ML and PR Project Report – Part 3

Like the other reports, this one pertains to the analysis of the fingerprint database.

This time we are going to analyze the distribution of the data, more precisely we are going to probe whether each feature’s distribution resembles a (univariate) gaussian distribution.

# Comparing Histogram and Gaussian pdf

## Entire Dataset

In order to check if a feature’s distribution resembles a gaussian we are going to plot, alongside the histogram of the fingerprint sample data projected feature by feature, a gaussian with mean and covariance equal to the ML estimation based on the sample data, meaning that we used the empirical mean and empirical covariance formulas. In order to be able to compare the two well we are normalizing the histogram.

Immagine che contiene schermata, diagramma, Diagramma

Descrizione generata automaticamenteImmagine che contiene diagramma, schermata, Diagramma

Descrizione generata automaticamente

Immagine che contiene testo, schermata, diagramma, Diagramma

Descrizione generata automaticamenteImmagine che contiene schermata, testo, diagramma, Diagramma

Descrizione generata automaticamente

Immagine che contiene schermata, Diagramma, diagramma, testo

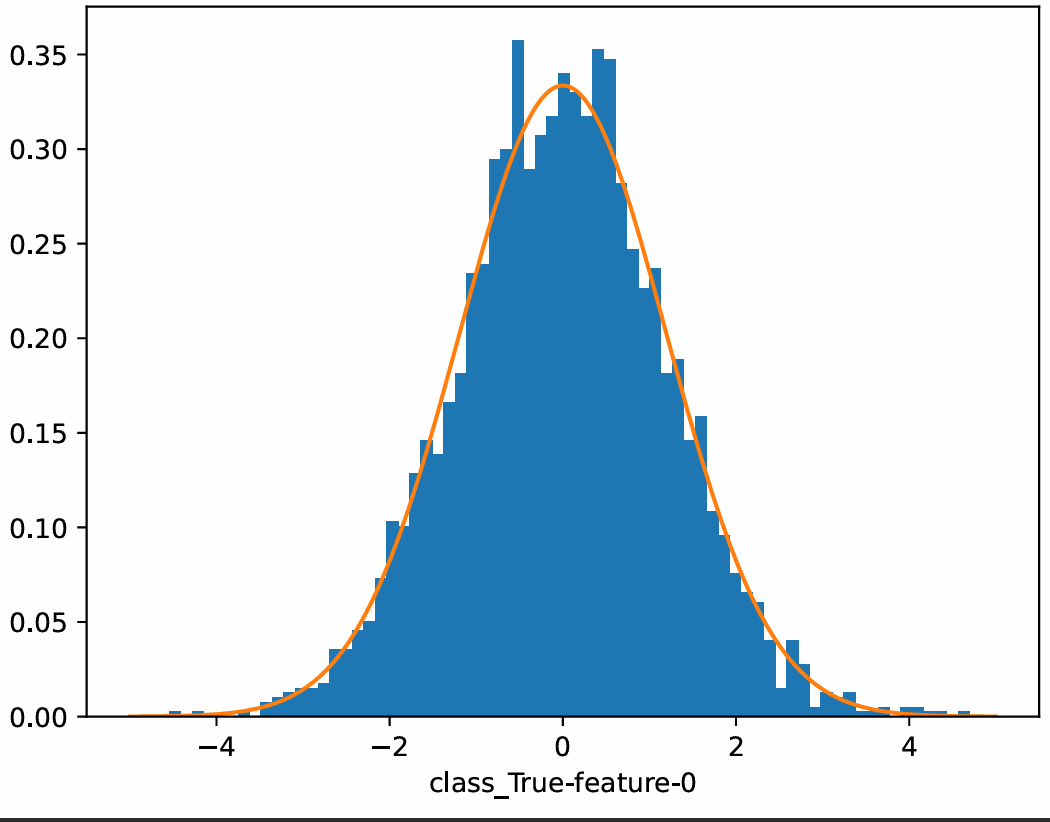
Descrizione generata automaticamenteImmagine che contiene schermata, diagramma, Diagramma

Descrizione generata automaticamente

We can observe that the first two features do seem to follow a gaussian distribution, just like the second two. The only difference between the two pairs is that the first one “overshoots” a little over the pdf, while the second one “undershoots” a little.

The last two features are clearly not compatible with a gaussian description.

## Class True

**Immagine che contiene schermata, diagramma, Diagramma

Descrizione generata automaticamente**

**Immagine che contiene schermata, diagramma, Diagramma

Descrizione generata automaticamenteImmagine che contiene schermata, diagramma, Diagramma

Descrizione generata automaticamente**

**Immagine che contiene schermata, Diagramma, diagramma

Descrizione generata automaticamenteImmagine che contiene schermata, Diagramma, testo, diagramma

Descrizione generata automaticamente**

This time the first four features fit the gaussian distribution well in the same fashion. The last two features don’t fit the gaussian distribution just like the whole dataset case, this time exhibiting two completely separate and almost equivalent peaks, mirrored in the origin.

**Class False**

Immagine che contiene schermata, Diagramma, diagramma

Descrizione generata automaticamente Immagine che contiene schermata, Diagramma, diagramma, testo

Descrizione generata automaticamente

Immagine che contiene schermata, diagramma, Diagramma, testo

Descrizione generata automaticamente Immagine che contiene schermata, Diagramma, diagramma

Descrizione generata automaticamente

Immagine che contiene schermata, Diagramma, diagramma, design

Descrizione generata automaticamente Immagine che contiene schermata, Diagramma, diagramma

Descrizione generata automaticamente

For this class, we have again a good fit for the first 4 features, but not for the last 2. What happens in this case is that we have a closer fit for these last two features, since we have three peaks of which the outer ones are quite smaller than the central one, but it’s clear that the associated pdf would not be the same as the gaussian plotted for reference.

**Conclusion**

Four out of six features (the first four features) fit a gaussian distribution quite well.