

Laboratory 2 - Binary Classification for fingerprint spoofing detection

Exploit all the data

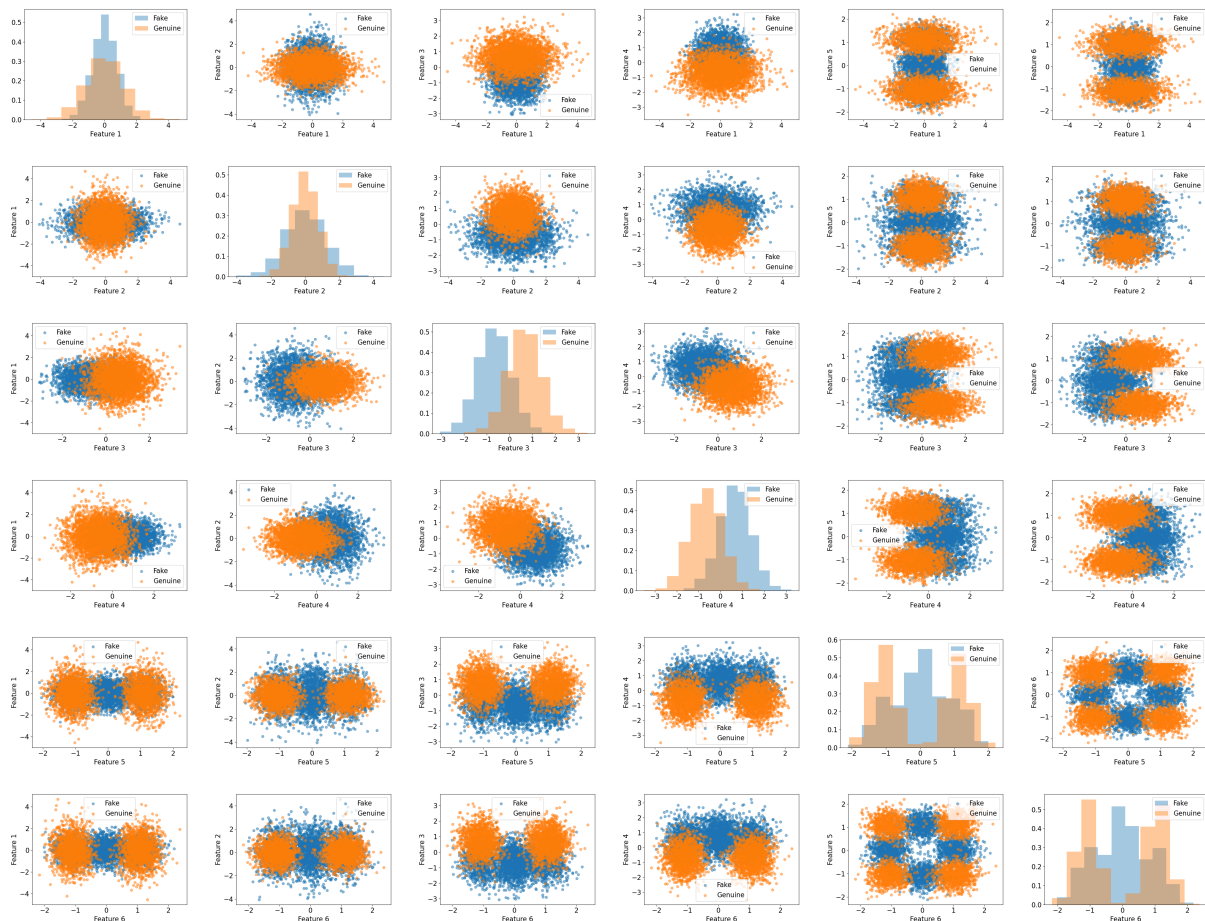


Figure 1: All features

First question

Analyze the first two features. What do you observe? Do the classes overlap? If so, where? Do the classes show similar mean for the first two features? Are the variances similar for the two classes? How many modes are evident from the histograms (i.e., how many “peaks” can be observed)?

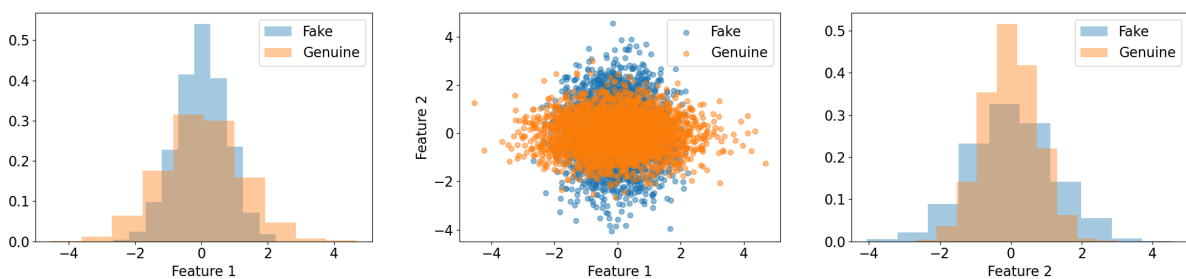


Figure 2: Analyse Feature 1 and Feature 2

Looking at the histograms, we can notice in both features classes overlap for each value assumed by the feature. Histograms have few peaks: in both features, peak is near 0, causing a class overlap around this value in the scatterplot. In particular, Feature 1's peak counts a larger number for fake detection, while Feature 2's peak is of genuine detection.

For class Fake, mean of Feature 1 and Feature 2 is respectively 0.00287744 (0.003) and 0.01869316 (0.019); variance is 0.56958105 and 1.42086571. For classe Genuine, mean of Feature 1 and Feature 2 is respectively 5.44547838e-04 (0.0005) and $-8.52437392 \times 10^{-3}$ (-0.008); variance is 1.43023345 and 0.57827792

Second question

Analyze the third and fourth features. What do you observe? Do the classes overlap? If so, where? Do the classes show similar mean for these two features? Are the variances similar for the two classes? How many modes are evident from the histograms?

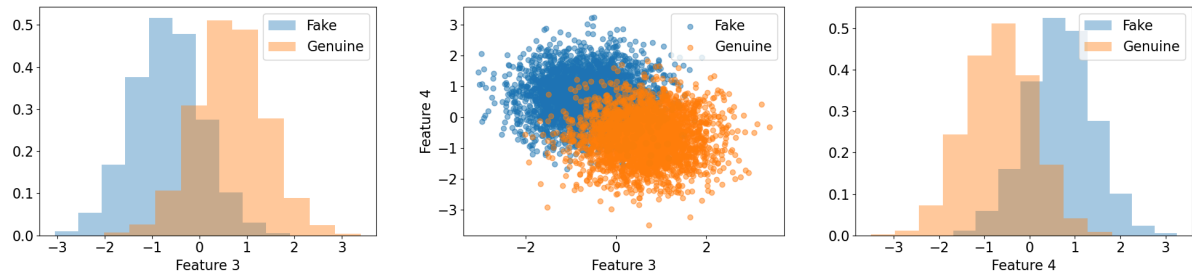


Figure 3: Analyse Feature 3 and Feature 4

Looking at histograms, we can see classes overlap for both features in the $[-2, 2]$ interval. In both features, histograms show two peaks representing both classes. For class Fake, mean of Feature 3 and Feature 4 is respectively -0.68094016 and 0.6708362 ; variance is 0.54997702 and 0.53604266. For class Genuine, mean of Feature 3 and Feature 4 is respectively $6.65237846 \times 10^{-1}$ and $-6.64195349 \times 10^{-1}$; variance is 0.5489026 and 0.55334275.

Third question

Analyze the last two features. What do you observe? Do the classes overlap? If so, where? How many modes are evident from the histograms? How many clusters can you notice from the scatter plots for each class?

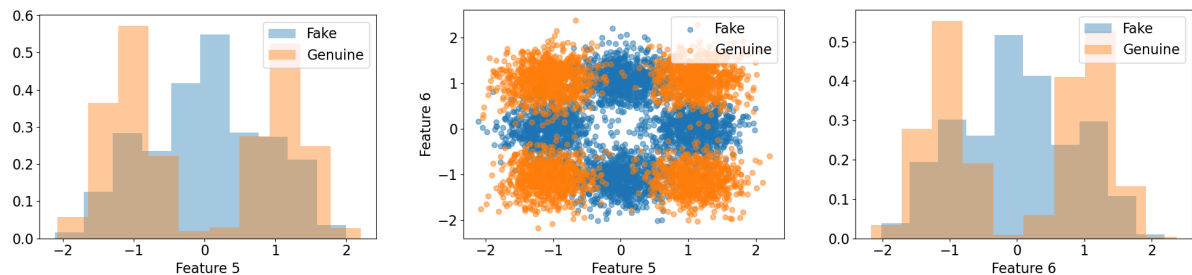


Figure 4: Analyse Feature 5 and Feature 6

Looking at histograms, we can see class overlap for both features in intervals $[-1.5, -0.5]$ and $[0.5, 1.5]$. In both features, histograms shows three peaks representing mostly Genuine class. Looking at the scatter plot, we can notice four cluster for each class. Cluster of Genuine class are in the intervals $[-2, -0.5]$ and $[0.5, 2]$, while in the residing interval (in the centre) there are only clusters of Fake class.