Session 4 project

Group 203

Participants:

- Nikolaos Gkloumpos
- Malthe H. Boelskift
- Louis Ildal
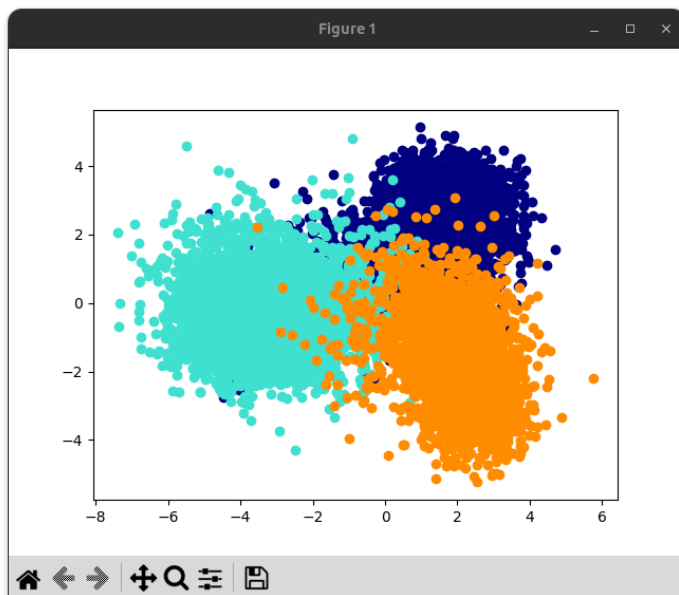- Guillermo V. Gutierrez-Bea

Overall description

Excersise on dimention reduction with both PCA and LDA methods, and comparison of the generated groups.

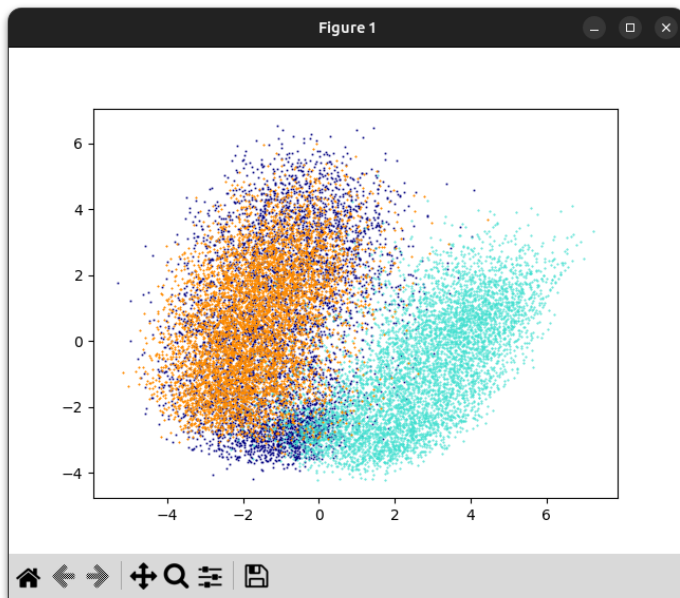### About the scripts

Part 1

To process the data we dimmentionally reduce them both with PCA and LDA to compare the results after classification.

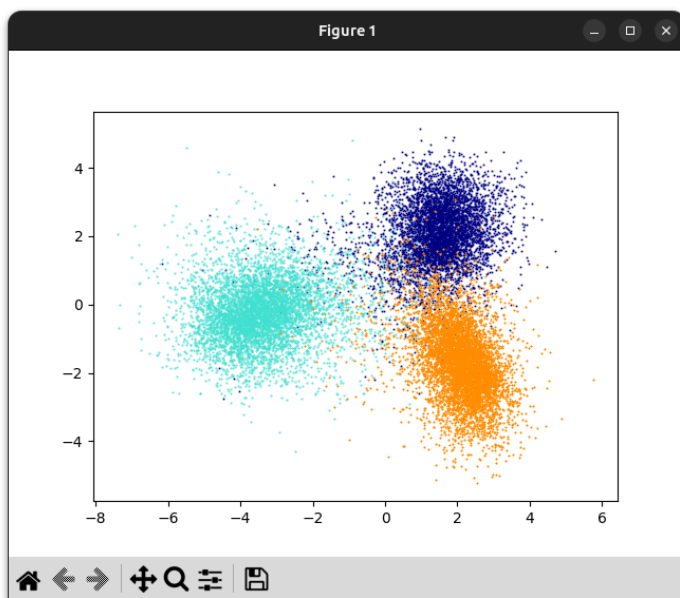A scatter plot was generated to help visualize the data.

Bellow we can see how different dimensional reductions affect the data when represented on 2 axis.

PCA reduction

LDA reduction

Part 2

Here we aim to make a bayesian classifier that is able to classify data into the three clases 5, 6 and 8. But this time we use dimensinally reduced data from the datasets using two methods (PCA and LDA) and comparing them. The dimentions of both the training set and dataset are reduced to 2. First, we use the training set to find the statistics of the two features of the three classes, we compute the likelihood and prior and finally the posterior probability of each class so we can perform classification taking the highest probability.

```
 The classifications for the test dataset using lda are [6 5 5 ... 8 8 8].
 The classifications for the test dataset using pca are [5 5 8 ... 5 8 8].
 The accuracy (pca) is 0.7103399433427762.
 The accuracy (lda) is 0.9444050991501416.
```

As expected lda gives a better performance since it tries to keep the distance between clases after the dimenstion reduction as much as it can.