Fingerprint Spoofing Detection - Lorenzo Bellino 309413

Introduction

Abstract

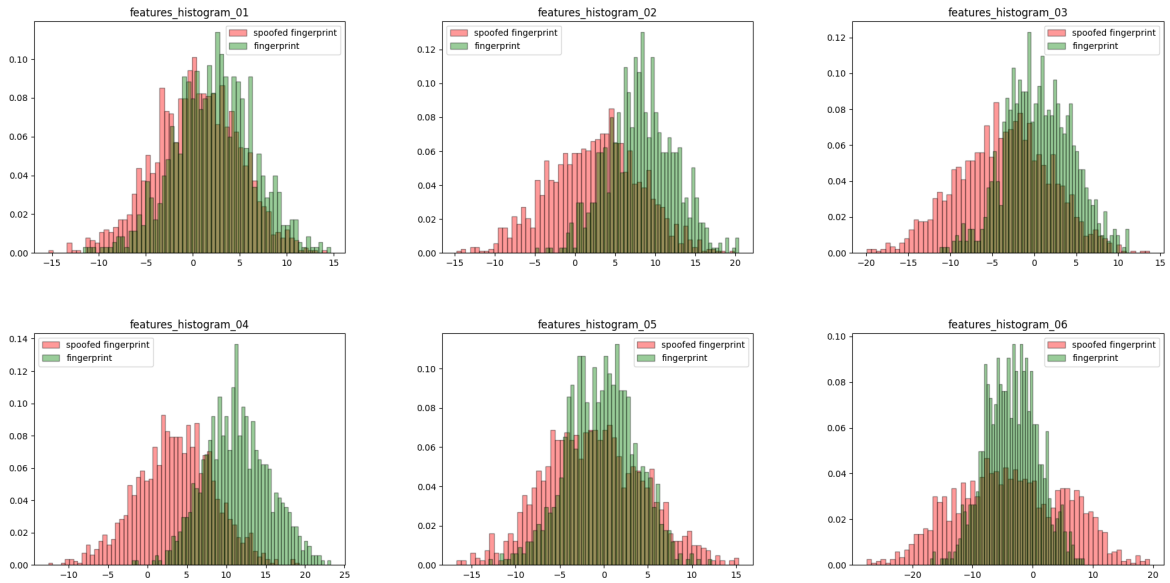
This report aims to examine a dataset composed of 2325 images of fingerprints using different Machine Learnin approaches. The dataset is composed of 2 classes: real and fake fingerprints. The goal is to classify the images in the correct class. The report is divided in 3 parts: the first one is about the preprocessing of the images and explore the dataset, the second one is about the training of the models and the third one is about the evaluation of them.

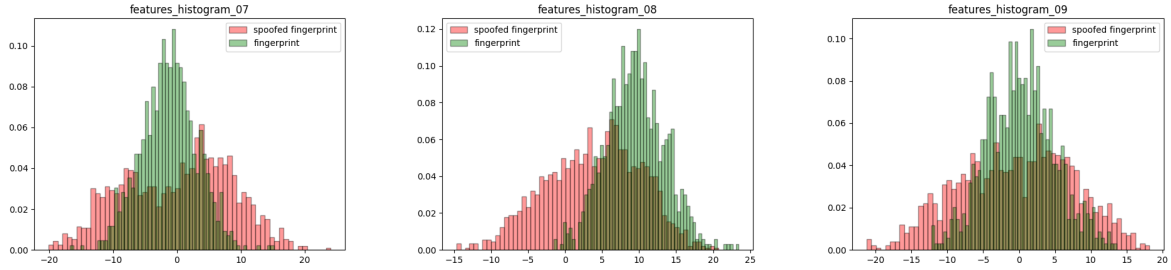
Considerations about the dataset

The dataset is composed of samples that represent fingerprint images through low-dimensional repre- sentations called embeddings. Each fingerprint is represented by a 10-dimensional vector of continuous real numbers, which is obtained by mapping the images to a lower-dimensional space. Real fingerprints are labeled as 1, while spoofed fingerprints are labeled as 0. Spoofed fingerprint samples can belong to one of six different sub-classes, each representing a different spoofing technique, but the specific technique used is not provided. The target application assumes that the prior probabilities for the two classes are equal, but this does not hold for the misclassification costs. The training set contains 2325 samples and the test set contains 7704 samples, with the fake fingerprint class being significantly overrepresented.

Feature Analysis

In the following images we can see the distribution of different features in the dataset diveded between authentic and spoofed fingerprint:

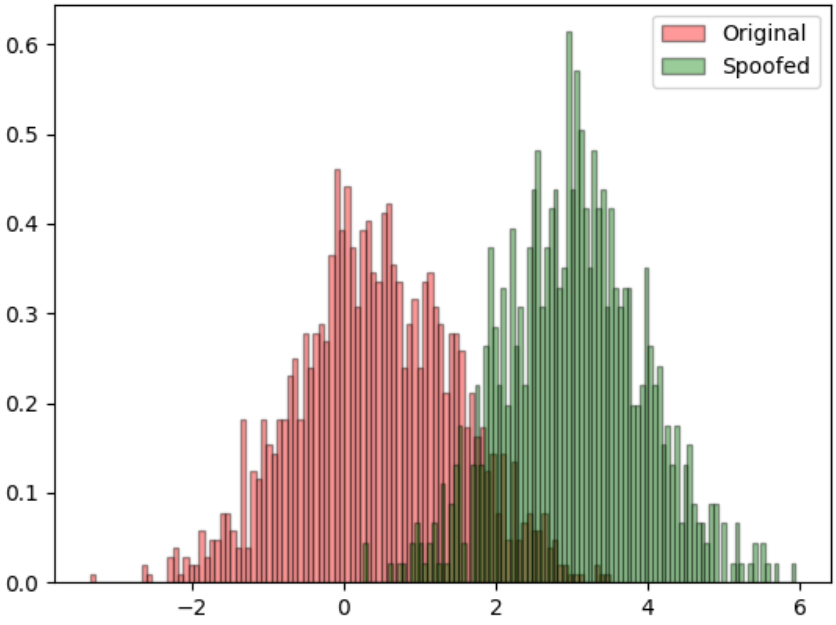




As we can see most of these features can be described by a Gaussian distribution, especially the first 4 and it is also evident that the distribution of the authentic fingerprint is much more similar to a normal distribution than the spoofed one, this can be explained by the fact that the spoofed fingerprint are actually a combination of different techniques and this could be the reason why the distribution is not so regular.

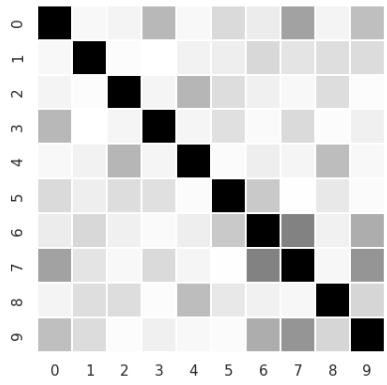
LDA

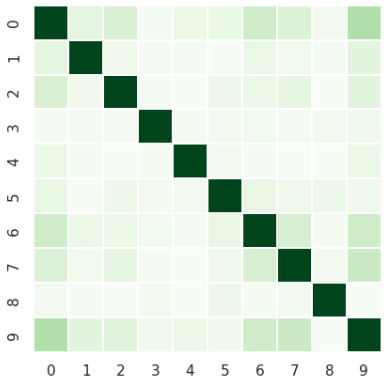
The first transformation that we can apply is the Linear Discriminant Analysis (LDA), this preprocessing step allow us to evaluate the features and decide if they can be linearly discriminble or if a gaussian model would be a better fit for this particular task. Looking at the graph we see how it is possible to use linear separation methods for the two classes but the separation would not be perfect and would lead to a number of errors.

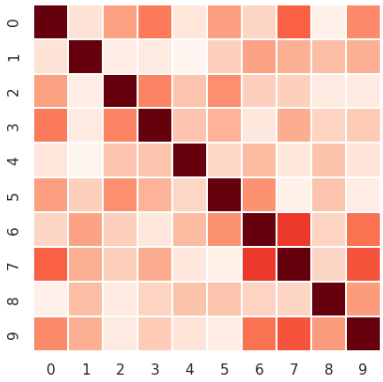


Correlation

Now we focus our attention of feature correlations using Pearson Correlations Plots


0 1 2 3 4 5 6 7 8 9
0 1 2 3 4 5 6 7 8 9


0 1 2 3 4 5 6 7 8 9
0 1 2 3 4 5 6 7 8 9


0 1 2 3 4 5 6 7 8 9
0 1 2 3 4 5 6 7 8 9