
REPORT FOR THE MACHINE LEARNING & PATTERN RECOGNITION PROJECT

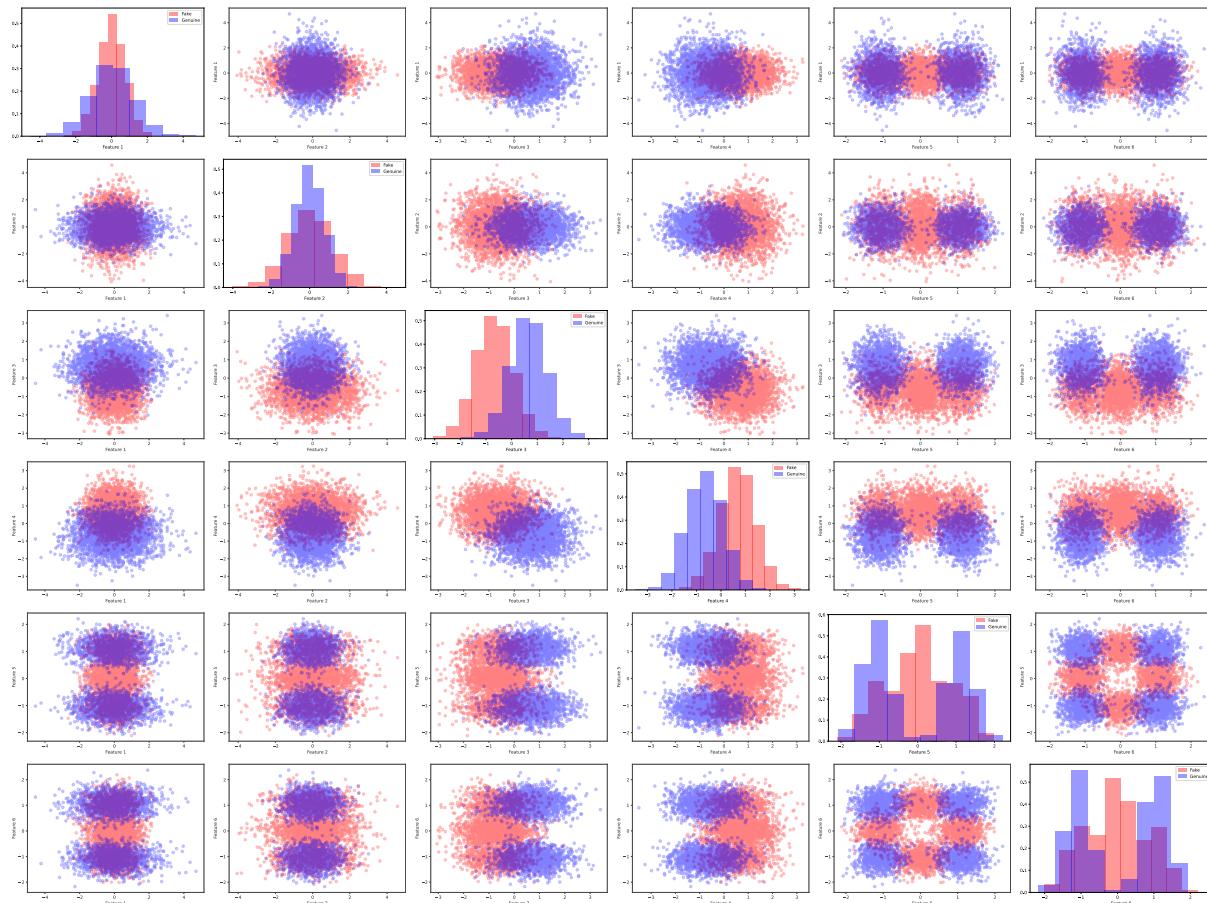
Eduard Antonovic Occhipinti

947847

INTRODUCTION

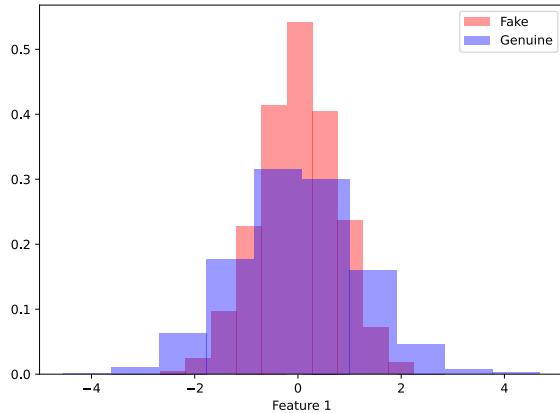
The task consists in a binary classification problem, the goal is to perform fingerprint spoofing detection (i.e. to distinguish between real and fake fingerprints).

The dataset consists of 6 features. In this first part we will analyze some statistics of the dataset and the correlation between the features.

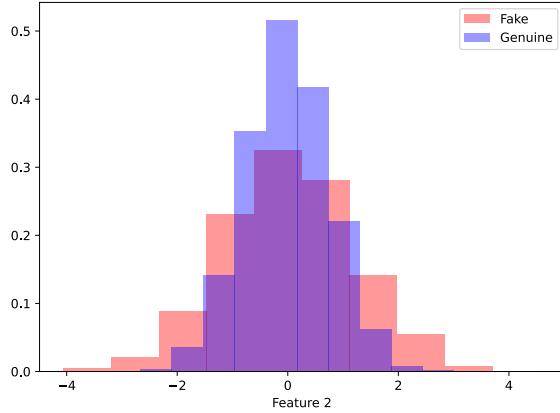


FEATURES COMPARED

FEATURES 1 AND 2



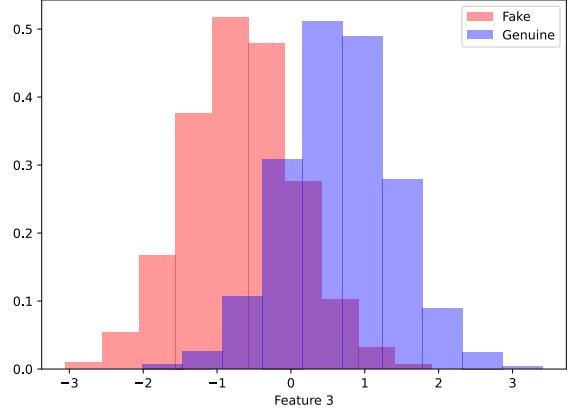
When looking at the first feature we can observe that the classes overlap almost completely. The **Genuine** label has higher variance than the **Fake** class but the mean is similar. Both classes exhibit one mode in the histogram but the **Fake** class has a higher peak.



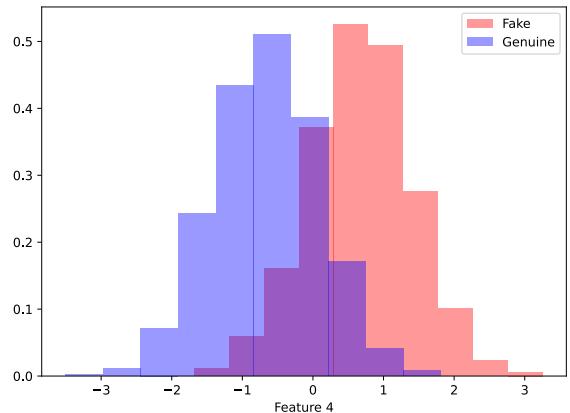
Looking at the second feature we can notice the opposite behavior. The **Fake** class has a higher variance than the **Genuine** class but the mean is similar. Both classes exhibit one mode in the histogram but

the **Genuine** class has a higher peak. Again, the classes overlap almost completely.

FEATURES 3 AND 4

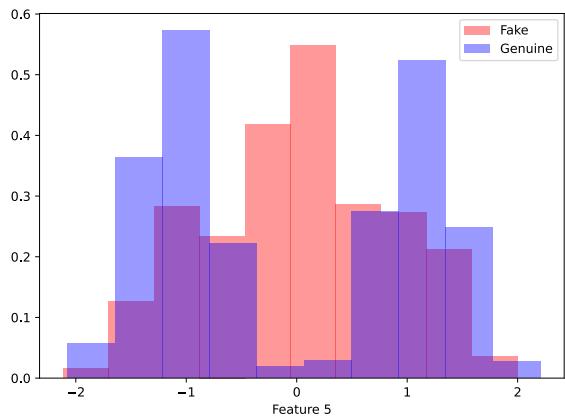


Looking at the plot for the third class we can notice that the two features are much more distinct, they overlap slightly in 0. The **Genuine** class has a peak in -1 while the **Fake** class has a peak in 1. They both have similar mean and variance. One mode for each class is evident from the histogram.

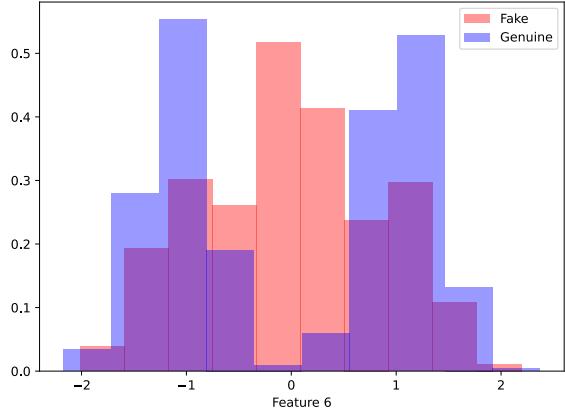


The fourth feature shows similar characteristics to the third feature.

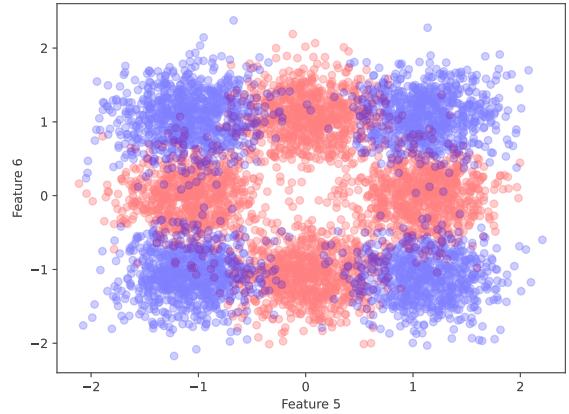
FEATURES 5 AND 6



The fifth feature also shows good distinction between the two classes with an overlap at the edges of the **Fake** class distribution. They exhibit similar variance but with a lower mean for the **Genuine** class. The **Fake** class peaks in 0 while the **Genuine** has two modes and peaks in -1 and 1 .



The last feature shows similar characteristics to the fifth feature.



Looking at the scatter plot we see that there are four distinct clusters for each of the labels, they overlap slightly at the edges of each cluster.

PCA TODO

