Programming Assignment 1
CSCI-860: Biometrics in a Networked Society

Presented To
Kiran Balagani
Associate Professor of Computer Science
New York Institute of Technology

Prepared By:
Pranav S. Krishnamurthy
Major – Computer Science
Student ID: 0704111

Due Date – 2nd December 2014
Submission Date – 4th December 2014
Modified Date – 20th December 2014

Problem Statement

This assignment requires the implementation of the Manhattan verifier to report *false accept (impostor pass)* and *false reject rates* on a publicly available keystroke biometric dataset. Any programming language may be used, as long as the program can be compiled on computers in Harry Schure Hall Room 212. In addition, a demonstration and explanation of our written code will be required.

My Approach

For this programming assignment, I have used the C#.net programming language, the Visual Studio 2013 Ultimate IDE for Windows 8.1 along with the PasswordData.csv file which I manipulated into 51 separate csv files, and when running the program, I have the code written such that depending upon the user number and the value of N, which represents the sample size: the training and testing vectors are extracted from the original data, the template vectors, genuine scores, impostor scores, and the various rates are calculated appropriately. The approach that I have taken for this assignment is detailed below:

1. Taking the CSV file from the link provided in the PDF file given, I took that file and broke that into 51 separate CSV files which represents each user.
2. Depending upon the user number that the end user enters (the user number will act as the "tag" for the subject ID), the program will then find the file in the respective directory, and convert the entire CSV file into a 2D double array.
3. When the end user enters a value for N, which can have the value of either 100, 200, or 300; the program will then parse through using the 2D double array, and the value of N and extract the testing samples which will be used to generate the mean vector.
4. The mean vector, the value of N, and the remaining samples are then parameters for generating the genuine scores. Here, I have assumed that the value of n in the formula for the Manhattan Distance is equal to the number of samples.
5. Then to generate the impostor scores, I will then be taking all the other user data which contain 400-N samples, and the mean vector of the user as parameters.
6. Finally, will be taking a look at the various genuine scores and using a threshold T and using the same threshold T, and looking at the various impostor scores, will calculate the Impostor Pass Rate and False Reject Rates.

Please note that I will be having three various plots and they will be properly denoted as Figure I, Figure II, and Figure III. These figures will be shown in the results section which follows.

False Accept Rate

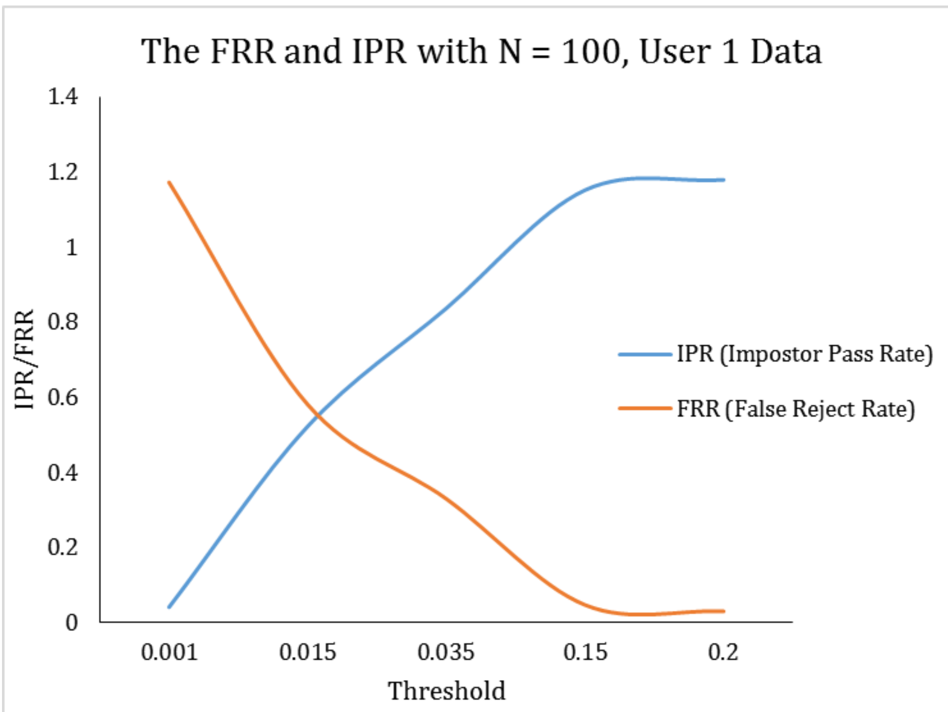
This is sometimes referred to as the impostor pass rate, to find this measure, I will be taking a threshold that the end user will put in and in all of the impostor scores that I have calculated, I will compare the value of the user entered threshold and the score. If the impostor score is less than or equal to the threshold value, that means the impostor has passed through the system. The screenshots below, I will show the values of thresholds that I have arbitrarily set. There will be three subsections which represent the various values of N, and also in each subset of screenshots, I will show the threshold values, and I will have five threshold values in total in each graph in the section entitled *results*.

Results

The following figures, Figures 1 through 3 are the results which have the tables and graphs. I have chosen five thresholds that I feel best meet the following criteria:

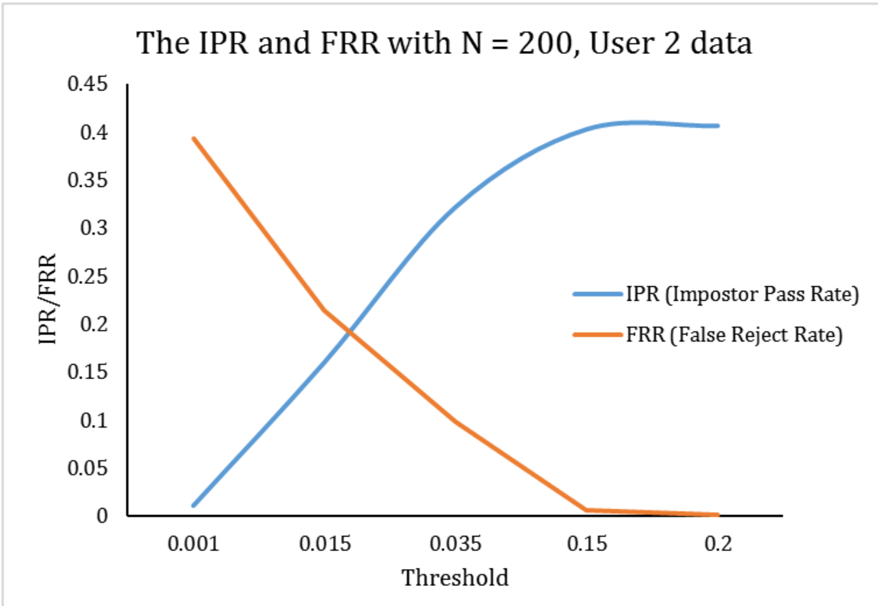
- Threshold which gives the highest possible False Reject Rate at 0 Impostor Pass Rate
- Threshold which gives the highest possible Impostor Pass Rate at 0 False Reject Rate

Figure 1. I have chosen to analyze the data from user 1, with the sample size of N = 100.



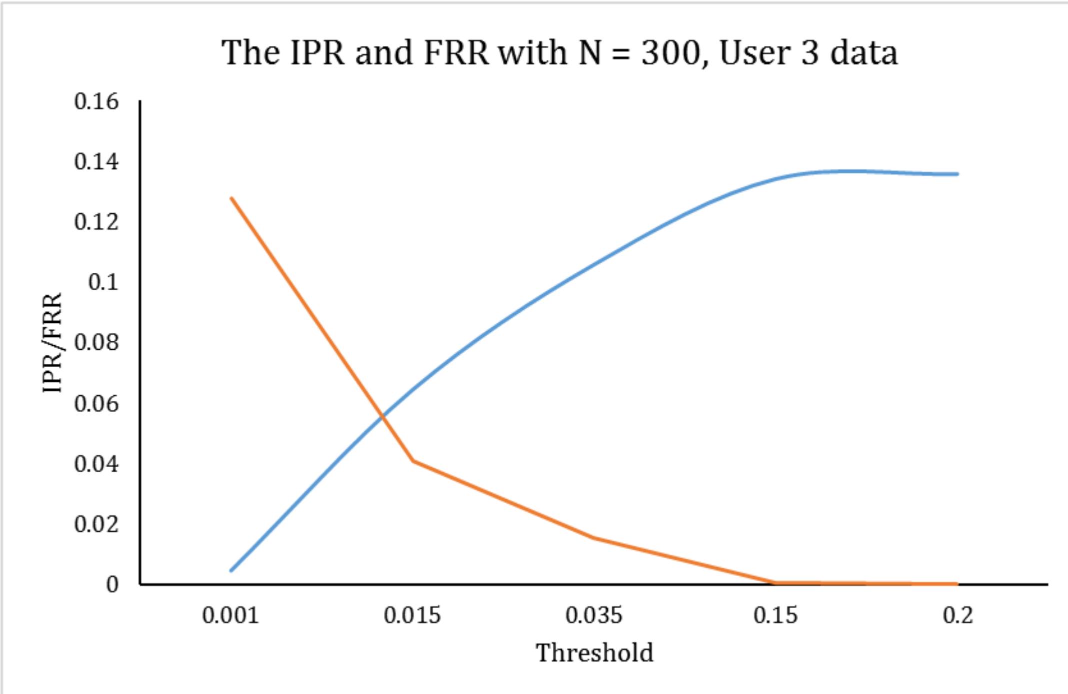
From the figure above, when I take into account the data from User 1, the equal error rate is around 60%, which is the intersection point of the IPR (Impostor Pass Rate) curve and the FRR (False Reject Rate) curve. Next, I have included Figure II below where I have chosen to analyze the data from user 2, with the sample size of N = 200.

Figure 2. Results of when I have chosen to analyze the data from user 2, with the sample size of N = 200.



Here since I have kept the thresholds the same, and I have increased the sample size from the value of N = 100, to N = 200, I have noticed that the equal error rate which is the intersection of the IPR (Impostor Pass Rate) curve and the FRR (False Reject Rate) curve has dropped approximately 40%. Before, the EER was approximately 60%, and now the EER is about 20% or so. Finally, I will be reporting when I choose to analyzed the data of user 3 in relation to all the other user data, and I also will be expanding the number of samples from the value of N = 200 to having the value of N = 300.

Figure 3. Results of when I have chosen to analyze the data from user 3, with the sample size of N = 300.



Now, finally from the above graph I have seen that the when I have increased the sample size from N = 200 to N = 300, I observer that the EER, which is the intersection between the IPR and FRR curves is equal to approximately 6%. Before when the sample size of N = 200, the EER was about 20%, but now has decreased to a very low 6%!

Comments or Concerns

Kiran, if you have any questions, comments, or concerns you can send me an email to pkrishna@nyit.edu, or pkrish19@outlook.com.