Machine Learning Lab 3

Di Wu

dw4y17@soton.ac.uk

1 Exercise 1 & 2

1.1 Contour plot and scatter plot

In the first and second exercises, we attempted to be familiar with the contour plot in two dimensional multivariable gaussian distribution, with means $\mathbf{m_1} = [0, 2]^t$,$\mathbf{m_2} = [1.7, 2.5]^t$ and covariance matrix $\mathbf{C_1} = \mathbf{C_2} = [2, 1; 1, 2]$. The contour plot and scatter plot are shown bellow as **Figure 1 & Figure 2** :

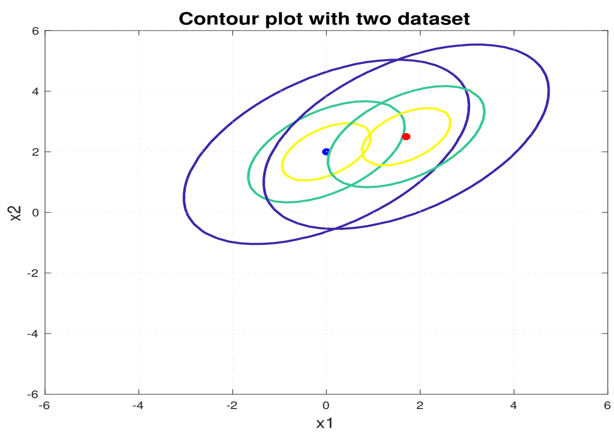

Figure 1 The contour plot with two datasets

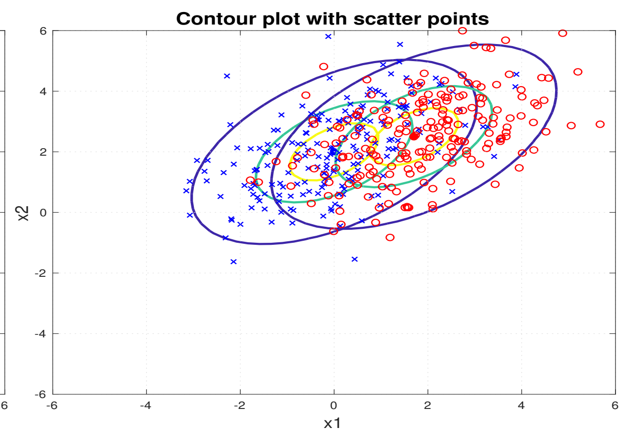

Figure 2 Respective scatter plot

2 Exercise 3 & 4

2.1 Fisher Linear Discriminant

The Fisher Linear Discriminant is a method that project the data to one dimension. The projected direction is chosen by maximising the fisher ratio which is actually maximising the distance between the means of two classes and minimising the variances of two classes. we could compute the direction by $\mathbf{w_f} = \lambda \mathbf{C_w^{-1}}(\mathbf{m_1} - \mathbf{m_2})$ with $\mathbf{m_1} = [0, 2]^t$,$\mathbf{m_2} = [1.7, 2.5]^t$

and $\mathbf{C_w^{-1}} = \mathbf{C_1} + \mathbf{C_2}$. Therefore in these exercises we could plot the direction of $\mathbf{w_f}$ and after projecting the data to this direction we could plot the histograms of distribution.The direction and distribution histogram are shown bellow as **Figure 3 & Figure 4** :

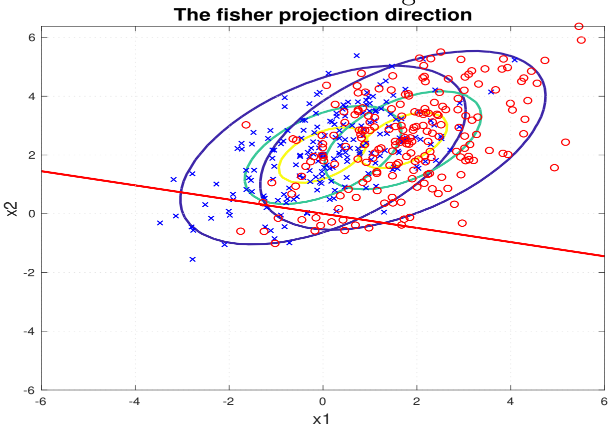

Figure 3 The projected direction of fisher

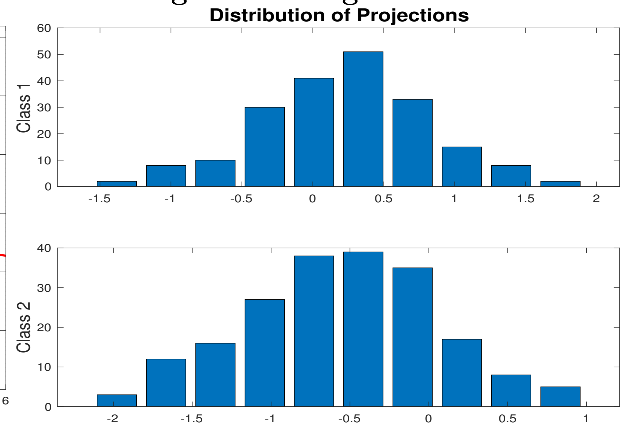

Figure 4 Distribution of projection

3 Exercise 5,6,7 & 8

3.1 More details of ROC curve and AUC

The ROC curve is a measure to compare different classifiers and the performance of one classifier. the curve is formed by computing different pairs of (FPR ,TPR) with changing the thresholds. It is superior to the metric accuracy, because it could reflect more details about the predicted situation vs true states than accuracy. It has many great properties. one is how to choose a suitable threshold. the 'suitable' means different criterion in different situations(relying on the cost of different error), but normally we want choose the threshold which has the largest accuracy. the way to find this point is to find the intersection of accuracy line(slope=#neg / #pos, in exercise 7 is equal to 1) and descending diagonal. The figure is shown as **Figure 5**, and the $\mathbf{accuracy = 75\%}$. Another great property of ROC is AUC(area under the ROC curve), it stand for the performance of this classifier. The larger the nember it is, the better the performance of classifier will be. Therefor in **Figure 6** we compare the AUC in three different projected directions: fisher(blue line), random(green line) and the connection of $\mathbf{m_1}$ and $\mathbf{m_2}$(red line). The AUC of them is $\mathbf{aucfisher = 7995}, \mathbf{aucm1m2 = 7690}$ and $\mathbf{aucrandom = 6059}$

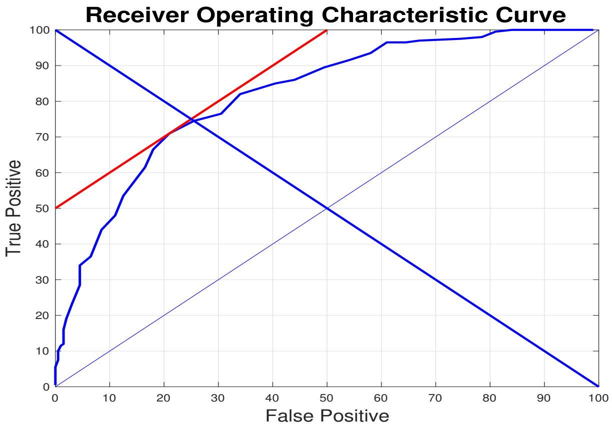

Figure 5 The ROC of fisher

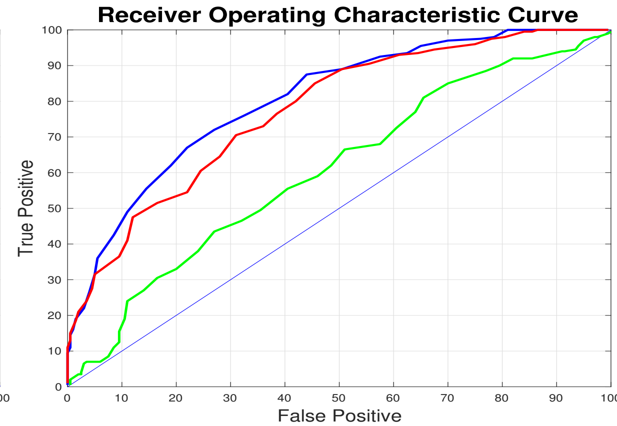

Figure 6 The ROC of different classifiers

4 Exercise 9 & 10

4.1 KNN, Euclidean and Mahalanobis classifiers

By using these classifiers, we could differentiate two classes with different accuracy. The first one is KNN, a non parameters classifier, the principle of KNN is simple we label the unknown $\mathbf{x}$ with the label which is generate by his nearest K neighbours. In this lab we simply consider just one nearest neighbour. The second and third classifiers is consider different types of distance. We give the label the same as the smaller distance between this unknown point $\mathbf{x}$ and means. the different is the measurement of distance, we are familiar with the Euclidean distance while the Mahalanobis distance is considering the distribution of datapoint and the covariance of two datasets. The accuracy is as fellow: $\mathbf{accKNN = 50}\%, \mathbf{accEuclidean = 69}\%$ and $\mathbf{accMahalanobis = 70}\%$.

5 Exercise 11 & 12

5.1 The shape of Bayesian decision boundary

The linear Bayesian decision boundary has strict assumptions that the datasets have the same covariance matrices(in this lab is $\mathbf{C_1} = \mathbf{C_2} = [2, 1; 1, 2]$) , in this case we can reasoning the expression that the Bayesian decision boundary is a straight line. But if we cancel the assumption that the two dataset have different covariance matrices like in Exercise 12 $\mathbf{C_1} = [2, 1; 1, 2]\mathbf{C_2} = [1.5, 0; 0, 1.5]$, The Bayesian decision boundary is not a straight line at all. The decision boundary is
$\mathbf{x^t(c_1^{-1} - c_2^{-1})x + 2(m_2^t c_2^{-1} - m_1^t c_1^{-1})x + m_1^t c_1^{-1} m_1 - m_2^t c_2^{-1} m_2 - ln(det(c_2)/det(c_1)) = 0}$
We verified this in Exercise 11 & 12 by plot 3d graphs of posterior probabilities. And the graph is shown as **Figure 7 & 8**:

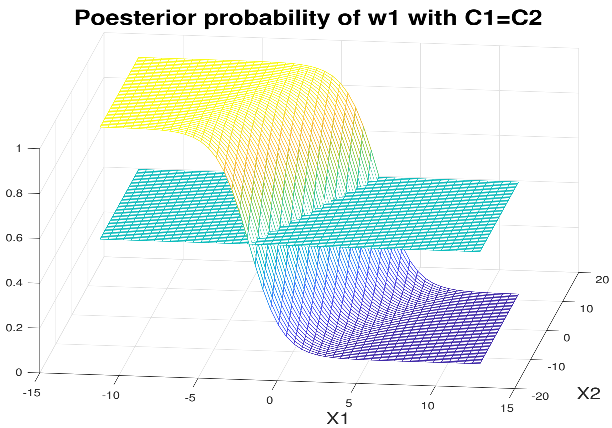

Figure 7 Situation with C1=C2

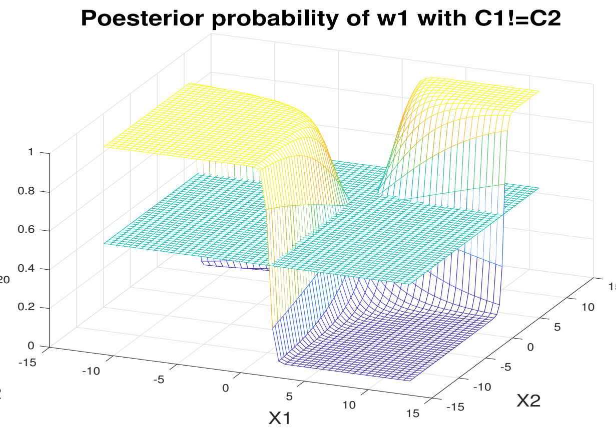

Figure 8 Situation with C1!=C2

As we could see in the graph above the shape of he Bayesian decision boundary is not a straight line when $\mathbf{C_1} = [2,1;1,2]\mathbf{C_2} = [1.5,0;0,1.5]$. Another clear graph could be seen as **Figure 9 & 10**:

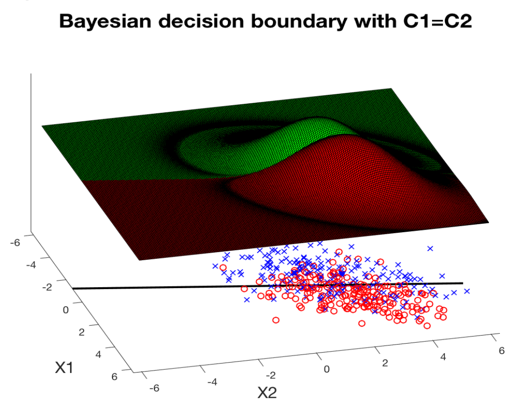

Figure 9 Situation with C1=C2

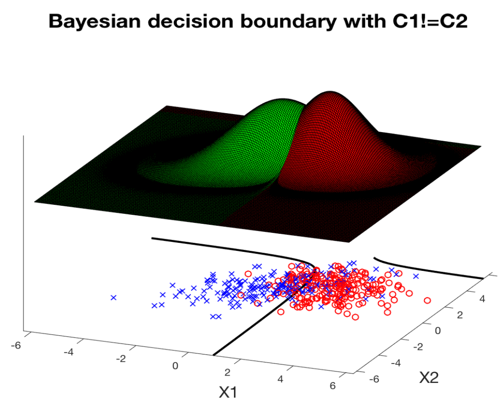

Figure 10 Situation with C1!=C2

6 Summary

6.1 Fisher Linear Discriminant, Perceptron, KNN, Bayesian decision boundary and ROC

We can use different linear classifiers algorithm to distinguish the dataset. but if we believe the Bayesian decision boundary is the optimal decision, we should be aware of that there are strict assumptions that the datasets have the same covariance matrices. It is the most important thing we need to know before we choose linear method to classify our data. After that we could use ROC curve to compare the performances of this classifiers.