Machine Learning Lab 3

Zhikun Zhu; ID: 29356822

October 25, 2017

I Question 1

Two bivariate Gaussian distributions are shown in Figure 1, with same covariance $C = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}$ and means $m_1 = \begin{pmatrix} 0 \\ 2 \end{pmatrix}$, $m_2 = \begin{pmatrix} 1.7 \\ 2.5 \end{pmatrix}$, respectively.

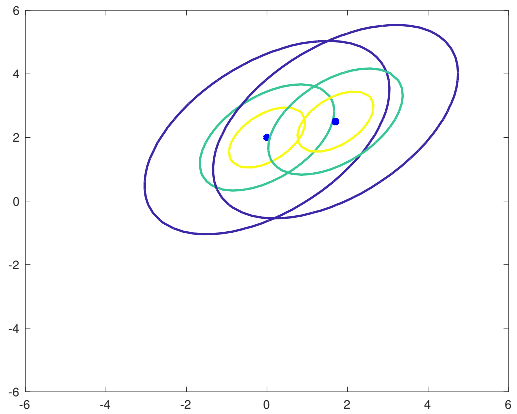

Figure 1: Contour map of two Gaussian distributions

Centres of both distributions are ploted, which are exactly their own means. Besides, since two covariance matrixs are same, their histograms looks parallel to each other. Therefore, the Fisher Liner Discriminate direction is: $W_F = C^{-1}(m_1 - m_2)$.

II Question 2

Figure 2 shows two random sequences which obey bivariate Gaussian normal distributions mentioned in previous question, both with a length of 500. The black and red dots represent Gaussian distributions with m_1 and m_2 means respectively.

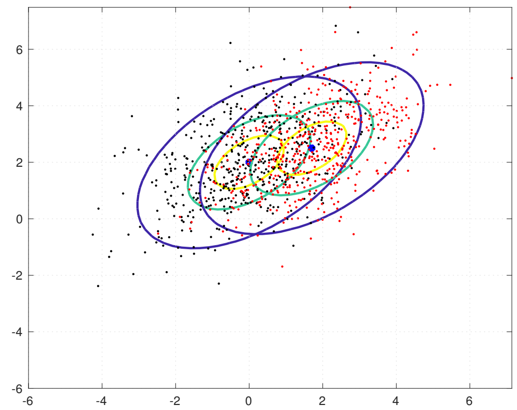

Figure 2: Two random Gaussian sequences follow distributions stated in Q.1

It is obvious from Figure 2 that most randoms are distribute around the inner circle of respective histogramwhich can be explained by looking at their Probability Distribution Function (PDF). Region around the centre have higher probability.

III Question 3

By applying gradient method to Fisher ration we get:

$$2C_B w \times (w^T C_W w) = C_w w \times (w^T C_B w) \quad (1)$$

Then we can immediately simplify above equation to:

$$C_B w w^T C_W w = C_w w w^T C_B w$$

Since $w^T C_W w$ and $w^T C_B w$ are both scalars and won't change w's direction, so we can eliminate them. Then substitute $C_B = (m_1 - m_2)(m_1 - m_2)^T$ into equation above we get:

$$(m_1 - m_2)(m_1 - m_2)^T w = C_w w$$

Similarly, we can eliminate $(m_1 - m_2)^T w$ as it is a scalar. Finally we get analytic expression for w_F:

$$w_F = \alpha C_w^{-1}(m_1 - m_2) \qquad (2)$$

Here, $\alpha \in R$ means w_F is a set of vectors with same direction. Finally, the Fisher Liner Discriminate direction is ploted in Figure 3 as a red line, which has same direction with w_F.

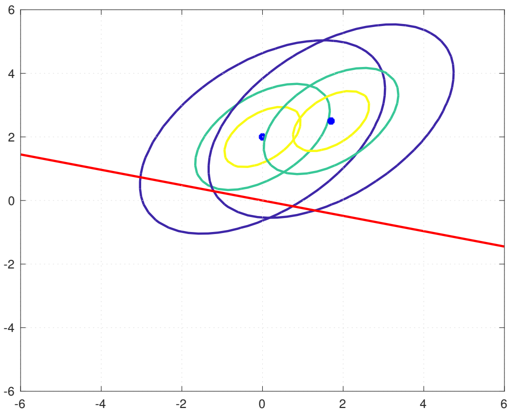

Figure 3: Fisher liner discriminate direction

It is easy from figure 3 to predict that even applying optimized projection directions, the distance between two classes data will still small. This can be explained by investigating their similiar means and same covariance.

IV Question 4

We can therefore project two classes x_1 and x_2 onto the Fisher Liner Discriminate direction by muptiply it:

$$x_{p_i} = x_i w_F \qquad (3)$$

The histogram of projected data is shown in figure 4. Horizontal axis is stricted by minimum and maximum of two projected class. Vertical axis is stricted between 0 and highest histogram for intutive comparsion.

Fisher Liner Discriminate method separate two class maximally. Figure 4 illustrate that projected data from class 1 and class 2

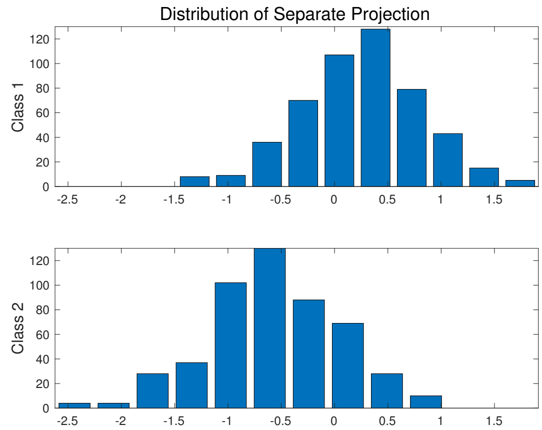

Figure 4: Histogram of projected distributions of two classes

centred around different values. Futhermore, figure 5 plot joint histogram distribution of two classes as well as their individual projected Probability Density Function (PDF), which can reflect real data distribution.

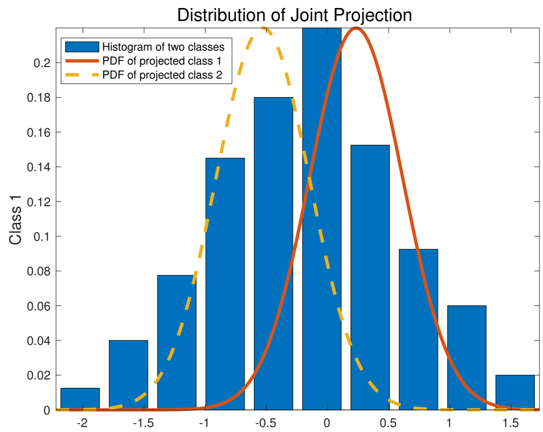

Figure 5: Projected PDF of two classes and joint histogram

V Question 5 & 6

Receiver operating characteristic curve are computed according to table 1. Applying function *trapz* $(\cdot)$ we can get the area under the ROC curve, which is 8.17×10^3. Decision threshold steps are generated by: $\frac{data_{max} - data_{min}}{M}$, where M is the resolution.

Real	Prediction	
	p_1	p_2
p_1	True Postive	False Negative
p_2	False Postive	True Negative

Table 1: Through three rounds at the Muni.

The result ROC curve is shown in figure 6, which is not smooth since the slow resolution.

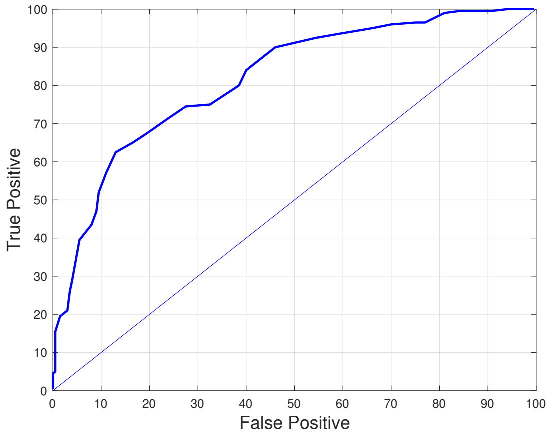

Figure 6: Receiver Operating Characteristic (ROC) curve

VI Question 7

Since two classes data have same covariance and different means. The decision threshold with bast performance to class the projected data into two class may be the middle of two means. We can calculate two projected means:

$$\begin{aligned} m_{1p} &= w_F^T m_1 \\ m_{2p} &= w_F^T m_2 \end{aligned} \tag{4}$$

Then, we can get the threshold by take median of $m_{1p} + m_{2p}$, which equals *-0.1483*. Therefore, we can calculate classification accuracy by following code.

```
threshold = (mm1+mm2)/2;
m = length(find(p1 > threshold));
n = length(find(p2 < threshold));
per = (m+n)/(2*N);
```

And the classification accuracy is $per = 0.735$.

VII Question 8

Let $w_R = \begin{pmatrix} sin(\theta) \\ cos(\theta) \end{pmatrix}$ and $\theta = \pi/4$. The new ROC curve with random project direction is shown in figure 7. And the area under the curve is 2.1631×10^3 and 8.1342×10^3 for the random and $m_1 - m_2$ direction, respectively.

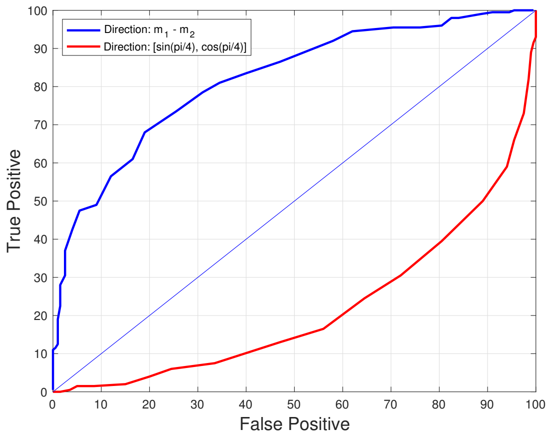

Figure 7: ROC curve of random and $m_1 - m_2$ project direction

VIII Question 9

Here we implement a 1 Nearest Neighbour Classifier to sort the data into two classes, based on the Euclidean norm between each data. The algorithm traverse the data to find the point with nearest distance to each point belong to the data set. Finally, the nearest neighbour accuracy in this case is *64.75%*, which performance is worse than fisher liner discriminate classifier.

IX Question 10

Two k-meanss algorithm using *Euclidean distance* and *Mahalanobis distance* respectively were generated to solve such clustering problem. Since there are two classes in the data, we choose $k = 2$.
The algorithm class the data into two class, the final result is shown in figure 8. And the correct rates for *Mahalanobis distance* and

Euclidean distance are *68.25%* and *73.50%*, respectively.

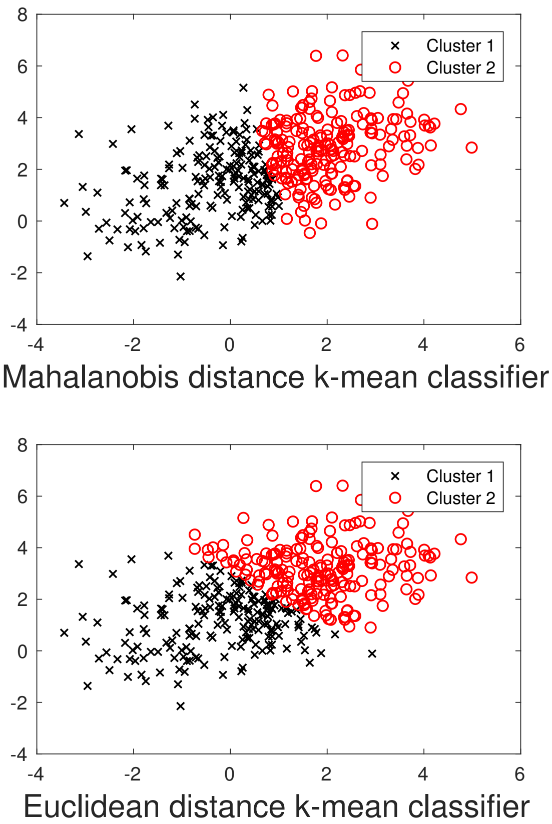

Figure 8: k-mean classifier under Euclidean & Mahalanobis distance

X Question 11

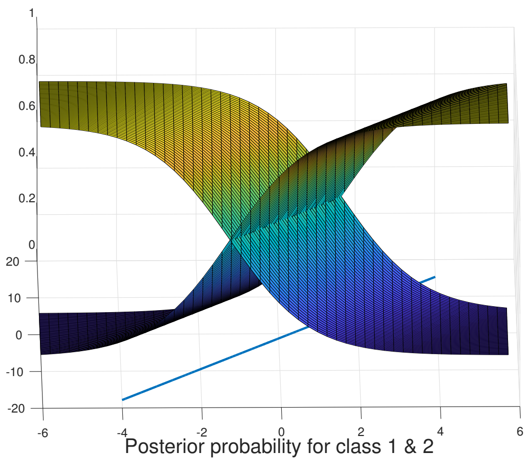

Figure 9: Posterior probabilities for class 1 & 2 and Baye's class boundary

Posterior probabilities for class 1 and 2 is shown in figure 9 as well as the Bayes liner classifier. It is clearly that the decision threshold line is same with the intersecting line of two probabilities' graph.

XI Question 12

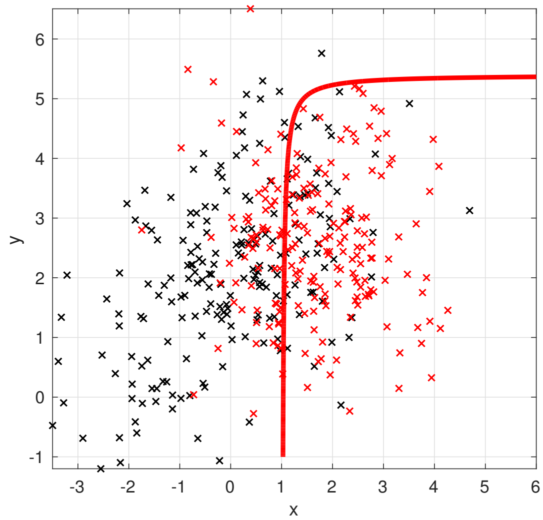

Figure 10: Baye's optimal class boundary $(C_2 = 1.5I)$

From Bayes classifier decision rule and bivariate Gaussian distribution we can get classifier decision threshold equation. Here we assume $w_1 = w_2$.

$$\frac{1}{(2\pi)^{p/2}det(C_1)^{1/2}}exp\{\frac{1}{2}(x-m_1)^tC_1^{-1}(x-m_1)\} \\ = \frac{1}{(2\pi)^{p/2}det(C_2)^{1/2}}exp\{\frac{1}{2}(x-m_2)^tC_2^{-1}(x-m_2)\} \quad (5)$$

Take $log_e(\cdot)$ on both side and simplify above equation:

$$x^t(C_1^{-1} - C_2^{-1})x + 2\omega^t x + b = log_e(\frac{|C_2|}{|C_1|}) \quad (6)$$

Here b $= (m_1^tC_1^{-1}m_1 - m_2^tC_2^{-1}m_2)$, and $\omega = m_2^tC_2^{-1} - m_1^tC_1^{-1}$ (For semi-definite matrix C, $C^t = C$).

Obviously, equation 7 containins element $x(1) \cdot x(2)$, so the Baye's optimal class boundary should be a curve instead of a straight line.