ML and PR Project Report – Part 4

The theme of the current laboratory experience is analyzing the results of **Linear** and **Quadratic** Classifiers.

# Classifying the whole dataset with MVG and variants

In the first part of this laboratory we analyze the usual fingerprints dataset in its entirety, by first splitting it in *training* and *validation* sets. We want to compare the results and performances of the following different Generative Models: Multivariate Gaussian, Naïve Bayes, Tied MVG, LDA.

The results are as follows:

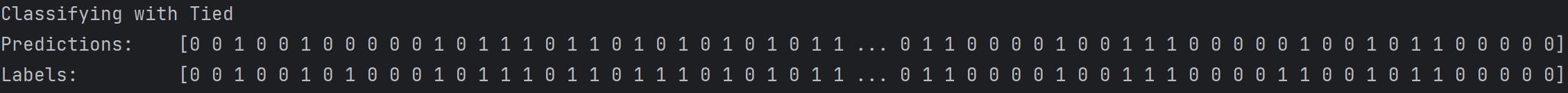
Immagine che contiene testo, Carattere, schermata, tipografia

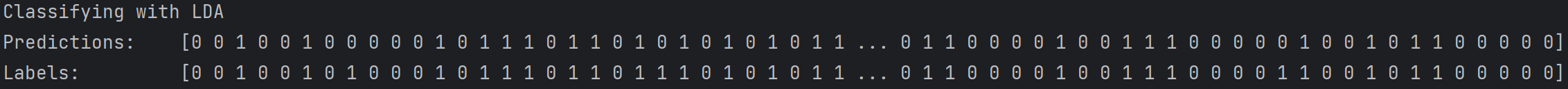
Descrizione generata automaticamente Immagine che contiene testo, Carattere, schermata, tipografia

Descrizione generata automaticamente

The models that performs better between the four is the standard Multivariate Gaussian. The fact that it outperforms the Tied version suggests that the tied hypothesis doesn’t hold true for the whole dataset.

We can also observe that the error rate for Tied and LDA is the same. Actually, the single predictions are also precisely the same, as one would expect, since the Linear Classifier that is the Tied MVG and LDA are actually one and the same.





Naïve Bayes’ performance is just slightly worse than that of MVG, suggesting that the features are uncorrelated for the most part. If we analyze the correlation matrices of the different classes, we can assess that the correlations are actually quite small.

Immagine che contiene testo, Carattere, schermata, numero

Descrizione generata automaticamente Immagine che contiene testo, Carattere, schermata, numero

Descrizione generata automaticamente

# Removing features that may not be Gaussian

In the previous laboratory we found out that the last two features do not really behave like Gaussians, especially in the case of the True class. For this reason we want to try and remove those features from the dataset and see if they contributed positively to the accuracy of the classifier, even though the gaussian hypothesis didn’t hold true for the most part.

Immagine che contiene testo, Carattere, schermata, tipografia

Descrizione generata automaticamente Immagine che contiene testo, Carattere, schermata, tipografia

Descrizione generata automaticamente

It turns out that they provided a positive contribution in all cases, meaning that their distributions’ behavior could be somehow approximated as the behavior of a gaussian.

# When the Tied MVG works

With the last laboratory in mind once again, we may remember that features 0 and 1 presented the same mean but different variances, while features 2 and 3 had the opposite behavior.

Having classes with same covariance was the necessary hypothesis for Tied MVG to work, meaning that we could expect a dataset comprising only of features 2 and 3 to behave much better under Tied MVG than one comprising only of features 0 and 1.

Immagine che contiene testo, Carattere, schermata, tipografia

Descrizione generata automaticamente Immagine che contiene testo, Carattere, schermata, tipografia

Descrizione generata automaticamente

features 0-1 features 2-3

The last hypothesis holds true. Also, since Tied MVG is a simplification of standard MVG, the latter also yields good results. The results are this good also probably thanks to the well distanced means, which features 0 and 1 did not have.

The features 0-1 dataset performs quite poorly with both classifiers, with the Tied alternative providing a coinflip result basically, and the standard MVG performing poorly.

As we anticipated, the problem probably arose from the excessive overlap of samples of the two classes, which made cluster division quite challenging.

The two classes shared the same mean, meaning that the likelihoods of the samples would be similar regardless of the class, and on top of that we also provided the same covariance, as one does when using Tied MVG, making that likelihood virtually the same. It’s not weird at all that the Tied MVG performs inadequately.

# Preprocessing the dataset with PCA

Let’s see how performance varies with PCA with different **m** values.

Immagine che contiene testo, schermata, Carattere, menu

Descrizione generata automaticamente Immagine che contiene testo, schermata, Carattere

Descrizione generata automaticamente Immagine che contiene testo, schermata, Carattere, menu

Descrizione generata automaticamente

PCA preprocessing never improves the classification performance and matches that of the non-pre-processed classifier only for MVG at 6 dimensions, which is the original number of features (Tied is also always the same, at whichever dimension).

Standard MVG performs better with more dimensions, hitting a minimum of 7% error rate when we don’t remove any direction (the result is the same if we did not apply PCA at all). This could be linked to the fact that MVG without additional hypotheses can get rid of its overfitting defect mainly through getting more information – in this case the whole feature space.

The Naïve model doesn’t seem to care that much about the number of dimensions: if fluctuates between 9.3% and 8.8%. Naïve Bayes is actually used when not that much data is available, so in light of this consideration the results seem to make sense.

There’s a weird result for Naïve Bayes though: the performance of Naïve Bayes with 6 dimensional PCA is 8.90% error rate, while the result without PCA at all (so still 6 features) is much lower, at 7.20%. Could the PCA projections have created direction that are correlated to one another to a higher degree than the original directions? Or if that’s not possible, did PCA modify the dataset in some other way that proved detrimental to the classification performance? (I will attach the results for no PCA again, for easier comparison)

Immagine che contiene testo, Carattere, schermata, tipografia

Descrizione generata automaticamente Immagine che contiene testo, Carattere, schermata, tipografia

Descrizione generata automaticamente

As for the Tied model, the results do not change almost at all, being always the worst of the three alternatives, at 9.25-9.30% error rate.

All in all MVG is always the best of the bunch, while Tied always the worst. At lower dimensions though MVG and Naïve Bayes performance are quite similar.

# Conclusion

Our dataset performs best on MVG, while yielding similar results with Naïve, hinting at an uncorrelation between the features, which is confirmed when analyzing the correlation matrices. The Tied classifier has worse results, although not tragically bad, and is actually quite good for features 2-3, since they share the same covariance and have non-overlapping classes.

PCA preprocessing worsens the performance at all values of **m**. Just at **m = 6** we get the same results for MVG, while Naïve is still worse than with no preprocessing.