CO544 – Lab 3

E/19/166

Jayathunga W.W.K.

1. **Class Boundaries:** In the feature space, these are the decision boundaries dividing several classes. **Posterior Probabilities:** Based on the observation of the feature, these probabilities indicate the chance that a data point will belong to a specific class.
The likelihood contours, which are the lines we've drawn below, and the predicted contours, which are what we anticipated seeing, cross at the points where both classes have equal chances. This is what Figures 01 and 03 show us. The expected outlines in Figure 02 are more in line with class 1. This is due to our assumption that class 1 would appear 70% of the time and class 2 just 30% of the time. Thus, the figure is consistent with our expectations derived from our computations. Densities and contours of probabilities on the posterior probability P[ω | x]

Case 1: Equal Covariance Matrices, Equal Prior Probabilities In this case, the decision boundary would be linear due to the equality of covariance matrices and prior probabilities. The scatter plot and contours would show a clear linear separation between the two classes. The posterior probability contours would also reflect this linear separation. This is consistent with the analytical derivation of the class boundaries for this case.

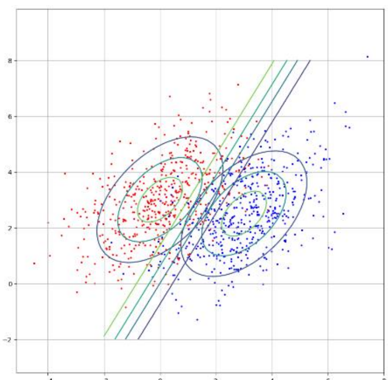


Figure 1:Probability densities and contours on posterior probability

Case 2: Equal Covariance Matrices, Different Prior Probabilities Here, the decision boundary would still be linear due to the equality of covariance matrices. However, the scatter plot and contours would show a shift in the decision boundary towards the class with the lower prior probability. This is because a lower prior probability makes it harder for a data point to be classified into that class. The posterior probability contours would also reflect this shift. This is consistent with the analytical derivation of the class boundaries for this case.

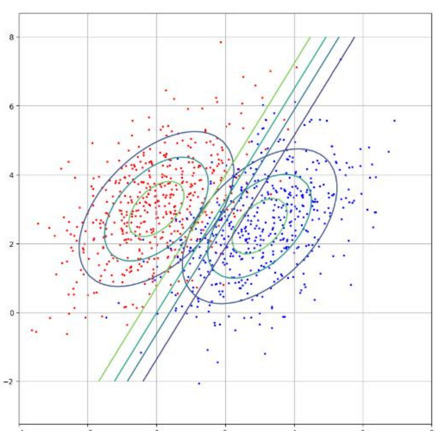


Figure 2:Probability densities and contours on posterior probability

Case 3: Different Covariance Matrices, Equal Prior Probabilities In this case, the decision boundary would be quadratic due to the difference in covariance matrices. The scatter plot and contours would show a curved separation between the two classes. The posterior probability contours would also reflect this curved separation. This is consistent with the analytical derivation of the class boundaries for this case.

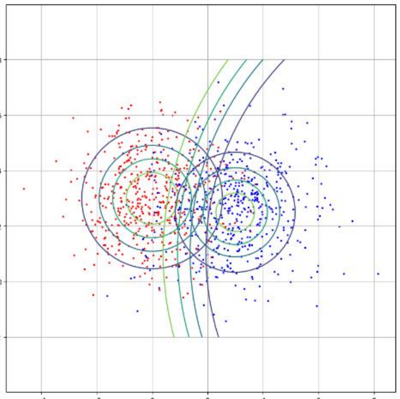


Figure 3:Probability densities and contours on posterior probability

2. Fisher's Linear Discriminant Analysis (LDA): This technique finds a linear combination of characteristics that distinguishes or characterizes two or more classes of objects or events. It is applied in machine learning, statistics, and pattern recognition. The resultant combination can be employed as a dimensionality reduction step before further classification, or more frequently, as a linear classifier.

$$w_F = (C_1 + C_2)^{-1}(m_1 - m_2)$$

$$(C_1 + C_2)^{-1} = (2C_1)^{-1} = \begin{bmatrix} 1/3 & -1/6 \\ -1/6 & 1/3 \end{bmatrix}$$

For this distribution,

$$w_F = \begin{bmatrix} 1.083 \\ -0.667 \end{bmatrix}$$

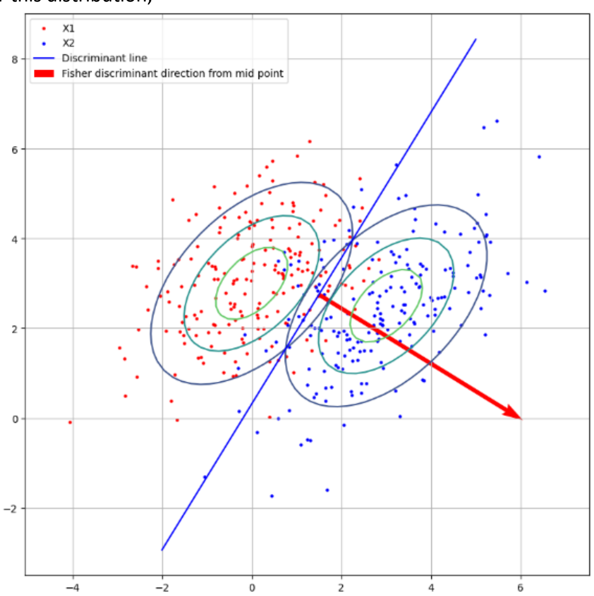


Figure 4:Fisher's Discriminant Line and Direction

The Receiver Operating Characteristic (ROC) Curve is a graphic that illustrates how well a binary classifier performs at various thresholds. The true positive rate (TPR) and false positive rate (FPR) are plotted against each other. The better the model, the closer the curve is to the upper left corner and the closer the Area Under the Curve (AUC) is to 1.

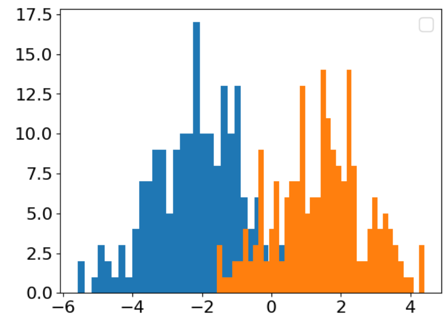


Figure 5:Histograms of projections of the two classes onto the Fisher discriminant direction

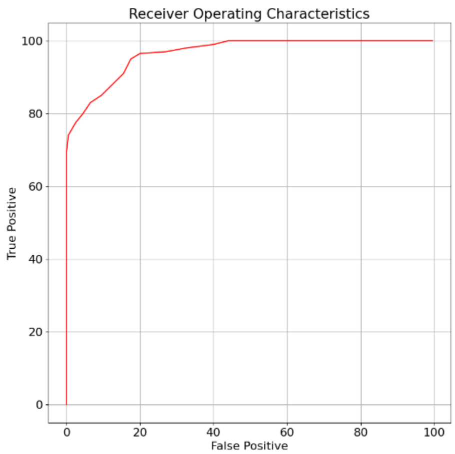
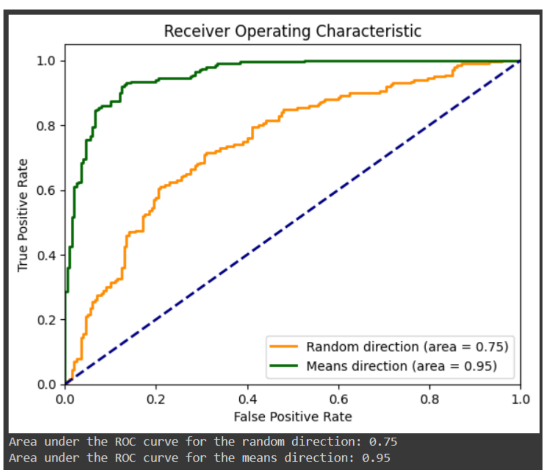


Figure 6: The ROC Cueve

Area under the curve = 9613.5

Classification Accuracy = 88.99999999999999



The Area Under the Curve (AUC) is a performance metric used in binary classification problems. It is a measure of a classifier's ability to distinguish between positive and negative classes.

The AUC is the area under the Receiver Operating Characteristic (ROC) curve, which plots the True Positive Rate (TPR) against the False Positive Rate (FPR) at various threshold settings. The AUC ranges from 0 to 1, where:

- An AUC of 1 indicates a perfect classifier; the model has a high true positive rate and a low false positive rate for all thresholds.
- An AUC of 0.5 suggests that the classifier is no better than random guessing.

- An AUC of 0 indicates a completely incorrect classifier; the model has a low true positive rate and a high false positive rate for all thresholds.

The AUC is used when you want to compare the performance of different models, or the same model with different parameters. It is especially useful when the classes are imbalanced, or when the costs of false positives and false negatives are very different. The AUC is independent of the threshold set for classification because it considers all possible thresholds. In terms of statistical interpretation, the AUC can be thought of as the probability that a randomly chosen positive instance will be ranked higher than a randomly chosen negative instance by the classifier. This interpretation is often used in the context of ranking and recommendation systems.

While the AUC is a useful metric, it should not be used in isolation. Other metrics such as precision, recall, F1 score, and confusion matrix should also be considered for a comprehensive evaluation of a model's performance.

3.

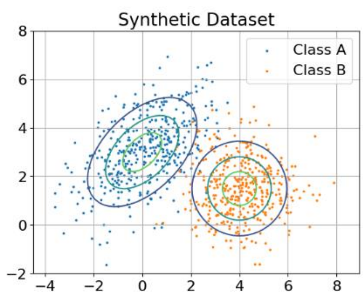


Figure 7:Distribution of the Sample Dataset

In a classification problem, both the distance-to-mean classifier and the Mahalanobis distance-to-mean classifier are used to classify data points into two classes based on their distances to the mean of each class. However, they calculate these distances differently and make different assumptions about the data, leading to different classification results.Distance-to-Mean Classifier The distance-to-mean classifier calculates the Euclidean distance from a data point to the mean of each class. It assumes that all features contribute equally to the distance calculation, which is equivalent to assuming that the features are uncorrelated and have the same variance. This classifier works well when these assumptions hold true. However, it can perform poorly when the features are correlated or have different variances, as it does not take these factors into account.Mahalanobis Distance-to-Mean Classifier The Mahalanobis distance-to-mean classifier calculates the Mahalanobis distance from a data point to the mean of each class. Unlike the distance-to-mean classifier, it takes into account the correlations between features and their variances. The Mahalanobis distance is essentially a scaled Euclidean distance, where the scaling is done according to the covariance matrix of the data. This classifier can therefore handle correlated features and features with different variances, making it more versatile than the distance-to-mean classifier.In summary, while both classifiers are used for the same purpose, they differ in how they calculate distances and the assumptions they make about the data. The distance-to-mean classifier is simpler but less flexible, while the Mahalanobis distance-to-mean classifier is more complex but can handle a wider range of data characteristics. The choice between the two depends on the specific characteristics of the data at hand.

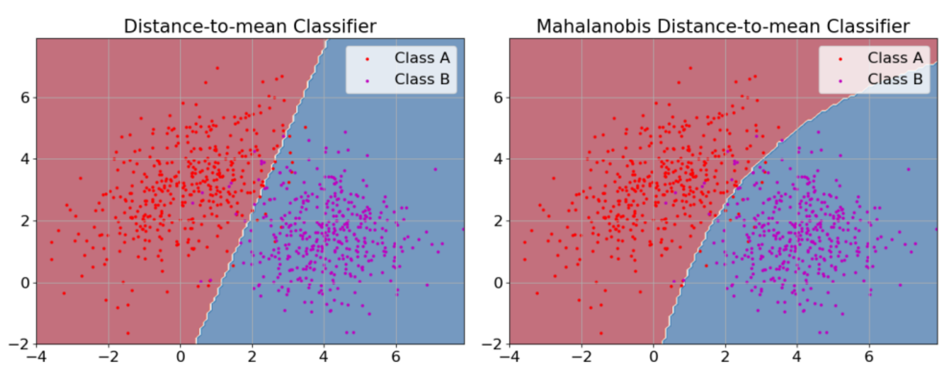


Figure 8:Illustration of the difference between a distance-to-mean classifier and Mahalanobis distance-to-mean classifier