Author:MAO YANG
Email:my4n20@soton.ac.uk

1 change the means and covariance matrices given above to illustrate the differences we are learning about.

(1)When $\boldsymbol{m}_1 = \begin{pmatrix} 0 \\ 3 \end{pmatrix}$ and $\boldsymbol{m}_1 = \begin{pmatrix} 3 \\ 2.5 \end{pmatrix}$, $\boldsymbol{C}_1 = \boldsymbol{C}_2 = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}$, $P_1 = P2 = 0.5$;

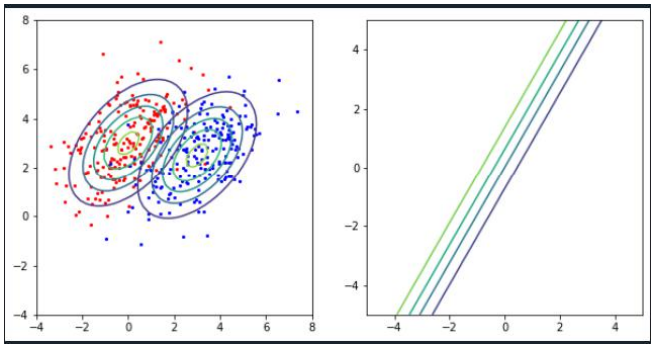

Figure 1

(2)When compare to the first one ,it change the P1 and P2, which is : P1=0.9 P2=0.1

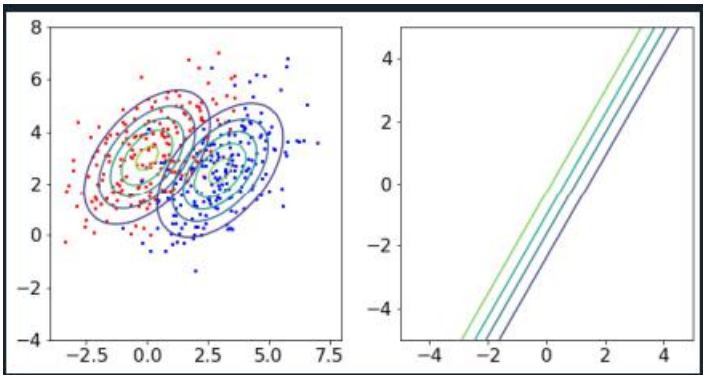

Figure 2

It is a plus one than the first one. And the posterior boundary is higher than the first one.

(3)When $\boldsymbol{C}_1 = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}$, $\boldsymbol{C}_2 = \begin{pmatrix} 1.5 & 0 \\ 0 & 1.5 \end{pmatrix}$, $P_1 = P2 = 0.5$.

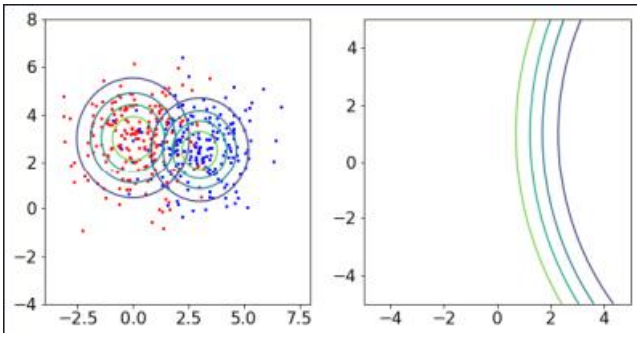

Figure 3

The shape of the data is a circle and posterior boundary lines are curved.

And the red one is lager than the blue one.

2.

(3) **Compute the Fisher Linear Discriminant direction**

The result of this:

[-1.08333337 0.66666669]

(4) Project the data onto the Fisher discriminant directions and plot histograms of the distri bution of projections:

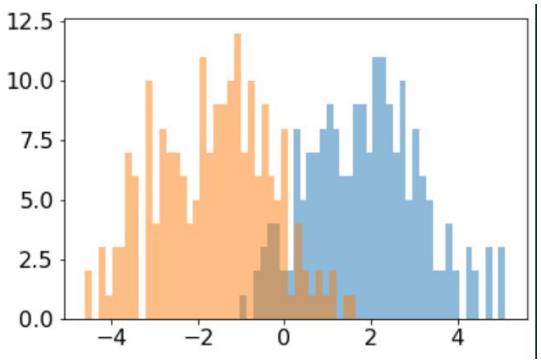

Figure 4

(5) Compute and plot the Receiver Operating Characteristic (ROC) curve, by sliding a decision threshold, and computing the True Positive and False Positive rates (see code snippet in Appendix.

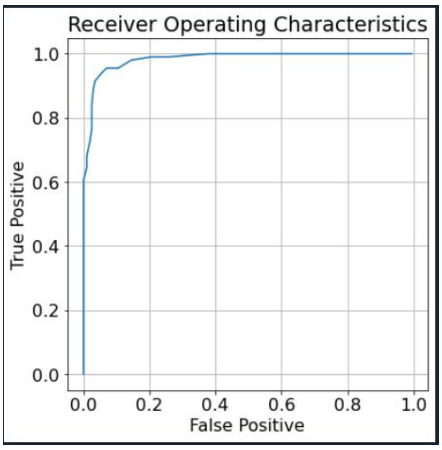

Figure 5

(6) Compute the area under the ROC curve

The area of ROC curve is 0.979.

AUC:0.979

(7) For a suitable choice of decision threshold, compute the classification accuracy.

As we can see from figure 4 and figure 5, when we make a choice around zero and I calculate it, the results is in blow:

Best Threshold:0.204
Best Accuracy:0.943

If we draw the"True Positive", "False Positive" and "Accuracy" in one picture, we can come to a same conclusion:

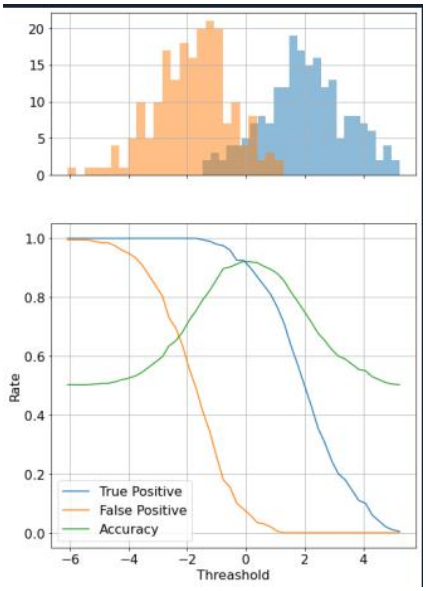

Figure 6 ”True Positive”, “False Positive” and “Accuracy”

Max Acc: 0.922 Thresh:-0.083

(8) Plot the ROC curve (on the same scale) for

- A random direction (instead of the Fisher discriminant direction).
- Projections onto the direction connecting the means of the two classes.

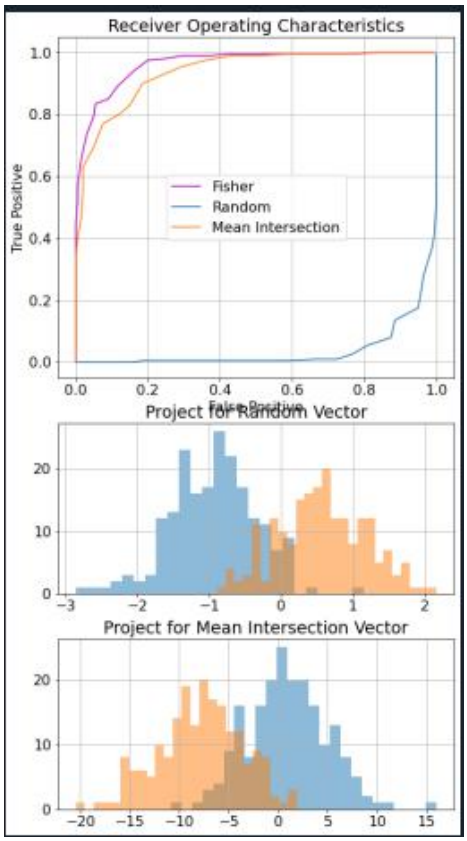

Figure 7

Compute the area under the ROC curve (AUC) for these two cases. Your report should explain what the precise statistical interpretation of AUC is and when it is used.

w Fisher: [-1.08333337 0.66666669]

w Rand: [0.39629657 0.13345883]
w Fisher: [-3. 0.5]
AUC Fisher:0.956
AUC Random:0.112
AUC Fisher:0.944

The area below the curve can be calculated by integrating along a given axis using the composite trapezoid rule.

3 Using a suitable classification problem (two-class in two dimensions, adapted from one of the above examples)

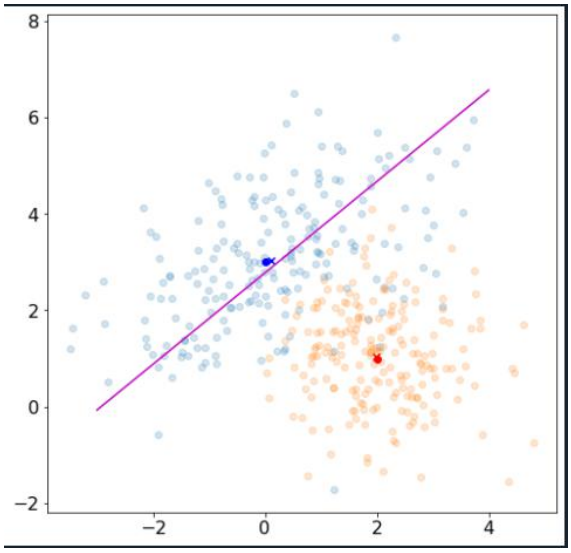

Figure 8

Pred m1: [0.12068804 3.02752718]
Pred m2: [1.99758226 1.04656628]
Pred c1: [[2.21314728 1.07596938][1.07596938 1.99712153]]
Pred c2: [[0.89084087 -0.11150014][-0.11150014 0.99167979]]
Euc Train Acc:0.919
Euc Test Acc:0.863
Mah Train Acc:0.919

The difference between a distance-to-mean classifier and a Maha lanobis distance-to-mean classifier:

The Mahalanobis distance-to-mean classifier takes into account the variances of different sample classes, not just the geometric distances.