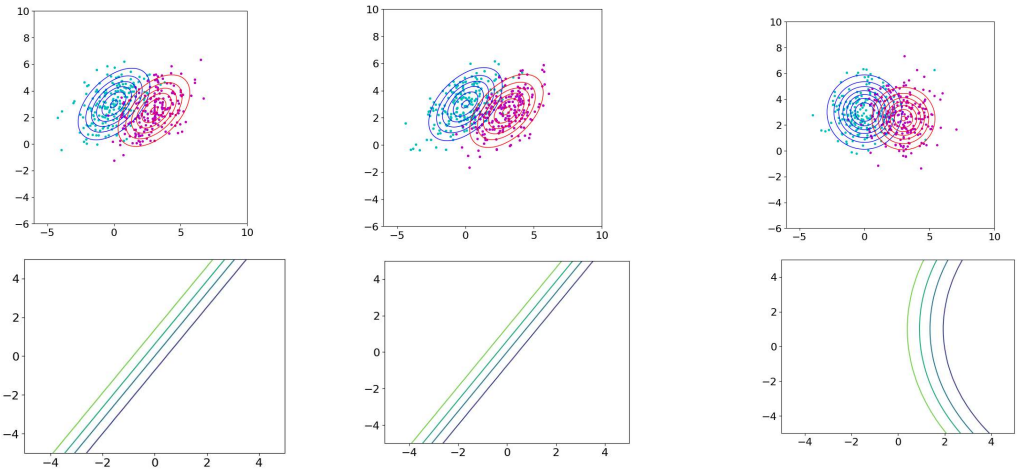
CO544 – Machine Learning
E/19/432 – Wickramaarachchi U.I.
Lab 03

Q1.

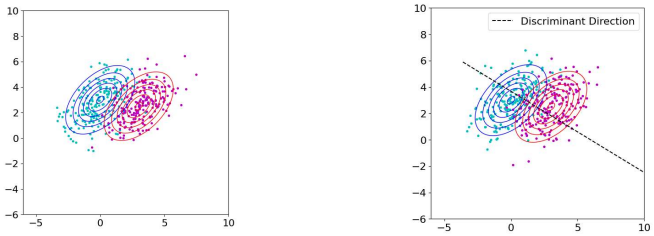


Discuss (in your report) if what you plot is consistent with your expectation from analytical derivation of the class boundaries.

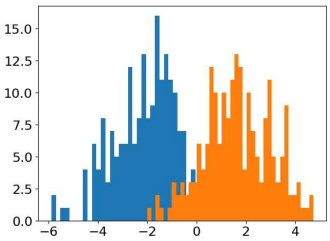
Data for each class is normally distributed with the same covariance matrix but different means. The separation between the clusters indicates that the class boundaries have been correctly derived and plotted.

Q2.

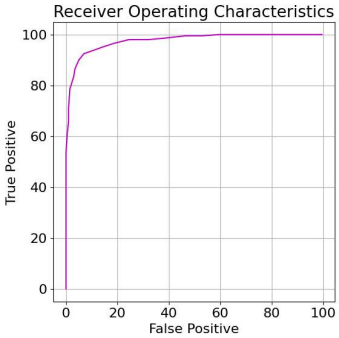
1. 2. 3.



4.



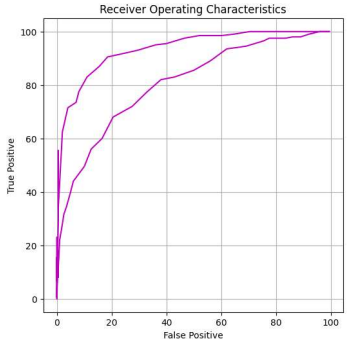
5. Pmin: -5.888194010837295
   Pmax: 4.735107920317232



6.
Area under the ROC curve: 9698.0

7.
Pmin: -5.888194010837295
Pmax: 4.735107920317232
Maximum accuracy of 0.93 achieved at threshold -0.25

8.



Area under the ROC for random curve: 8003.0
Area under the ROC for direction connecting the means curve: 9293.875

Explain what the precise statistical interpretation of AUC is and when it is used.
Since it is more accurate when the AUC is larger from above the direction connecting the means of the two classes is precise statistical interpretation of AUC than a random curve.
The AUC is a commonly used metric for binary classification problems because it provides a single number summary of the model's performance across all thresholds, and it's robust against imbalanced classes.

Q3.

Euclidean distances: 0.9928680732332427, 0.8914272573298286
Mahalanobis distances: 0.9813805302861112, 0.6975605462003125

Give a clear and succinct description of the differences.

While Euclidean distance is a distance metric for vectors in a Euclidean space, Mahalanobis distance is a measure of how many standard deviations away a point is from the mean of a multivariate distribution.

Euclidean distance doesn't take into account the correlation between variables, while Mahalanobis distance does. This makes Mahalanobis distance more useful for multivariate data, as it can balance the scales of the different variables.

In this case, the Mahalanobis distances are smaller than the Euclidean distances. This is because the data points are closer to the mean of the distribution when taking into account the covariance of the data.

All files can be found in : https://github.com/uthsaraiw/Machine-Learning-Labs/tree/main/ML_Lab03

Colab Links: https://colab.research.google.com/drive/1f0mZgF0CdsMrd-pnehL_VVhx3PM0mrTe?usp=sharing