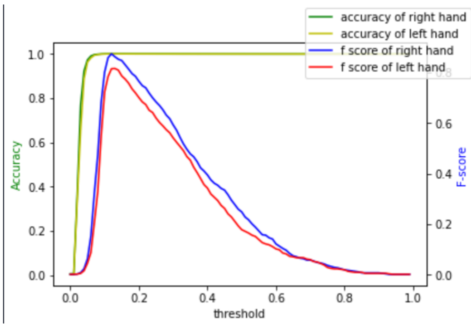
3.1 Plot F1 and accuracy as a function of the decision thresholds on the similarity score.

Accuracy is $\frac{TP+TN}{TP+TN+FP+FN}$

F-score is $\frac{2}{precision^{-1}+recall^{-1}}$. It is basically the average of precision and recall.

Sklearn toolbox has specific function to calculate accuracy scores and f1 scores.

The result graph is shown below:



3.2 Calculate the threshold and accuracy for which F1 is maximal. Is it an interesting operating point?

Using sklearn library is super easy to calculate the best threshold for which F1 is maximal.

For the right hand, threshold is 0.12 and accuracy is 0.999769 when F1 is maximal. For the left hand it is 0.13 and 0.999681.

 It is an interesting operating point since we need to keep precision and recall both high. But in reality we need make a trade-off between precision and recall. So when F1 score is maximal, it can represent that the system is able to keep high precision while keeps recall also high. Though it doesn’t consider a lot about the true negative, the true positive is more important for verification system.

3.3 Do the same for the classification error (accuracy). Is there a difference?

The result is 0.12 and 0.13 for right and left hand, which are the same as using F1 score.

3.4 Is accuracy a good performance metric in this case?

In this case, accuracy is not a good performance metric due to the unbalance between genuine and imposter. The proportion between 0 and 1 is 999:1. The zero is much more than 1. So accuracy is not able to express how much 1 is predicted. And for verification system, we need to make FPR as low as possible.

4. AUC and EER as summary measures

4.1 Calculate ROC AUC. Is this a good metric? What does it reveal about the system?

By using library `sklearn.metrics.roc_auc_score` , it is easy to directly calculate ROC AUC.

It is a good metric for imbalanced data such as this situation. Since the number of 1 is very small compared with 0, the ROC is very sensitive for 1, which is related to FPR and TPR. The more closer AUC is to1, the better the system is. In this situation, the AUC for right hand and left hand is 0.983 and 0.971 respectively. They are both good classifier. But the right hand is slightly better than left hand.

4.2 Calculate (by approximation) the EER and plot it on the FAR-FRR curve. Is this a good operation point?

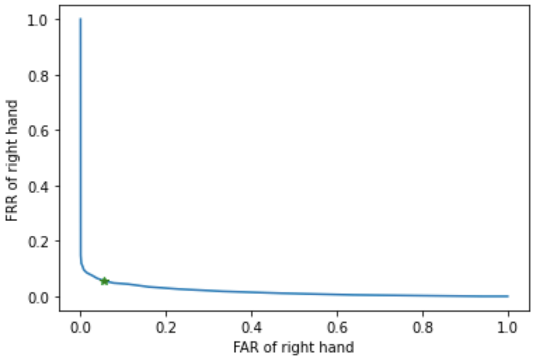
The EER curve is FPR + TPR -1 = 0. By finding the point on AUC curve that is closest to this curve, EER point can be found. The code is shown below:

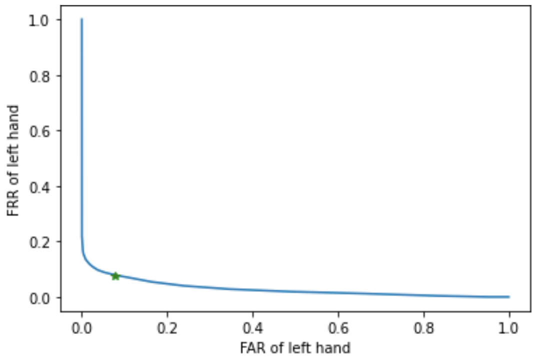
```
FPR, TPR, thresholds = metrics.roc_curve(ri_genuine_id, ri_scores, pos_label=1)

distances = abs(FPR + TPR -1)
# print(np.argmin(distances))
index = np.argmin(distances)
ERR = [FPR[index], TPR[index]]
print(ERR)
```

The EER points for right hand and left hand are [0.054403403403403404, 0.944] and [0.07793493493493493, 0.921] respectively.

Since the using situation is not given, EER is a good metric to balance FRR and FAR.

The graph is shown below:





4.3 Calculate the decision threshold for which the sum of FRR and FAR is minimal. Is this point similar to the total classification error?The code use to calculate threshold for which the sum of FRR and FAR is minimal is shown below:

The result threshold for right hand and left hand is 0.059 and 0.057.

$error = \frac{FN+FP}{FN+FP+TN+TP}$.

$FPR = \frac{FP}{FP+TN}$ and $FRR = \frac{FN}{TP+FN}$

It is clear that error is not equal to FPR + FRR

4.4 Can you suggest other strategies that give you an "optimal" performance? Calculate and discuss their (de)merits.

Cross entropy loss is another way to compare the performance of classification error. The formula is shown below:

$$L = -\frac{1}{N}[\sum_{J=1}^{N}[t_j log(p_j) + (1-t_j)log(1-p_j)]]$$

The cross entropy loss for right and left index finger is 0.026 and 0.029 respectively. The right hand has less loss which means less error.

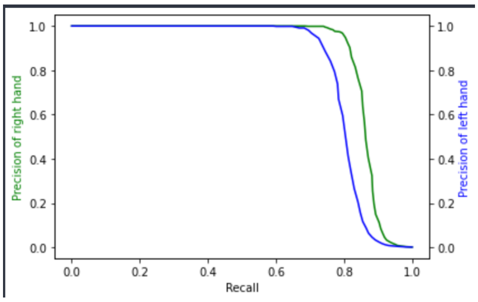
Question5: Evaluating using Precision and Recall

5.1 Calculate and plot the Precision-Recall curve for this system. What does it reveal about the performance of the system?

$Precision = \frac{TP}{TP+FP}$

$Recall = \frac{TP}{TP+FN}$

The resulting curve is shown below:



Compared with ROC curve, PR curve focuses more on positive example. Thus it can avoid the problem of imbalanced data such as this situation.

In this graph, Precision can sustain a very high level, which means that it is hard to treat an imposter as a genuine sample. But recall can vary a lot, which means that it is hard to recognise all positive number.

It is pretty clear that the curve of left hand is below right hand. It can be concluded that the right classifier can better recognise positive sample.

5.2 Calculate the Area Under the PR-curve. Discuss.

The values for right hand and left hand are 0.86 and 0.80 respectively. Since larger precision and recall number are better, the classifier for right hand is better than left hand.

5.3 Calculate the average precision scores. Discuss its value.

The resulting score for right and left hand is 0.8595 and 0.7989 respectively.

Calculating average precision score is similar to calculating AOC of PR curve. The only difference is that AP doesn't use curve fitting. The formula is $\sum^{n}(P_n - P_{n-1}) * R_n$. The result is similar to ROC. So right hand has better performance than left hand.

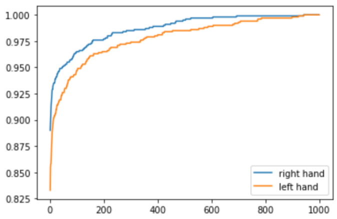
Question6: CMC Curves

6.1 Calculate the Cumulative Matching Characteristic curve

The code to calculate the curve is shown below:

```
ri_matrix = ri_similarity_matrix.to_numpy()
probability = []
for rank  in range(1,1000):
    correct = 0
    for row in range(0,1000):
        columnArfSort = np.argsort(-ri_matrix[row])
        candidates = columnArfSort[0:rank]
        if row in candidates:
            correct +=1
    probability.append(correct/1000)

plt.plot(range(1,1000), probability)
```

The resulting curve is below:



6.2 Compute the Rank-1 Recognition Rate.

The code to compute the rank-1 recognition rate is shown below:

```
correct = 0
for row in range(0,1000):
        columnArfSort = np.argsort(-ri_matrix[row])
        candidates = columnArfSort[0:1]
        if row in candidates:
            correct +=1
probability_1 = correct/1000
print(probability_1)
```

The result for right hand and left hand is 0.89 and 0.833 respectively.

Question7 Evaluating different biometric systems

For verification and identification, there are different metrics.

Verification:

1. ROC curve: the curve of right hand is above left hand. It is fair to say that right hand perform better than left hand. And the AUC of ROC is also larger for right hand.
2. FPR and FRR curve: The FRR curve is similar. But the FPR curve of right hand is lower than left hand. So right hand is better.
3. The accuracy curve is similar. But right hand has higher f score curve. So it is better
4. Precision and recall curve: The curve of right hand is higher than left hand. Right hand is better.

Identification:

CMC curve: right hand is higher than left hand which means given a series of candidates, the prediction result of right hand would more fall into the range of the candidates.

Right hand is better.

Overall right hand is better for both situation.