**MACHINE LEARNING FROM DATA**

**Report: Lab Session 1 – MAP and Gaussian data**

**Classification criteria based on maximizing posterior probability**

**Names:**

**Instructions**

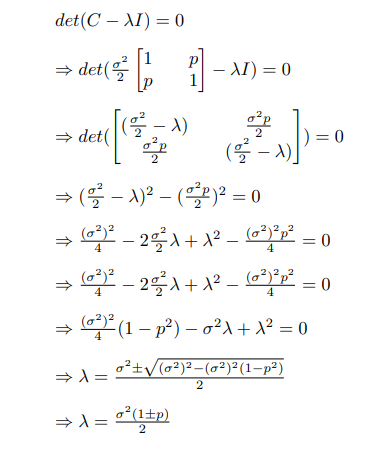
Handling your work:

* Answer the questions in this document with the name **Mlearn\_Lab1\_report\_team\_surnames.doc**

**Questions**

Q1: Derive the expression for the eigenvalues of the matrix as a function of the parameters and . (edit equations or solve by hand and scan and insert an image with the solution)





Q2. Create a table including error probabilities obtained by the linear classifier (LC) and error probabilities obtained by the quadratic classifier (QC), for each SNR value on the test set. Discuss the results.

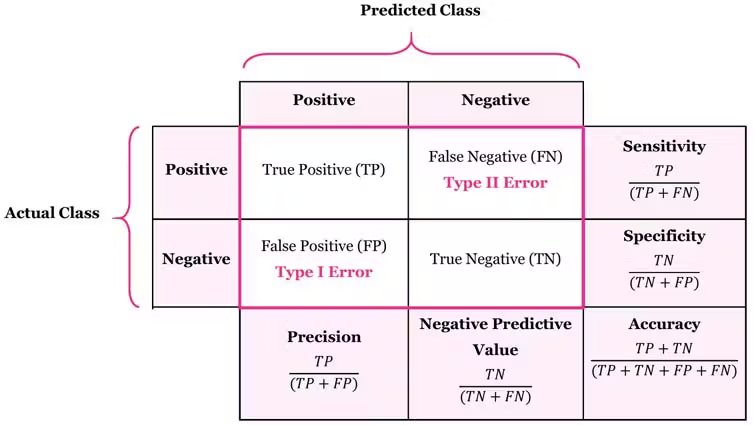
|  | 3 dB | 0 dB | -3 dB | -10 dB |
| --- | --- | --- | --- | --- |
| LC | 0.0095 | 0.0425 | 0.111 | 0.289 |
| QC | 0.01 | 0.0425 | 0.111 | 0.2925 |

|  | 3 dB | 0 dB | -3 dB | -10 dB |
| --- | --- | --- | --- | --- |
| Sigma | 0.041 | … | … | 0.8333333… |

For a two class classification, we find that as SNR is reduced the variance augments, in these case the data is more distributed therefore it is possible for there to be overlapping. Due to this -10db probability error is a lot worse than 3.

Q3. Include in the report the confusion matrices obtained for SNR=-10db and SNR=-3dB and the two classifiers on the test set. Discuss the results.

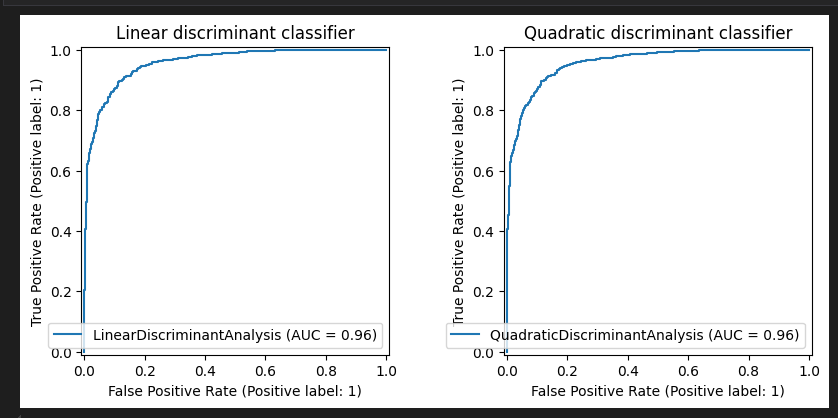
|  | -3 dB | -10 dB |
| --- | --- | --- |
| LC | [[881 119]  [103 897]] | [[708 292]  [286 714]] |
| QC | [[881 119]  [103 897]] | [[702 298]  [287 713]] |

(Sensitivity == Recall)

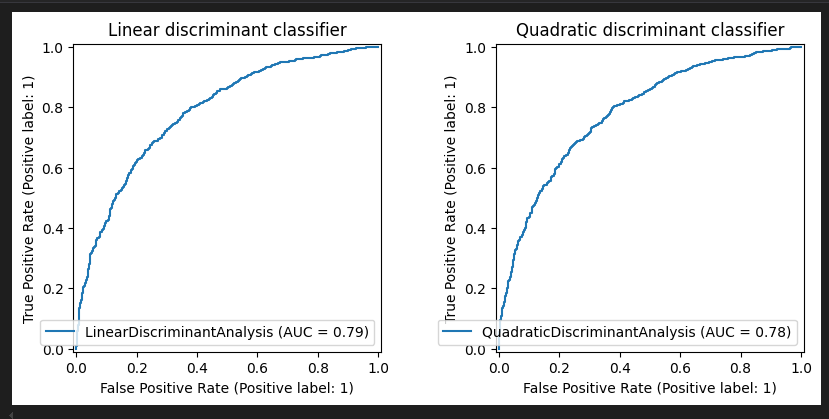
We can see that since there is more separability within classes we have better sensitivity, accuracy and error for the -3dB compared to -10db.

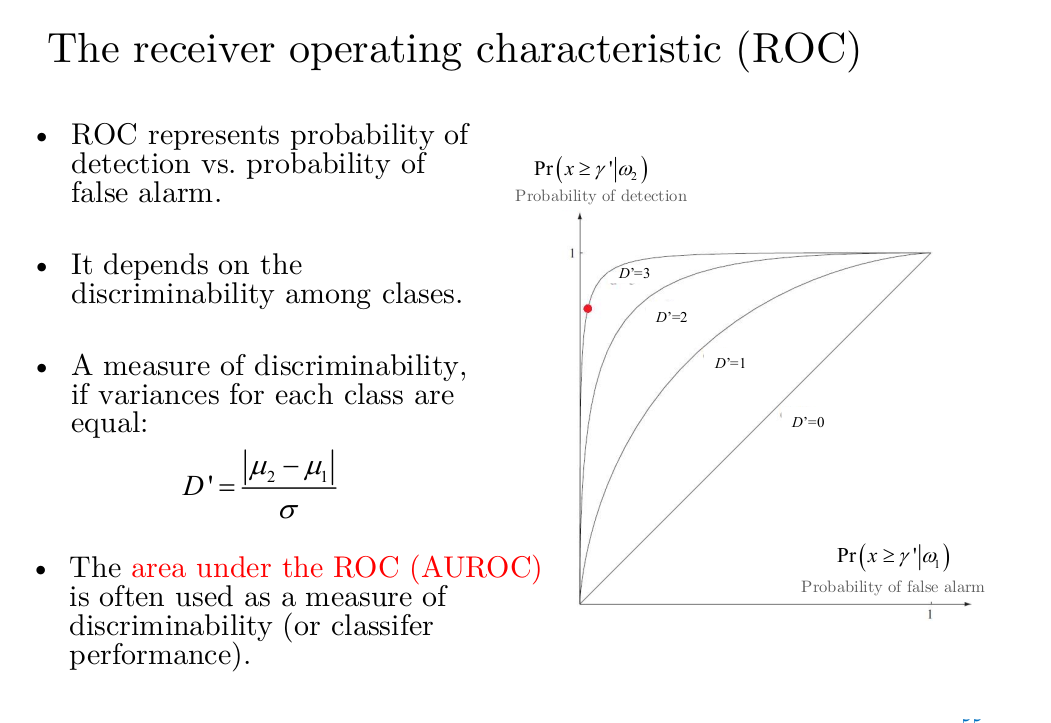
Q4. Include in the report the ROC curves obtained for SNR=-10db and SNR=-3dB and the two classifiers on the test set. Discuss the results.

* -3 db



* -10 db

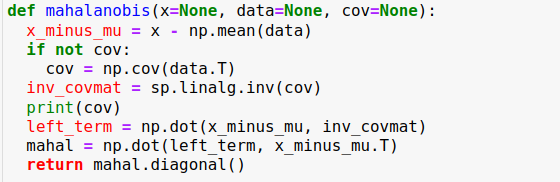




Since means do not vary, what varies is sigma, as shown, sigma increases as SNR decreases, and the closer it is to 0 the flatter it will be. This is why -10 is a lot flatter than -3, because it has a greater sigma value.

Q5. Compute the Mahalanobis distance between the two classes on the test set for SNR= 3, 0, -3,-10 dB. Compare the results. Explain why the result differs depending on the order of the parameters.

|  | 3 dB | 0 dB | -3 dB | -10 dB |
| --- | --- | --- | --- | --- |
| 01 | 26.46 | 14.79 | 8.92 | 4.16 |
| 10 | 26.82 | 15.05 | 9.13 | 4.34 |



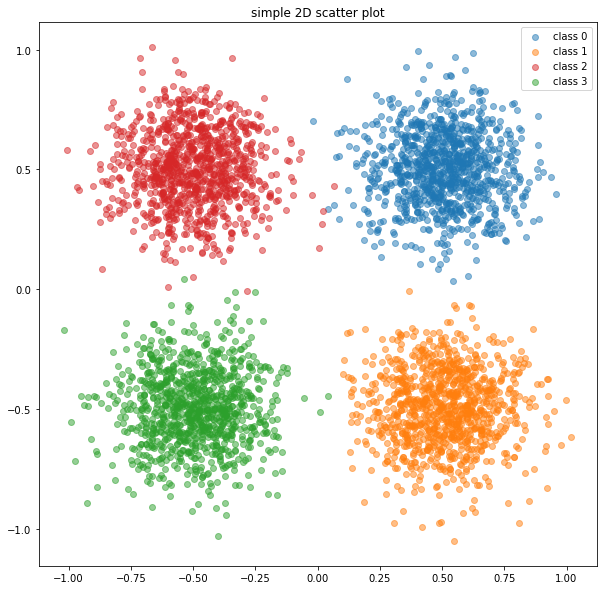
As we said before, as we decrease the value of SNR the variance increases. This means that the data is more distributed and thus we find more overlapping, resulting in a decrease of the Mahalanobis distance between the data. Since the way we calculate the Mahalanobis distance is not symmetric, the results differ depending on the order.

**QPSK and covariances of all classes identical but arbitrary (case 2)**

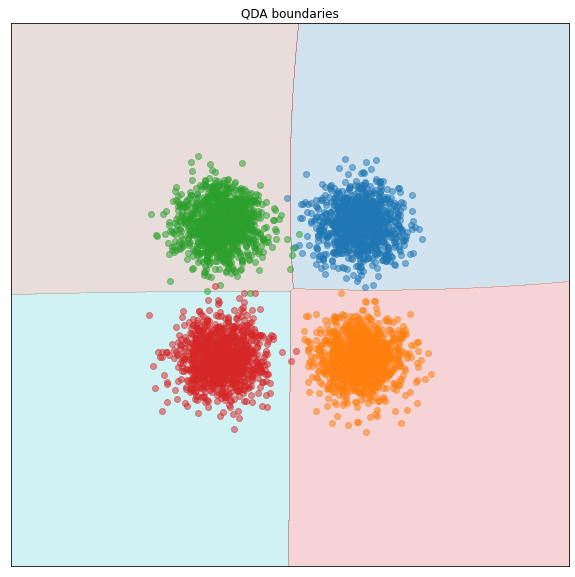
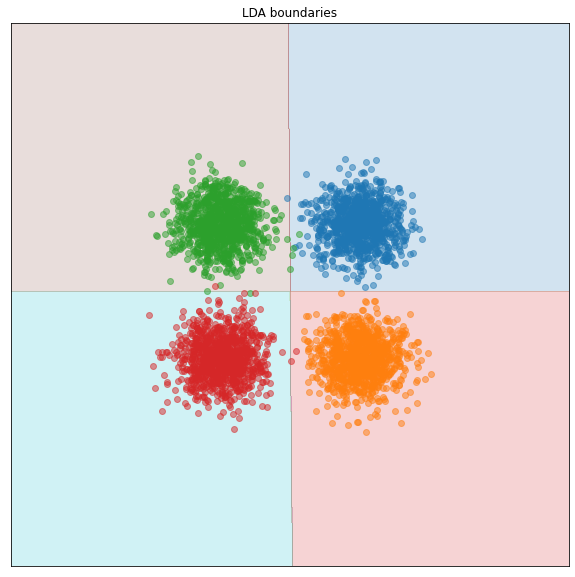
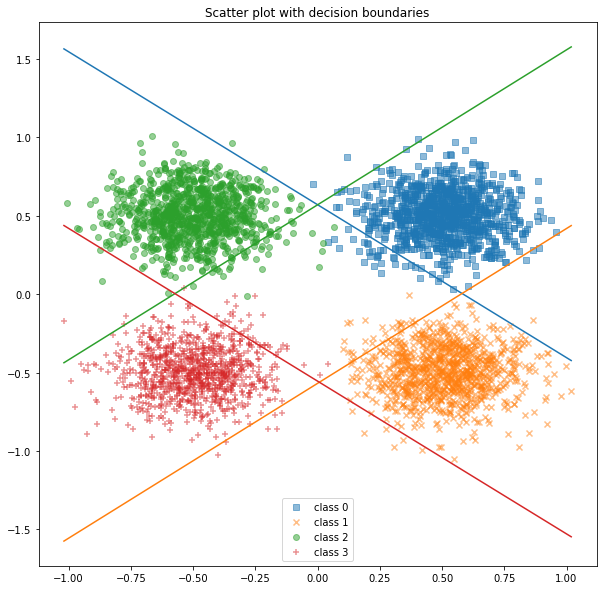
Q6. Include the scatter plot, decision boundary, confusion matrices and error probabilities obtained using the linear classifier (LC) and the quadratic classifier (QC) for *ρ* = 0. Compare the metrics for the two classifiers and discuss the results.

https://scikit-learn.org/stable/auto\_examples/classification/plot\_lda\_qda.html

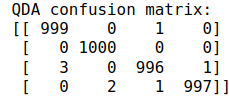
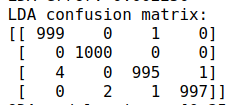
**Scatter Plot**

****

**Decision Boundary**



**Confusion Matrixes**

****

**Error Probabilities**

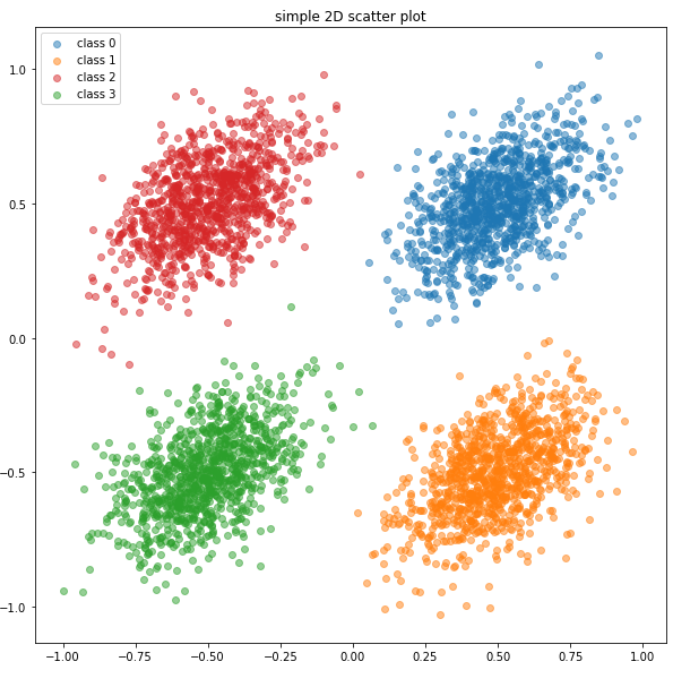
LDA error: 0.002250

QDA error: 0.002000

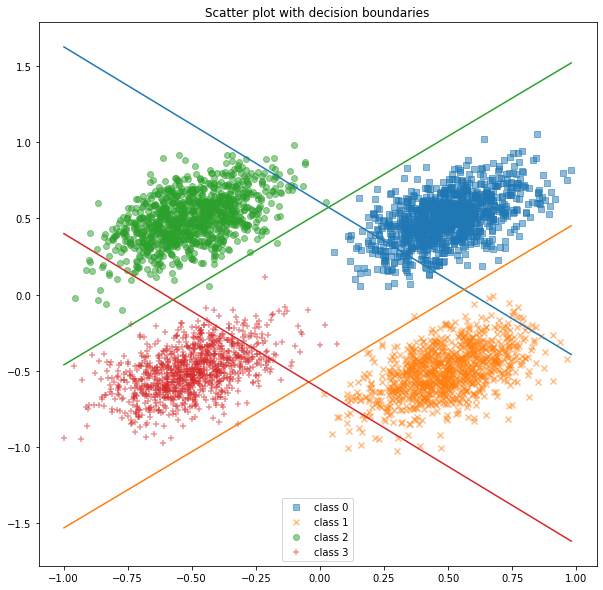
Since p=0 we know that there will be no correlation between features, hence the circular shape of the distributions. Since there isn’t much overlap between classes due to the high SNR value both classifiers seem to work very well.

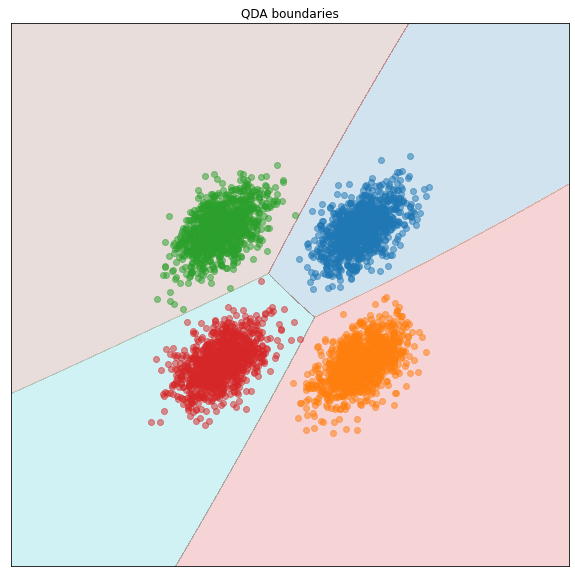
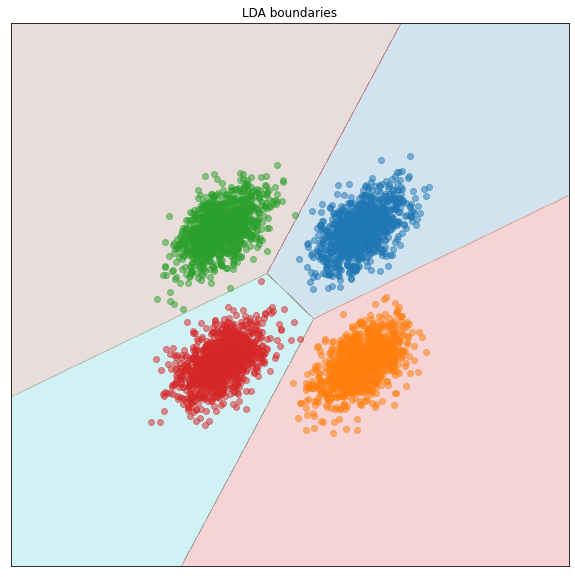
Q7. Repeat the previous analysis (Q6) for *ρ* = 0,5. Compare the metrics for the two classifiers and discuss the results.

**Scatter**

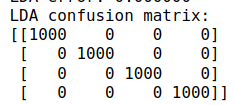


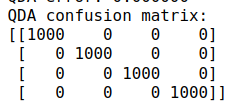
**Decision Boundary**





**Confusion Matrixes**





Errors:

0.0000 for both cases

We can observe that with p=0.5 there is a correlation between features, giving the scatter plots a more ellipsoidal shape. This results in different looking boundaries.

Q8. Compare and discuss the results obtained in Q6 and Q7

As mentioned, we can see that having a higher correlation between classes results in a more ellipsoidal shape, whereas with no correlation we have circles.

PREGUNTAR – CORRELATION?

**QPSK and different covariance matrices (case 3)**

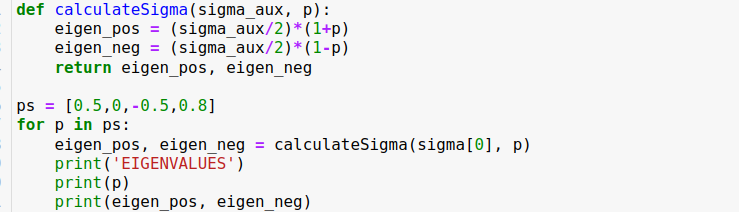
Q9. Include the error probabilities obtained using the linear classifier (LC) and the quadratic classifier (QC) for SNR = +5 dB and +10 dB. Compare the metrics for the two classifiers and discuss the results.

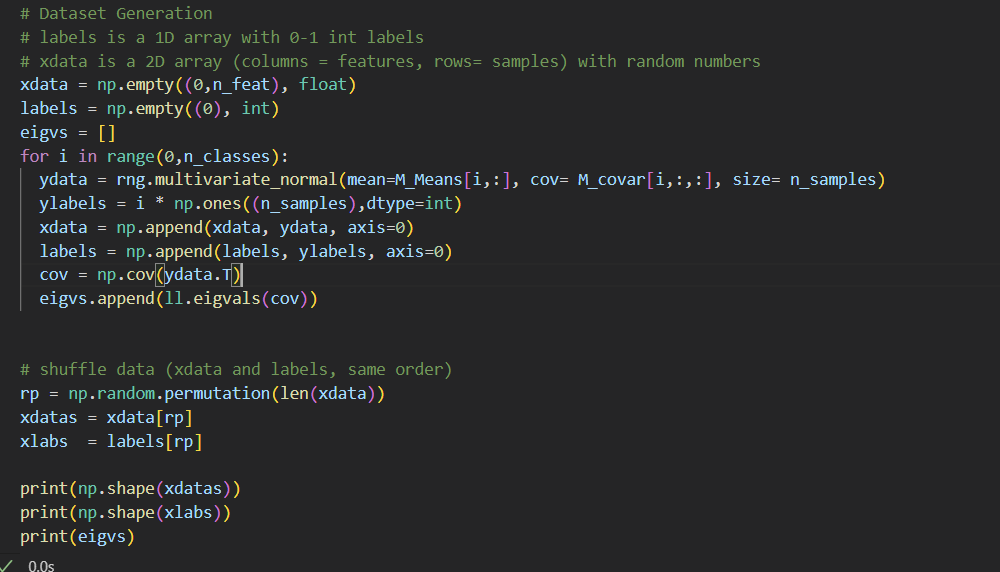
|  | 5 dB | 10 dB |
| --- | --- | --- |
| LC | 0.062750 | 0.001500 |
| QC | 0.059000 | 0.000500 |

We can see a major difference between the quadratic and the linear, the covariance matrices are all different so the hyper-quadratic decision boundaries are a better fit for our model.

Q10. Complete the table with the theoretical eigenvalues using the formula obtained when answering Q1, and the eigenvalues computed using the **sample** data covariance matrices. Add the code to compute the eigenvalues of each covariance matrix; use scipy.linalg.eigvals (for just one SNR value)

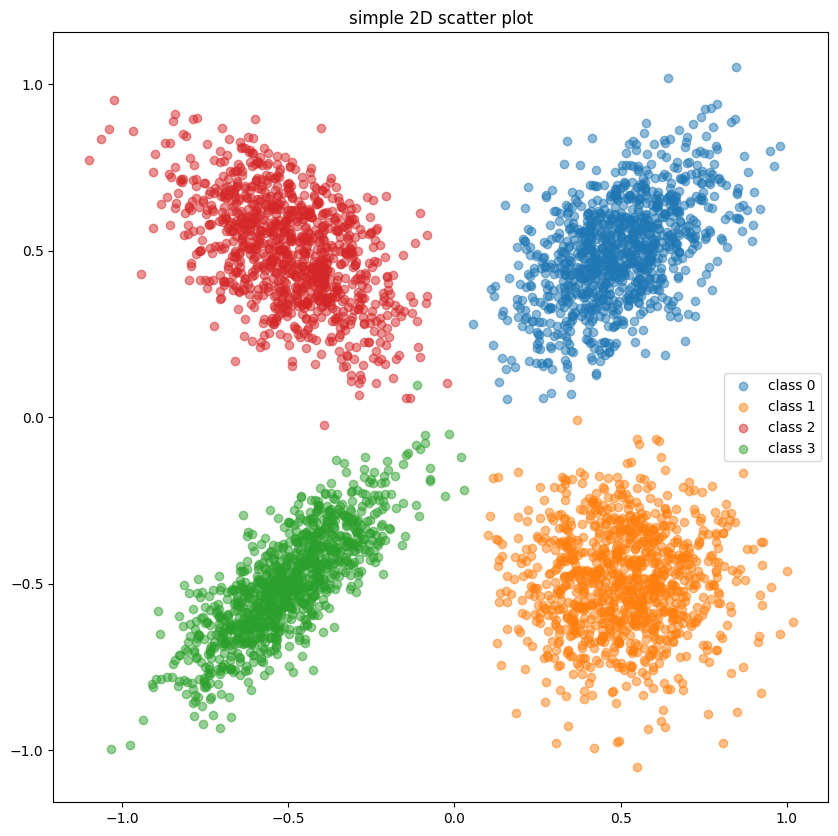
| SNR=5 | Class 1 = 0.5 | Class 2=0 | Class 3=-0.5 | Class 4=0.8 |
| --- | --- | --- | --- | --- |
| Theoretical eigenvalues | 0.11858541225631422 0.03952847075210474 | 0.07905694150420949 0.07905694150420949 | 0.03952847075210474 0.11858541225631422 | 0.14230249470757708 0.015811388300841892 |
| Eigenvalues from sample covariance matrices | 0.12316786, 0.04012092 | 0.08756795, 0.0760404 | 0.03804094, 0.11019306 | 0.12880184, 0.01609249 |



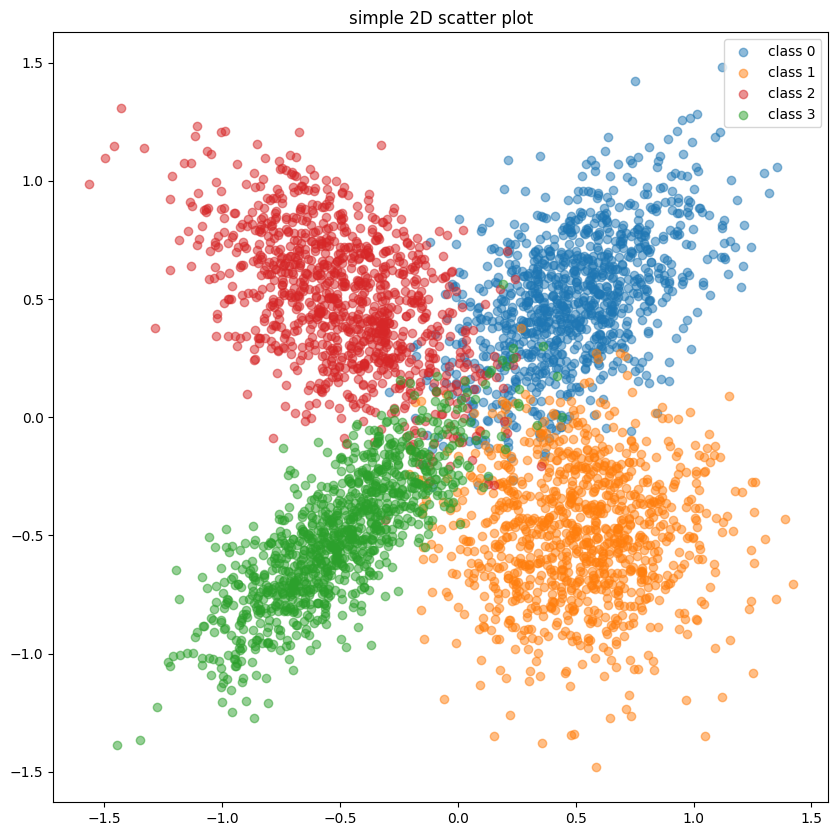


Q11. Include scatter plots for the linear and quadratic classifiers using SNR= +5 dB and SNR= +10 dB. Relate the shape of the clusters with the eigenvalues of the covariance matrices.

* SNR = 10



* SNR = 5

**Eigenvalues** (λ\lambdaλ): These values indicate the amount of variance that can be explained by each principal component (direction in which the data varies).

The orange class or class 1 has a p=0 (the one with the spherical shape) where the eigenvalues are the same.

For class 0 or the blue class (p=0.5) or class 2 (red, p=-0.5) the eigenvalues are the same but transposed, and the same happens in the graph. That’s why they look mirrored.

Finally, for class 3 or the green class, the difference between the eigenvalues is the greatest, which is why the ellipse looks elongated.

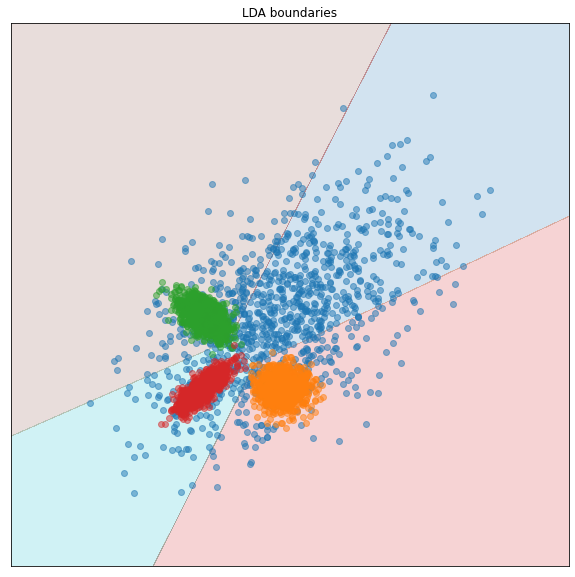
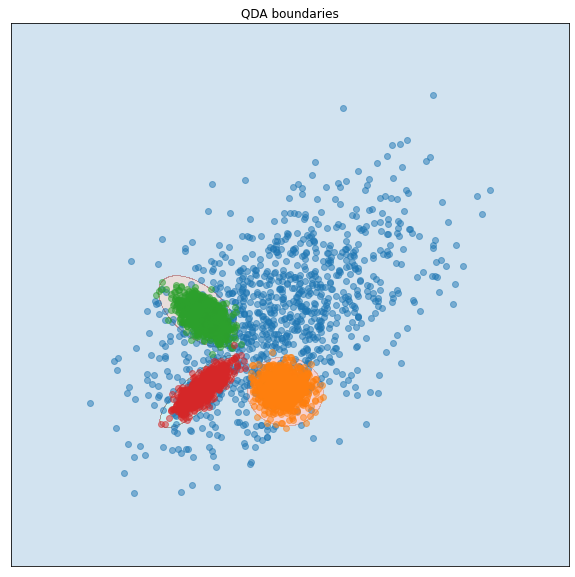
Q12. For SNR = 10 dB, multiply the covariance matrix of class 1 by a large number (for example sigma(0)\*=30). Compute the classification error, scatter plots and boundaries for the linear and the quadratic classifiers. Observe that in this case the quadratic discriminant outperforms the linear one.

Include error probabilities, scatter plots and decision boundaries. Compare the performance of the classifier and justify the results. Include in your answer the new value of sigma[0].

**SIGMA:**

[1.5 0.05 0.05 0.05]

|  | LDA | QDA |
| --- | --- | --- |
| 10db | 0.127000 | 0.048750 |



As we can see the QDA’s decision boundaries are hyper-quadratic instead of only hyper-planes, that is why in the QDA we have hyper-spheres for class 1,2 and 3 and for LDA we have hyper-planes and the QDA outperforms LDA. This phenomenon is produced by having a 30 times the original variance (sigma[0] = sigma[0] \* 30)

**Questions:**

* Does correlation between classes improve the model? Ex: p0 vs p5 reduces error to 0. Why is this happening?
* Mahalanobis distance