Machine Learning and Pattern Recognition Project Report

Paolo Magliano s314867

July 9, 2024

Abstract

This report describes the work done on Machine Learning and Pattern Recognition project. The report follows the structure of course laboratories, which studies and analyses the dataset and the models:

- Multivariate Gaussian Model
- Logistic Regression
- Support Vector Machines
- Gaussian Mixture Model

1   Lab 2: Dataset

The samples are computed by a feature extractor that summarizes high-level characteristics of a fingerprint image. The data is 6-dimensional and it consists of labeled samples corresponding to the genuine (True, label 1) class and the fake (False, label 0) class.

The dataset can be summarized by the image in Figures 1. It shows the data distribution of features and their correlation with each other.

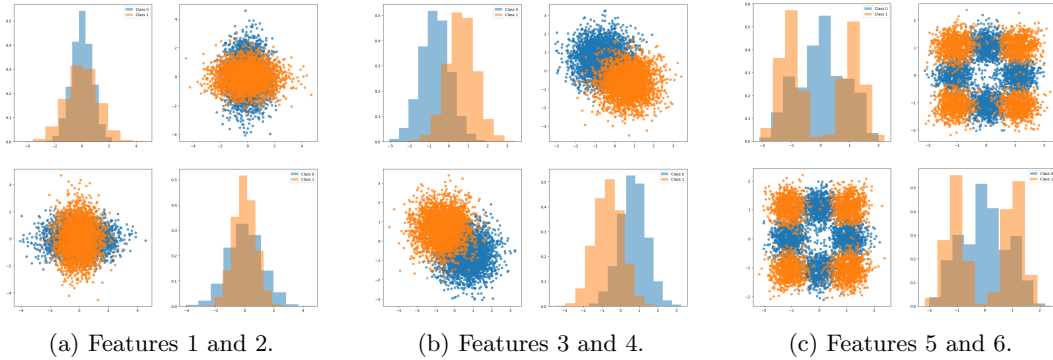(a) Features 1 and 2.   (b) Features 3 and 4.   (c) Features 5 and 6.

Figure 1: The figure shows histograms of the projected data on each feature. The scatter plot shows the correlation between the two features.

The first two features are mostly overlapping, so it is difficult to separate the two classes. The histograms show that they can be probably well approximated by a Gaussian density function. Further more means are very close to each other, but the variances are different.

The third and fourth features are more separated, but still overlapping. Also these features can be approximated by a Gaussian density function but the means are different and variance very similar. These features are more useful in a classification task as they are more discriminative.

The last two features are the most discriminative, they can't be approximated by a simple unimodal Gaussian density function because they seems to be bimodal or trimodal. Thanks to this behavior they form 4 distinct clusters (two for each feature) that can be easily separated. They are probably the most useful features if the classifier is able to capture the complex structure.

2 Lab 3: PCA and LDA

2.1 Dimensionality Reduction

Principal Component Analysis (PCA) is a dimensionality reduction technique that finds the directions of maximum variance in the data. It projects the data onto a lower-dimensional subspace while preserving as much variance as possible. The Figure 2 shows the results of PCA on the dataset. The first principal component captures most of the variance and it separates decently the two classes. The other principal components overlap, so they are not very useful for classification.
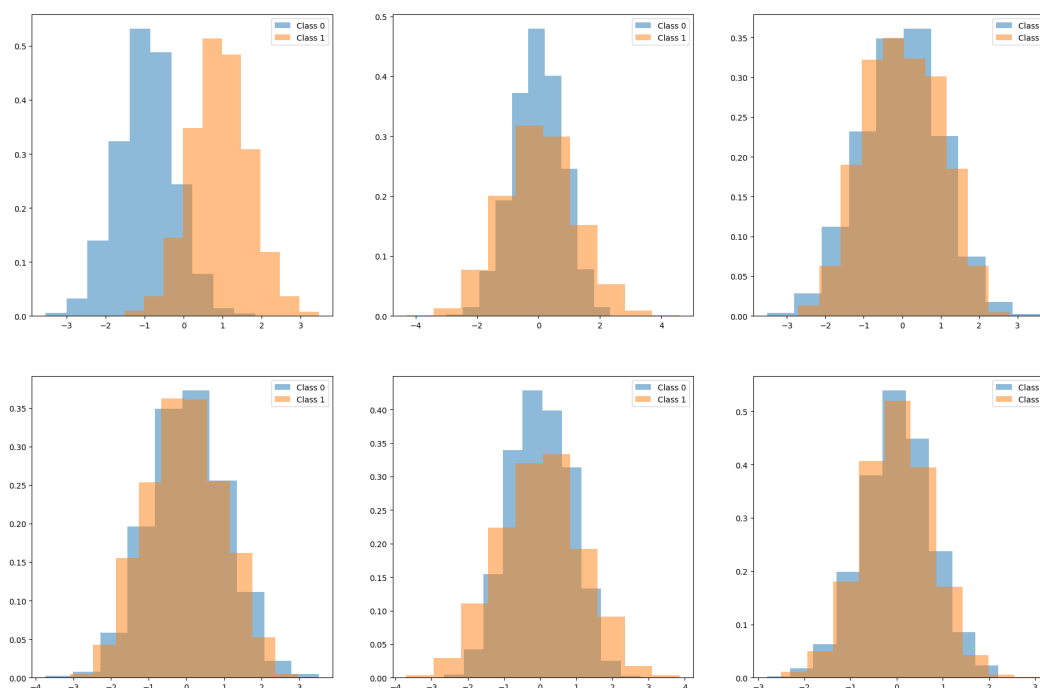

Figure 2: The figure shows the results of PCA on the dataset. The histograms show data distribution on each principal component. The top left one represents the first principal component, the bottom right is the last one.

Linear Discriminant Analysis (LDA) is a supervised dimensionality reduction technique that finds the directions that maximize the separation between classes. The Figure 3 shows the results of LDA on the dataset. Only one linear discriminant is found because the number of classes is 2. The linear discriminant separates the two classes partially. It is comparable to the first principal component of PCA.

Both cases simplify the dataset as much that the well-distinctive cluster of the last two features is lost. So, the preprocessing techniques cloud be useful to speed up the computation but they could also lose important information. In very simple dataset ad this one, with only 6 dimensions, presumably the best choice is to use the original dataset.
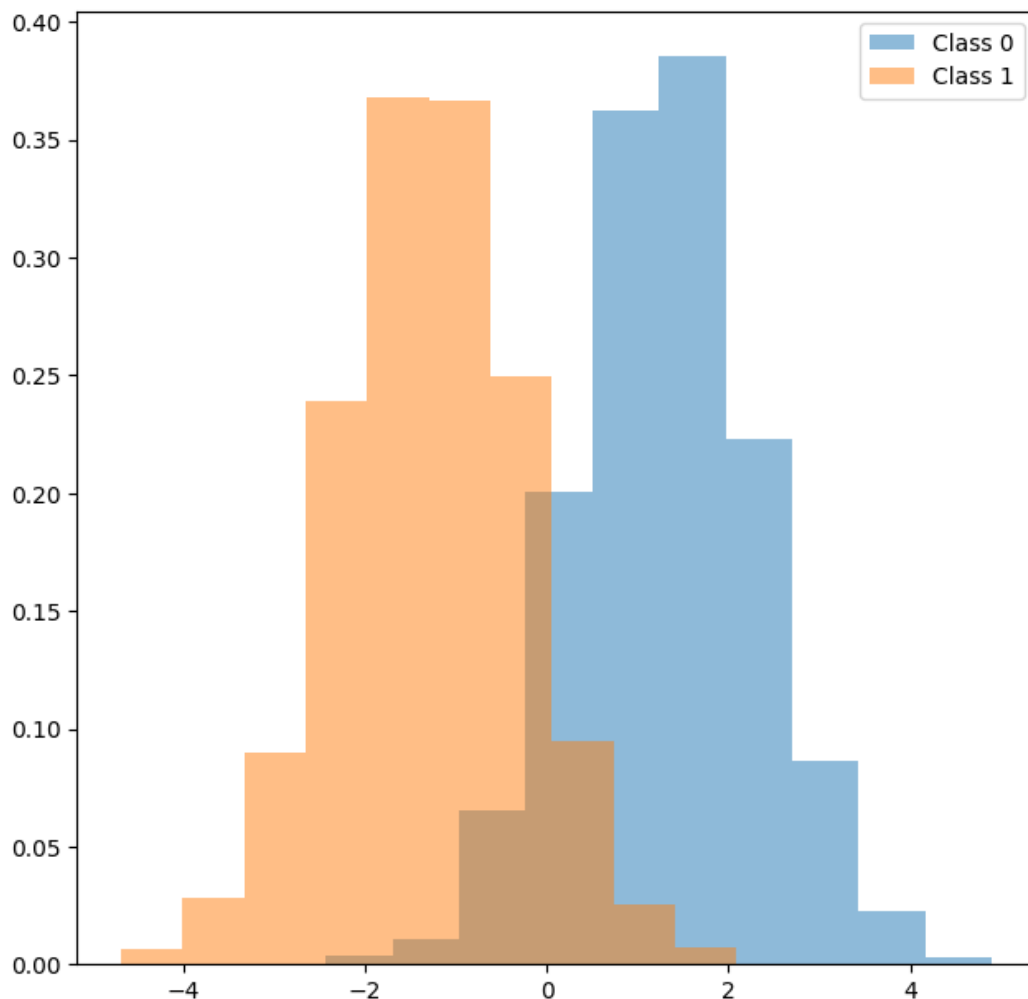
Figure 3: The figure shows the results of LDA on the dataset. The histogram shows data distribution on the only linear discriminant found.

2.2 Classification

The project analyses the performance of LDA method as a classifier. In order to get reliable results, the experiments' performance is evaluated with k-fold cross-validation with k equal to 10. It is also inspect the improvement of the classifier by using PCA as a preprocessing technique.

The Table 1 shows the performance results as error rate, Detection Cost Function (DCF) and Minimum DCF. The last one is useful to understand the gains of changing the threshold to optimal value.

As intuitively retrieved from PCA and LDA graphs, the capabilities of both methods are similar. In fact, the performance of the classifier is not improved by using PCA as a preprocessing technique.

In general the performance are terrible considering the simplicity of the dataset. Further more the threshold selection is not optimal, so the performance could be improved by tuning it a little bit.

3

PCA	Error Rate (%)	DCF	Min DCF
None	9.58	0.192	0.176
1 PC	9.68	0.194	0.177
2 PC	9.63	0.193	0.178
3 PC	9.43	0.189	0.176
4 PC	9.57	0.192	0.175
5 PC	9.58	0.192	0.175
6 PC	9.62	0.193	0.176

Table 1: Performance metrics for LDA models with various dimensionality reduction techniques.

3 Lab 4: Multivariate Gaussian density

As preliminary step, the project analyses the correspondence between the features distribution and the Gaussian density function. This is helpful to understand the potential efficacy of model based on Gaussian density. The figure 4 shows the histograms of the dataset and the Gaussian density function that best fits the data.
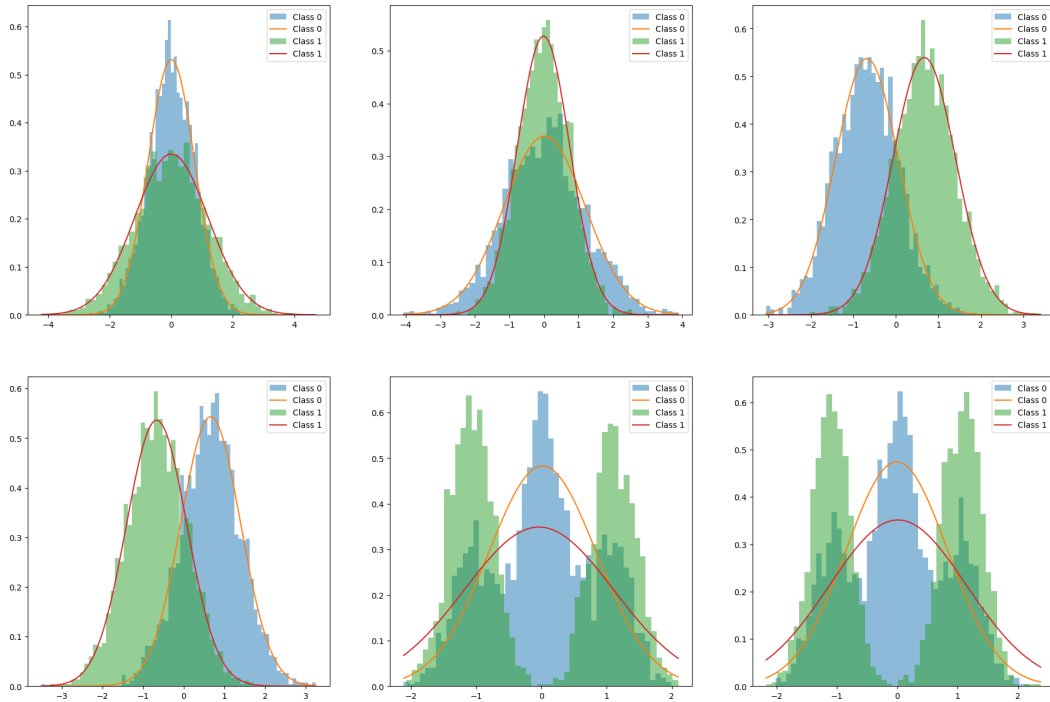

Figure 4: The figure shows the histograms of the dataset and the Gaussian density function that best fits each feature.

The results are consistent with the previous analysis. The first four features can be approximated by a unimodal Gaussian density function. Instead, the last two can not be well represented thanks to their multimodality.

4 Lab 5: Multivariate Gaussian Model

The project analyses the performance of Multivariate Gaussian (MVG) models with different covariance matrices: full, tied, and diagonal. The experiments are conducted with k-fold cross-validation with k equal to 10. The results are shown in Table 2.

Model	Error Rate (%)	DCF	Min DCF
MVG	7.45	0.15	0.131
TiedMVG	9.62	0.193	0.176
NaiveMVG	7.48	0.15	0.13

Table 2: Performance metrics for various MVG models without dimensionality reduction.

The results show that the tied MVG performs worse than the others. This is due to the fact that the covariance matrix change considerably along the features and so it can't be approximated by a single matrix. To visualize this better, the Figure 5 shows the estimated Gaussian density function for the tied MVG model.
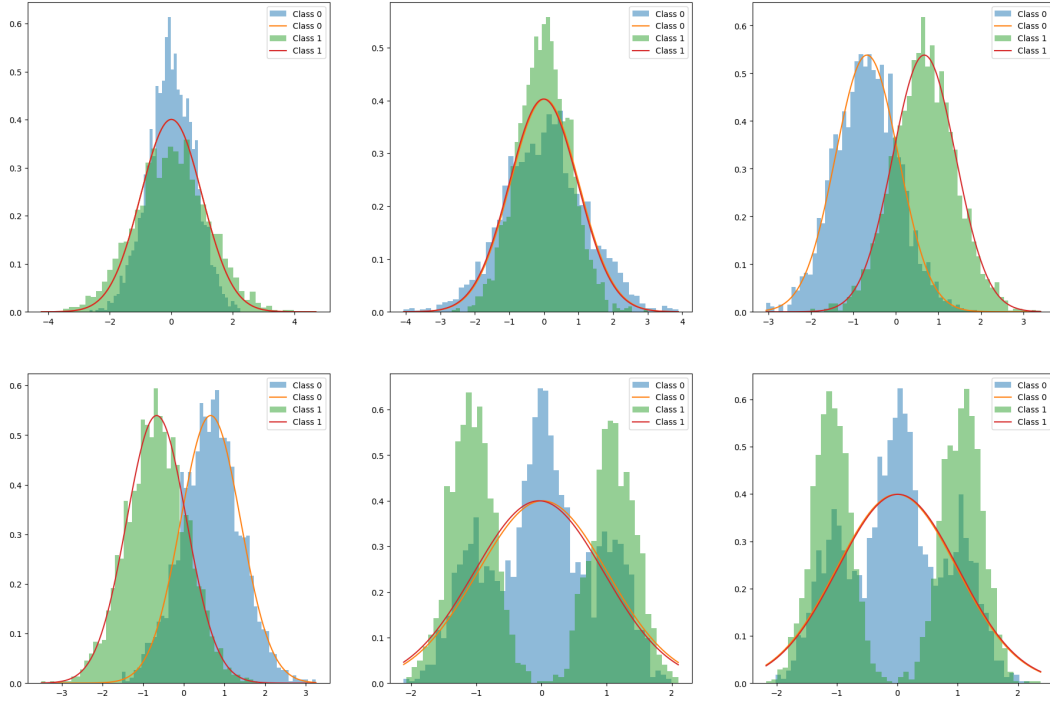

Figure 5: The figure shows the histograms of the dataset and the Gaussian density function with tied covariance matrix for each feature.

The graphs show that the tied MVG model can practically use only features 3 and 4 as discriminative ones. The other features are approximated with the same gaussian function fro both classes.

4.1 Correlation Analysis

In order to analyze further the performance of Naive MVG model respect to the full MVG model, the project computes the pearson correlation coefficient between the features for each class. The results are shown in Figure 6.
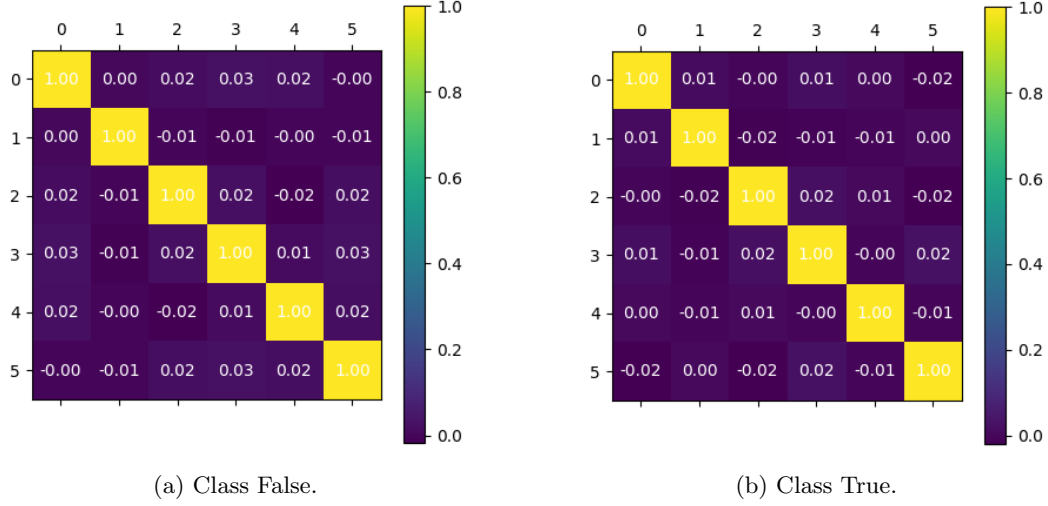
(a) Class False. (b) Class True.

Figure 6: The figure shows the Pearson correlation coefficient between the features for each class. Correlated features are close to 1 or -1.

The graphs show the reason why the Naive MVG model performs as well as the full MVG model. The features are not correlated at all, so the assumption of independence is not violated.

4.2 Assumption Analysis

The Gaussian model assumes that features can be jointly modeled by Gaussian distributions. The good-ness of the model is therefore strongly affected by the accuracy of this assumption. As already discussed, the last two features do not satisfy this assumption. To analyze if indeed the last set of features negatively affects our classifier, we can try repeating the classification using only feature 1 to 4. The results are shown in Table 3.

Model	Error Rate (%)	DCF	Min DCF
MVG	8.35	0.167	0.148
TiedMVG	9.73	0.195	0.175
NaiveMVG	8.35	0.167	0.149

Table 3: Performance metrics for various MVG models with only the first 4 features.

For both MVG and Naive MVG models, the performance is slightly worse than the full dataset. This implies that the last two features are useful anyway, even if they are not well approximated by a Gaussian.

4.3 Mean and Variance of Features Analysis

To further explore how means and variance of the features affect the performance of different approaches, the project repeats the classification using only features 1 and 2 and then only features 3 and 4. This is done because features 1 and 2 means are similar but variances are not, whereas for features 3 and 4 the two classes mainly differ for the feature mean, but show similar variance. The results are shown in Table 4 and Table 5.

Lets first analyze the generale performance between the two cases. The performance of the first two features is terrible, this is due to the fact that the two classes are not separable at all and this shows how important is the mean of class distribution. Instead,

6

Model	Error Rate (%)	DCF	Min DCF
MVG	36.83	0.738	0.454
TiedMVG	49.55	0.99	0.488
NaiveMVG	36.80	0.737	0.454

Table 4: Performance metrics for various MVG models with only the first 2 features.

Model	Error Rate (%)	DCF	Min DCF
MVG	9.67	0.194	0.175
TiedMVG	9.68	0.194	0.175
NaiveMVG	9.68	0.194	0.175

Table 5: Performance metrics for various MVG models with only the 3 and 4 features.

the performance of features 3 and 4 is very good, this is due to the fact that the two classes mean are very different and this remark how much these features are discriminative.

More over, in the first case Tied MVG model performs even worse because the features' variances are unsimilar. Instead in the second case, the performances are the same for all models because the features' variances are pretty the same.

4.4 Dimensionality reduction

Finally the project investigates the performance of MVG models with PCA as a preprocessing technique. The results are shown in Table 6.

Model	Metrics	PCA 1	PCA 2	PCA 3	PCA 4	PCA 5	PCA 6
MVG	Error	9.7	9.23	8.78	8.33	7.38	7.45
	DCF	0.195	0.185	0.176	0.167	0.148	0.15
	Min DCF	0.177	0.171	0.159	0.149	0.131	0.131
TiedMVG	Error Rate	9.68	9.63	9.45	9.58	9.57	9.62
	DCF	0.194	0.193	0.19	0.192	0.192	0.193
	Min DCF	0.177	0.178	0.176	0.175	0.175	0.176
NaiveMVG	Error Rate	9.7	9.32	9.3	9.02	8.58	8.6
	DCF	0.195	0.187	0.186	0.181	0.172	0.172
	Min DCF	0.177	0.171	0.162	0.159	0.154	0.154

Table 6: Performance metrics for various MVG models with PCA as a preprocessing technique.

The results confirm the hypothesis done in Dimensionality Reduction section. The performance of the classifier is not improved by using PCA as a preprocessing technique due to the lostness of the information. Indeed the best results are obtained with higher number of principal components. Tied MVG model is the only one that performs the same with and without PCA, this is due to PCA that provides features with similar variances.

4.5 Evaluation Metrics

Explain the metrics used to evaluate the performance of your models.

5   Experiments

Describe the experiments conducted to validate your models. Include details such as experimental setup, datasets used, and any specific configurations.

6   Results

Present the results of your experiments. Use tables, figures, and charts to illustrate the performance of your models.

Model	Accuracy	F1 Score
Model 1	0.90	0.88
Model 2	0.85	0.84
Model 3	0.92	0.90

Table 7: Model performance comparison

7   Discussion

Interpret the results. Discuss any patterns, insights, or anomalies observed. Compare the performance of different models and explain any differences.

8   Conclusion

Summarize the findings of your project. Discuss the implications of your results, any limitations of your study, and possible future work.