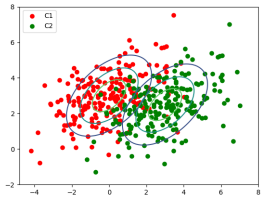
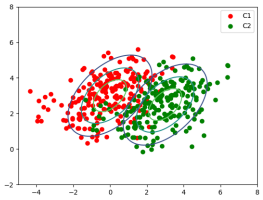
COMP6245 : Lab 3 Report

Thanakorn Panyapiang(tp2n19@soton.ac.uk)

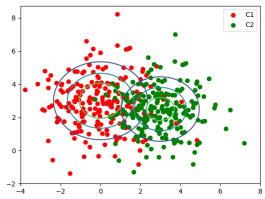
1 Class Boundaries and Posterior Probability



$\Sigma_1 = \Sigma_2 = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$
$P_1 = P_2 = 0.5$



$\Sigma_1 = \Sigma_2 = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$
$P_1 = 0.7, P_2 = 0.3$



$\Sigma_1 = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix} \Sigma_2 = \begin{bmatrix} 1.5 & 0 \\ 0 & 1.5 \end{bmatrix}$
$P_1 = P_2 = 0.5$

Figure 1

Figure 1 illustrates the scatter of three classification problems which are drawn from different distributions.

The class boundary is the area where the posterior probability of ω_1 and ω_2 are equal which can be written as an equation below:

$$P(\omega_1|x) = P(\omega_2|x) \tag{1}$$

Since both classes have a Gaussian distribution, Eq. 1 can be define as:

$$P_1 \frac{1}{|\Sigma_1|2\pi} e^{-(x-\mu_1)^T \Sigma_1^{-1}(x-\mu_1)/2} = P_2 \frac{1}{|\Sigma_2|2\pi} e^{-(x-\mu_2)^T \Sigma_2^{-1}(x-\mu_2)/2} \tag{2}$$

As covariance matrix and the prior probabilities are equal on the first dataset. Eq. 2 can be simplified by eliminating all equal terms which results as follow.

$$(x-\mu_1)^T (x-\mu_1) = (x-\mu_2)^T (x-\mu_2)$$

$$x^T x - \mu_1^T \mu_1 - \mu_1^T x - x^T \mu_1 = x^T x - \mu_2^T \mu_2 - \mu_2^T x - x^T \mu_2$$

$x^T x$ can be removed as it is on both sides, $\mu_1^T \mu_1$ and $\mu_2^T \mu_2$ are constant, and $\mu_1^T x$ and $\mu_2^T x$ are scalar, the decision boundary can be simplified to

$$x^T \mu_1 + C_1 = x^T \mu_2 + C_2$$

$$x^T(\mu_1 - \mu_2) + C = 0 \tag{3}$$

Eq.3 suggests that the class boundary should be a straight line similar to Figure 2a.

On the second problem, although the covariance matrix is the same as the first problem, the prior probabilities are different. By solving Eq. 2, the decision boundary will be as follow:

$$(x^T x - \mu_1^T \mu_1 - \mu_1^T x - x^T \mu_1) + log(P_1) = (x^T x - \mu_2^T \mu_2 - \mu_2^T x - x^T \mu_2) + log(P_2)$$

$$x^T(\mu_1 - \mu_2) + C + log(\frac{P1}{P_2}) = 0$$

As the prior probability of ω_1 is higher than ω_2, $log(\frac{P1}{P_2})$ is more than zero. Therefore the decision boundary is similar to the first problem but slightly shifted to the right as shown in Figure 2b.

On the third dataset, unlike the first two, the covariance metrices are different so they can't be eliminated from the Eq. 2. The decision boundary will be as the following equation.

$$-log(\Sigma_1) + \frac{1}{2}(x - \mu_1)^T \Sigma_1^{-1}(x - \mu_1) = -log(\Sigma_2) + \frac{1}{2}(x - \mu_2)^T \Sigma_2^{-1}(x - \mu_2)$$

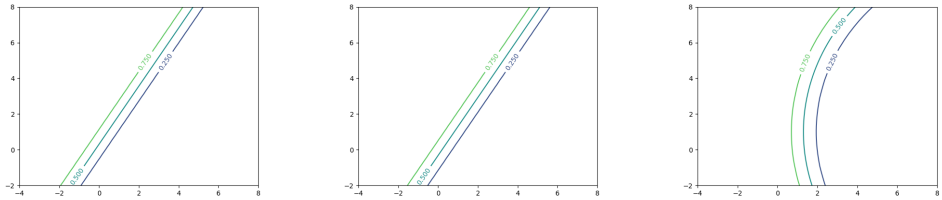
Solving the equation above will give the following equation:

$$x^T(\Sigma_1^{-1} - \Sigma_2^{-1})x + 2(\Sigma_1^{-1}\mu_1 - \Sigma_1^{-1}\mu_1)x + \mu_1^T \Sigma_1^{-1}\mu_1 - \mu_2^T \Sigma_2^{-1}\mu_2 - log(\frac{\Sigma_2}{\Sigma_1}) = 0$$

As $\mu_1^T \Sigma_1^{-1}\mu_1$, $\mu_1^T \Sigma_2^{-1}\mu_2$, and $log(\frac{\Sigma_2}{\Sigma_1})$ are scalar. It can be simplified to C. Moreover, $\Sigma_1^{-1} - \Sigma_2^{-1}$ and $\Sigma_1^{-1}\mu_1 - \Sigma_1^{-1}\mu_1$ are independent from x so it can be reduced to A and B respectively. Therefore, the decision boundary can be written as follow:

$$x^T Ax + 2Bx + C = 0$$

The equation above indicates that the decision boundary is a quadratic function which is consistent with the graph on Figure 2C.

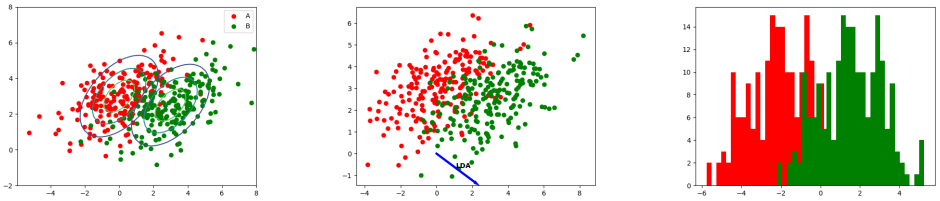


A B C

Figure 2

2 Fisher LDA and ROC Curve

Ditributions of original data, Fisher Linear Discriminant direction, and histograms of projected data are shown in Figure 3. The ROC curve and classification accuracy of Fisher LDA on several threshold values are illustrated in Figure 4.



Data distribution Fisher LDA direction Projected data distribution

Figure 3

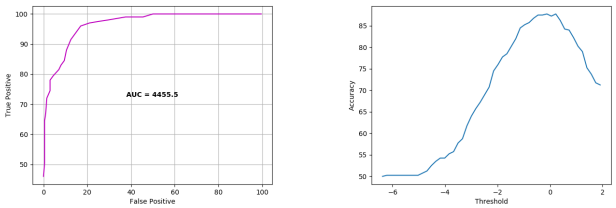
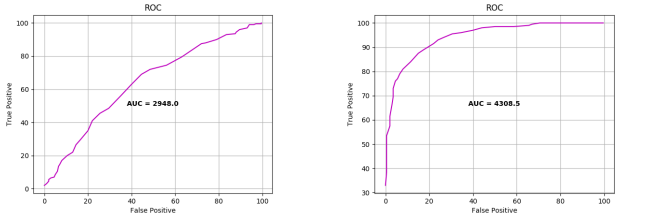


Figure 4

Figure 5. shows ROC curves of another two classifiers which project the original data onto the random direction and the direction connecting means respectively.



Random Direction Direction Connecting m_1 and m_2

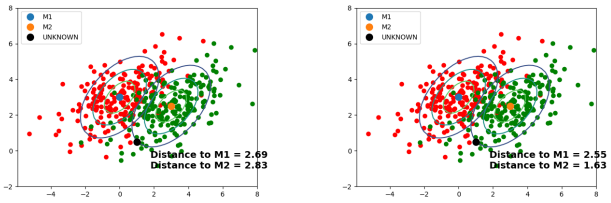
Figure 5

The area under the curve(AUC) of ROC is computed by adding True Positive of different thresholds. In other word, this area is the sum of positive samples which are correctly classified using several thresholds. This number reflects the

accuracy of the classifier on the positive class. The classifier with high AUC is more likely to label a positive data point as positive than negative.

AUC of ROC curve is normally used to compare the performance of a classifiers. A model with high AUC tends to do a better job than a model which has low AUC.

3 Mahalanobis Distance



Euclidean Distance Mahalanobis Distance

Figure 6

The Euclidean distance-to-mean classifier will classify the unknown sample in Figure 6. as C_1 while the Mahalanobis distance-to-mean classifier will classify it as C_2. Based on the decision boundary which is shown in Figure 1., the Mahalanobis distance-to-mean classifies is correct in this case.

From the picture, although the unknown data point is geometricly closer to m_1 than m_2, it can be observed that the point lies on the outmost contour of C_2 which means it is 2 standard deviation away from m_2. On the other hand, the data point is located outside the outmost contour of C_1 so the distance to m_1 is more than 2 standard deviation. This observation suggests that the unknown data point is more likely to be C_2 than C_1 because both classes have a Gaussian ditribution.

The above conclusion explains the difference between two classifiers. While the Euclidean distance only tells geometric distance between two data points, the Mahalanobis distance reflects the distance between a data point and the distribution as it takes covariance matrix into account.