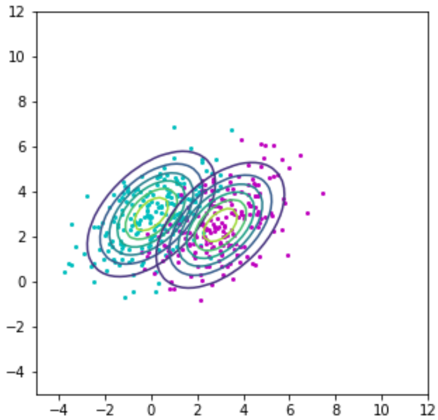
Foundations of Machine Learning Lab 3 Report

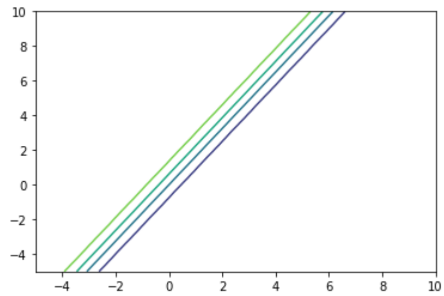
Abhilash Pulickal Scaria
aps1n21@soton.ac.uk
Student Id:33124639

Observation

1)

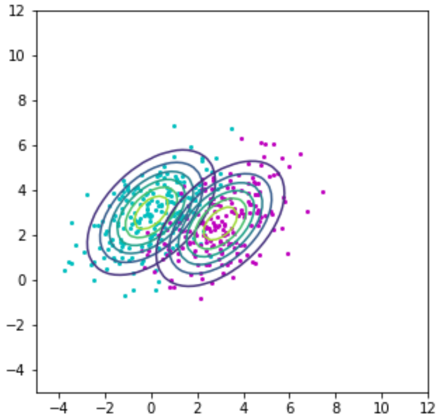


(a) Contour and scatter plot for the distributions

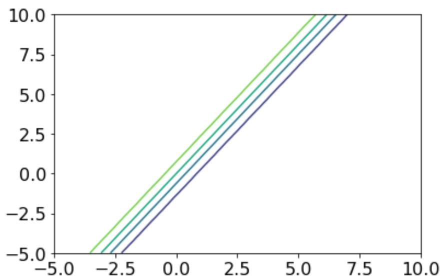


(b) Contour plot for posterior probability for class 1(m1) with prior probability P1 = P2 = 0.5

Figure 1: Plots for the distributions with mean m1 = $\begin{pmatrix} 0 \\ 3 \end{pmatrix}$, m2 = $\begin{pmatrix} 3 \\ 2.5 \end{pmatrix}$ and covariance $\begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}$

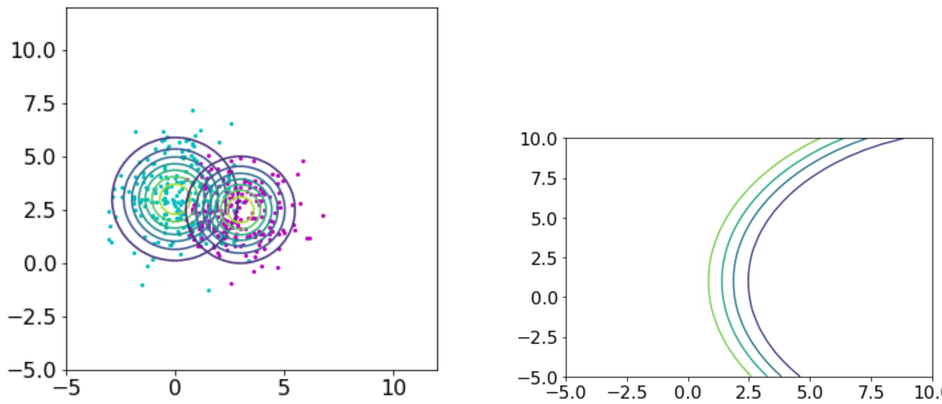


(a) Contour and scatter plot for the distributions



(b) Contour plot for posterior probability for class 1(m1) with prior probability P1 = 0.7, P2 = 0.3

Figure 2: Plots for the distributions with mean m1 = $\begin{pmatrix} 0 \\ 3 \end{pmatrix}$, m2 = $\begin{pmatrix} 3 \\ 2.5 \end{pmatrix}$ and covariance $\begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}$



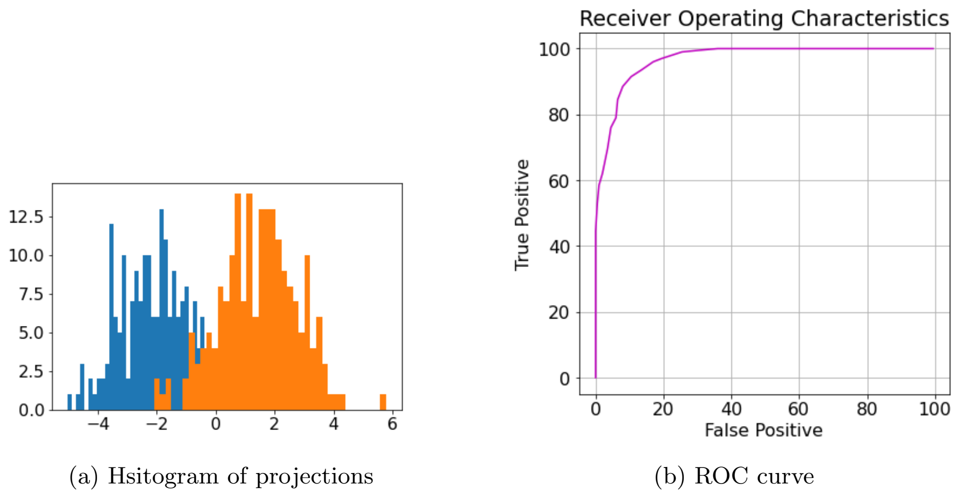
(a) Contour and scatter plot for the distributions

(b) Contour plot for posterior probability for class 1(m1) with prior probability P1 = 0.5, P2 = 0.5

Figure 3: Plots for the distributions with mean m1 = $\begin{pmatrix} 0 \\ 3 \end{pmatrix}$, m2 = $\begin{pmatrix} 3 \\ 2.5 \end{pmatrix}$ and covariance C1 = $\begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}$, C2 = $\begin{pmatrix} 1.5 & 0 \\ 0 & 1.5 \end{pmatrix}$

From Figure 1 and 2 when we consider points from each class and cross check it with the posterior probability it can be seen that the plot is consistent with our expectation. Consider the point (0,4), from posterior probability contour it is observed that the point has higher probability to be in class1 by crosschecking with the fig(a) we can see that it belongs to class1. Similarly for figure 3 we can check with point (0,2.5), it can be seen that the posterior probability contour suggests the point to be in class1 and from figure 3(a) it can be observed that the point belongs to class1.

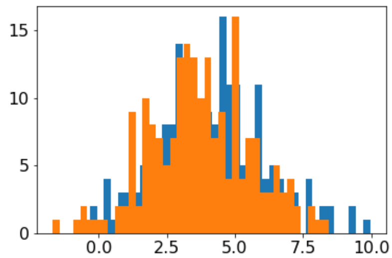
2) Using the code snippet provided we can compute the Fischer linear discriminant and project the data in this direction. Fischer linear discriminant gives the direction in which there is maximum seperation between the projected means of the 2 distributions and minimum variance wihtiin individual distributions. Taking a decision threshold value of -1.08 we obtain a classification accuracy of 84.5%. Using the next code snippet we can compute the ROC curve



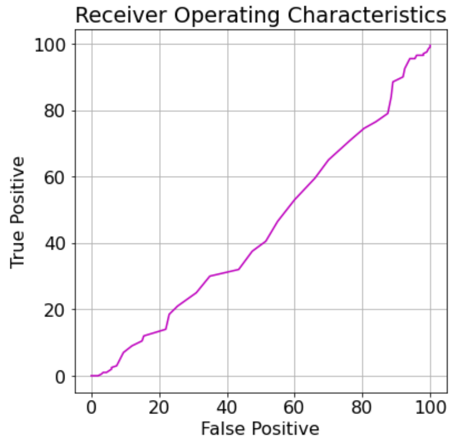
(a) Hsitogram of projections

(b) ROC curve

Figure 4: Histogram and ROC curve for projection of data in Fischer Discriminant direction

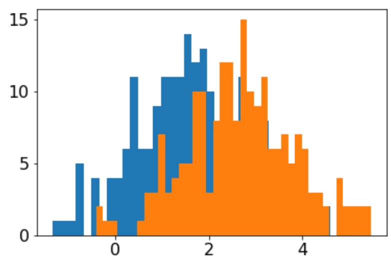


(a) Hsitogram of projections

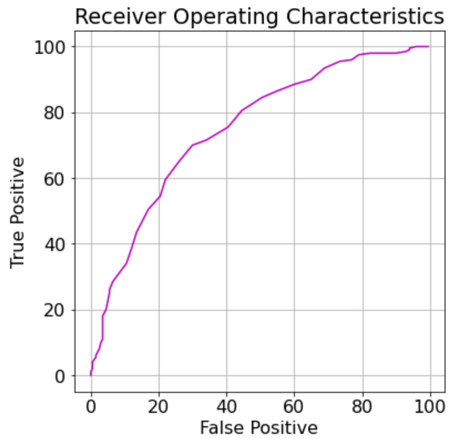


(b) ROC curve

Figure 5: Histogram and ROC curve for projection of data in direction connceting means of two classes(means are added)



(a) Hsitogram of projections



(b) ROC curve

Figure 6: Histogram and ROC curve for projection of data in random direction

The Area under curve(AUC) for figure 4, figure 5 and 6 are 9711.5, 4433.75 and 7461.87 respectively. It can be seen that the ROC curve for projection in Fischer discriminant direction has the largest AUC. This implies that the best seperation(classification) of the two classes happens when it is projected in Fischer discriminant direction. The AUC value is used to compare different ROC curves. Models with higher AUC values seperates the classes better with less missclassification.

3) Consider two distributions with m1 = $\begin{pmatrix} 1 \\ 2 \end{pmatrix}$, m2 = $\begin{pmatrix} -3 \\ 2 \end{pmatrix}$ and covariance C1 = $\begin{pmatrix} 0.5 & 0 \\ 0 & 0.5 \end{pmatrix}$, C2 = $\begin{pmatrix} 3 & 0 \\ 0 & 3 \end{pmatrix}$. From the scatterplot(Figure 7) it can be observed that the class 2 is more spread out(bigger variance). Consider a point (0, 7) if we use Euclidean distance to check which class this point belongs to then we get distance to class 1 as 5.09 and distance to class 2 as 5.83. The distance to mean classifier suggest the point should be in class 1 but from figure 7 we can observe the point belongs to class 2. When we consider the mahalanobis distance we get distance to class 1 as 7.21 and distance to class 2 as 3.36.

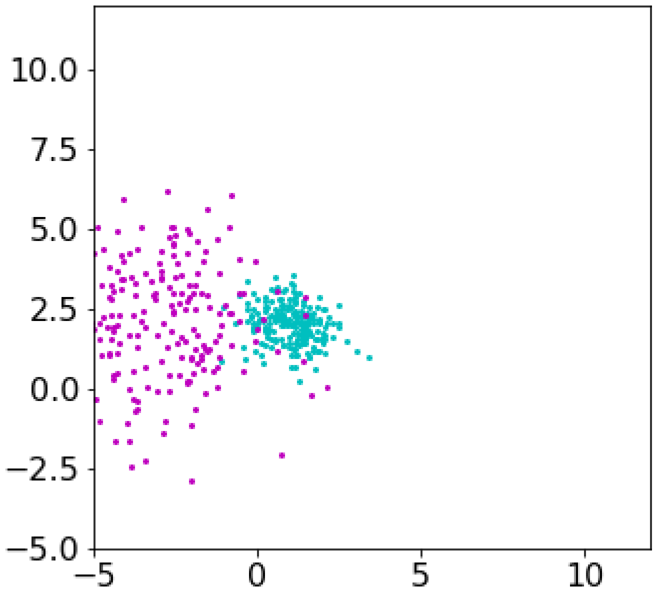


Figure 7: Scatterplot of the two distributions with m1 = $\left(\begin{smallmatrix}1\\2\end{smallmatrix}\right)$, m2 = $\left(\begin{smallmatrix}-3\\2\end{smallmatrix}\right)$ and covariance C1 = $\left(\begin{smallmatrix}0.5 & 0\\0 & 0.5\end{smallmatrix}\right)$, C2 = $\left(\begin{smallmatrix}3 & 0\\0 & 3\end{smallmatrix}\right)$.

The mahalanobis distance to mean classifier correctly classifies the point. This is because unlike the distance to mean classifier which only considered distance, the mahalanobis distance has inverse covariance term which helps in decorrelating the two distributions. This becomes significant in cases like that shown in figure 7 in which one class has much higher variance compared to other.