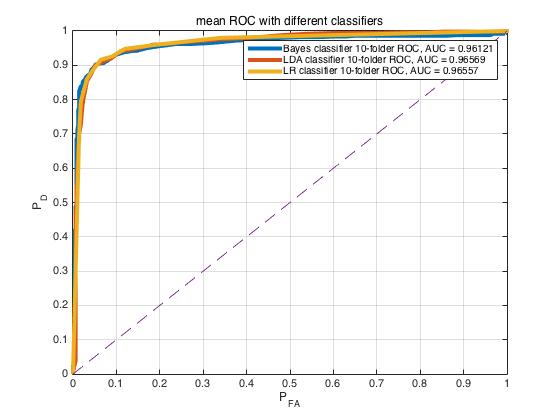
**ECE681 HW4 REPORT**

SHENGXIN QIAN

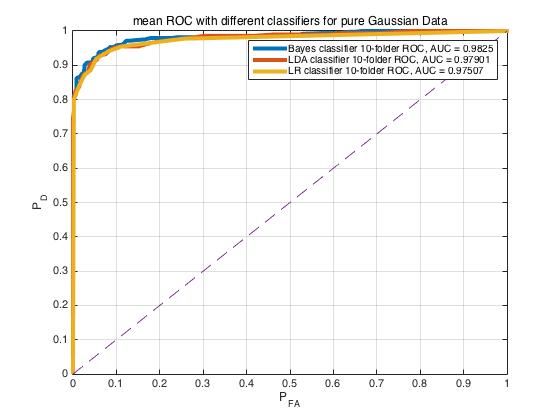
1. **Gaussian Data**

**Figure.1**

As we can see in Figure.1, it shows the cross-validated ROC curve and AUC of Bayes, FLD, Logistic Discriminant Classifier dealing with Gaussian data with a few outliers. All tests used 10-folder cross validation. All ROC curves are pretty close to each other which means that the performance of each classifier is pretty close when dealing with the required data set. Figure.2 shows the cross-validated ROC curve and AUC of those three classifiers dealing with pure Gaussian data. If we compared the ROC curves and AUC values of pure Gaussian data and impure Gaussian data, as we can see in Table.1, we could find the outliers do affect the performance of those classifiers.

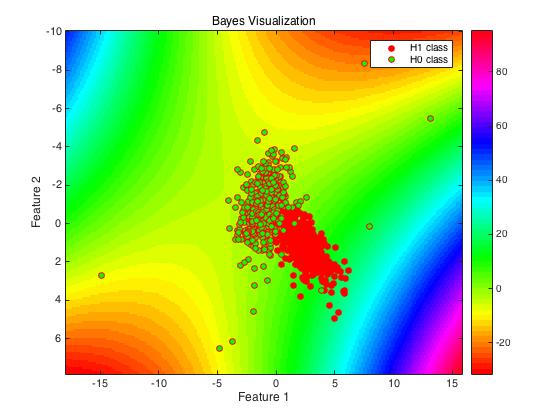
|  |  |  |  |
| --- | --- | --- | --- |
| Classifier | Bayes | FLD | Logistic Discriminant |
| Pure Gaussian AUC | 0.9825 | 0.9790 | 0.9751 |
| Gaussian with outlier AUC | 0.9612 | 0.9657 | 0.9656 |

**Table.1**

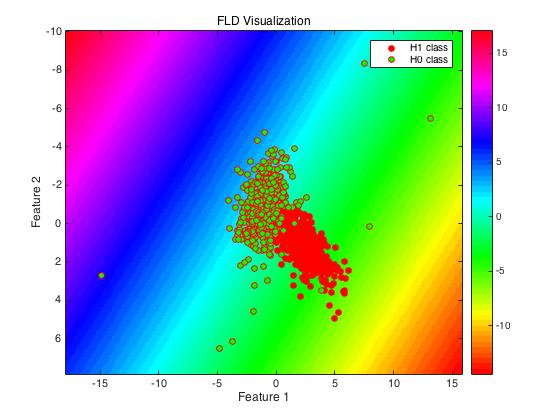


**Figure.2**

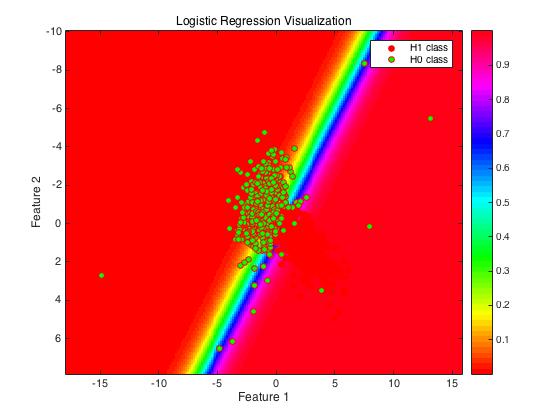
The Bayes classifier has highest AUC when dealing with pure Gaussian data and has lowest AUC when the data set has outliers. This result does make sense because Bayes classifier is sensitive to the distribution of dataset. If the tested dataset is not consistent with Gaussian distribution we predefined, the performance would drop.



**Figure.3**

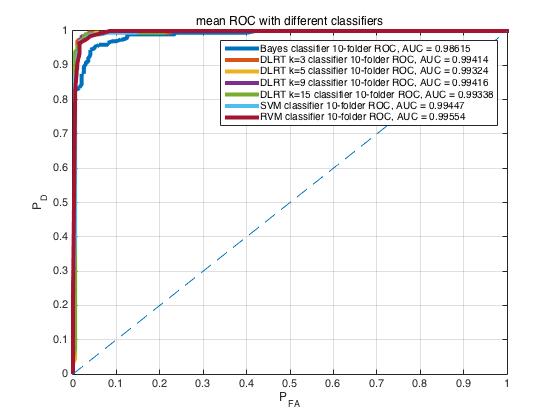
As we can see in Figure.3, the visualization result of Bayes classifier does make sense. In this case, we assume the probability of each class is 1/2 and probability density function of test data is 2-D Gaussian distribution. So, the decision statistics are determined by the ratio of Mahalanobis distance from test data points to the center of training data points and the ratio of covariance matrix determinant of each class. That is why the contour line of decision surface is similar to hyperbola based on Mahalanobis distance and it is symmetrical with respect to the diagonal. When we add several outliers, the distribution of contour line changed. Because the outlier would significantly affect the expectation and variance of dataset, if we still use the method of moments to estimate the parameters of distribution, the distribution we predefined would be inconsistent with actual distribution. However, Bayes Classifier still has advantages. The best part of Bayes Classifier is that it only need partial statistical information of the training data like mean, covariance and assumed probability distribution. So, when we finished the parameter estimation of assumed probability distribution, the computational complexity of calculating the decision statistic is pretty low. More than that, when we only know the partial statistical parameter and data’s probability distribution without whole training data, Bayes Classifier could still work. Moreover, because it is a high bias/low ****variance classifier, it works well on a small training set.

**Figure.4**

As we can see in Figure.4, the visualization of FLD classifier does make sense. The FLD classifier transforms 2-D data into one dimension and makes two classes have the best separability after transformation. It is obvious that the diagonal direction is the best choice. So, the corresponding optional decision boundaries are perpendicular to the diagonal. The FLD classifier still has good performance when dataset has outliers. This is because the samples with a few outliers could still be easily separated by a straight line and the overlap area of two classes is not very large. The strengths of FLD classifier is that it still based on the partial statistical information of training data but it does not need assumed probability distribution. It means that as long as the data is linearly separable, FLD classifier could work with low computational complexity and only need to know the mean and covariance of training data. If the overlap area is larger or the boundary between two classes is curved, the performance would drop because FLD suffers multicollinearity.

**Figure.5**

As we can see in Figure.5, the lower right is H1 class and the upper left is H0 class. Figure.5 clearly shows the visualization of Logistic Discriminant Classifier decision surface. Essentially, Logistic Discriminant Classification is still a regression problem. The core of this classifier is a vector of weights. When we calculate the decision statistics, we encapsulate the linear combination of features into sigmoid function. Because the sigmoid function is monotonically ascending, the decision boundaries are still straight lines. The strength part of Logistic Discriminant Classifier is that when we finished the training process, the time cost of prediction is pretty small. The weakness is the training speed is pretty slow because it need to solve optimization problems. More than that, same as FLD, it required dataset linearly separable. So, it suffers from multicollinearity as well.

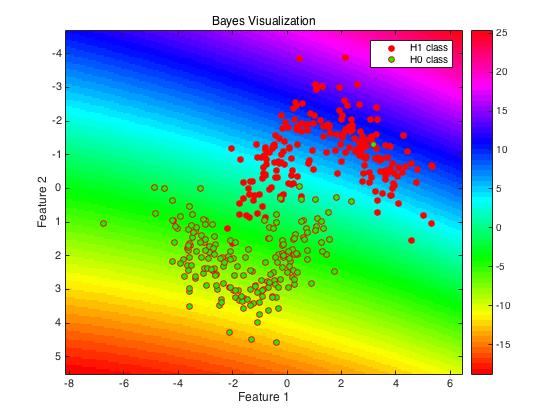
1. **Gaussian Mixture Data**

**Figure.6**

As we can see in Figure.6, it shows the cross-validated ROC curve and AUC of Bayes, SVM, RVM, DLRT with k=3,5,9,15 dealing with Gaussian Mixture Data. All tests used 10-folder cross validation. As we can see in Table.2, DLRT, SVM and RVM classifier have the best performance.

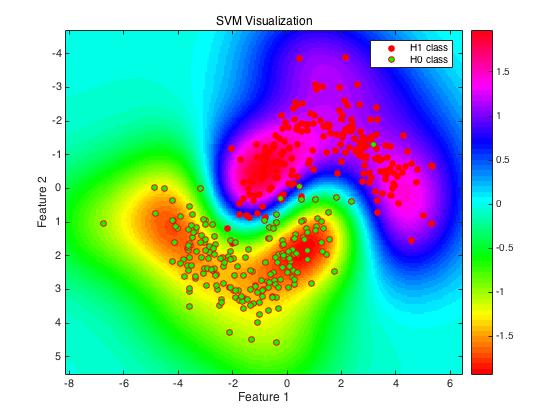
|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Classifier | Bayes | SVM | RVM | DLRT k=3 | DLRT k=5 | DLRT k=9 | DLRT k=15 |
| AUC | 0.9862 | 0.9945 | 0.9955 | 0.9941 | 0.9932 | 0.9942 | 0.9934 |

**Table.2**

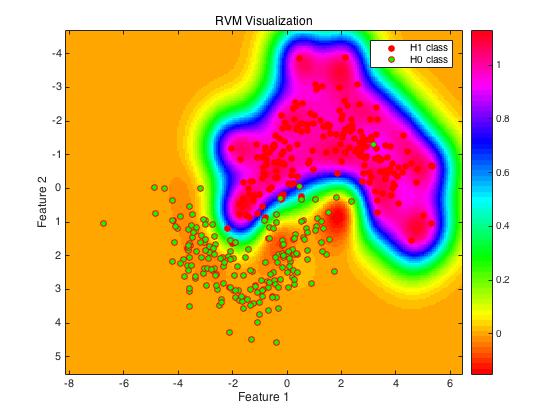
****

**Figure.7**

As we can see in Figure.7, the contour lines could partly separate two classes. The result does make sense. Even though the probability distribution of test samples we assumed is unimodal Gaussian which is inconsistent with the probability distribution of GMM data, because two classes do not overlap too much, Bayes classifier still has good performance. However, GMM data changed decision boundaries into straight lines. It means if the GMM dataset is not linearly separable, the performance would drop a lot. So, we should not use Bayes Classifier when the actual distribution is inconsistent with what we assumed.

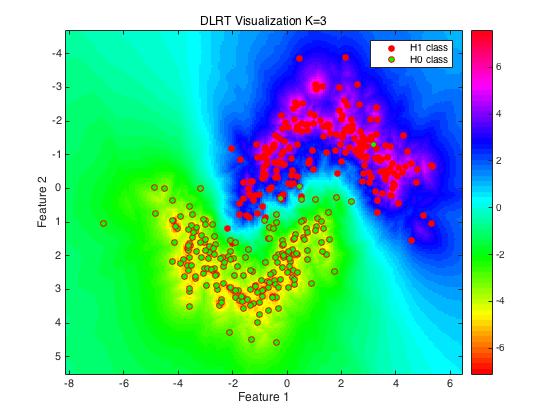
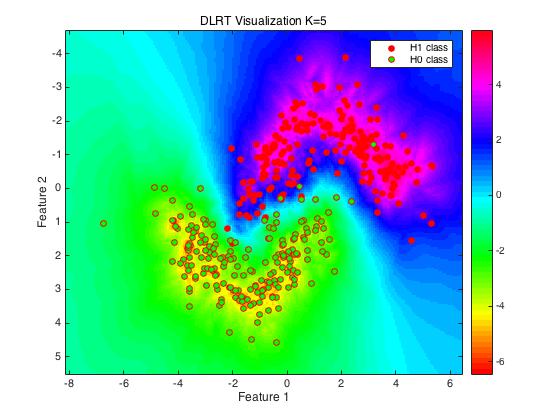
**Figure.8**

As we can see in Figure.8, SVM classifier perfectly separates two classes. In this case, I chose 0.7 as in radial basis kernel function. Because the kernel function we used is radial basis function, the decision statistics are determined by the distance between test data points and support vectors. At the same time, because the prediction of SVM classifier is only affected by the support vectors which are only small part of training points, the computational cost is pretty low compared with KNN or DLRT. The strength of SVM is it could be suitable for many types of data models as long as we choose the right kernel function. The weakness is that even though support vectors occupy a small part of training data, when the number of training points increased a lot, the computational cost would also increase.

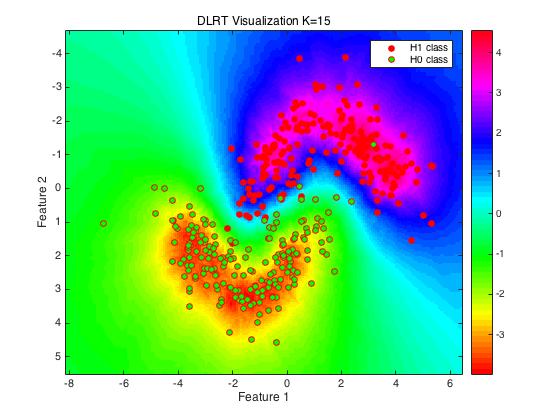
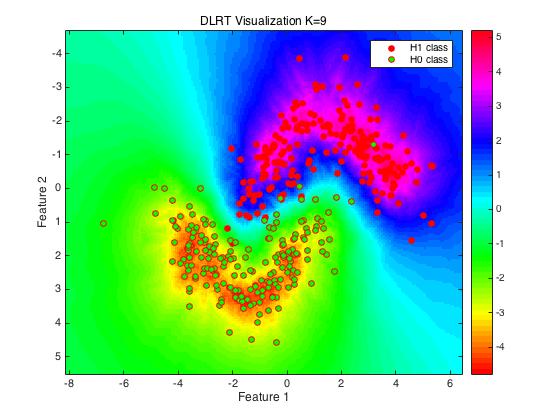
**Figure.9**

As we can see in Figure.9, RVM classifier handled GMM data pretty well.

When we use radial basis kernel function, it is important to choose the right . In this case, the parameter of radial basis function was searched from the . Finally, I chose which could make RVM achieve highest AUC. If the is larger, the classifier may tend to be underfitting. If the is lower, the classifier may tend to be overfitting. The visualization of decision surface does make sense. Different from SVM classifier, RVM classifier decreased the number of support vectors. The left lower side points are not support vectors any more compared with Figure.8. The strength of RVM classifier is that it significantly decreased the number of support vectors compared with SVM. So, RVM decreased the computational cost. The weakness is the training part need O(n2) space complexity.

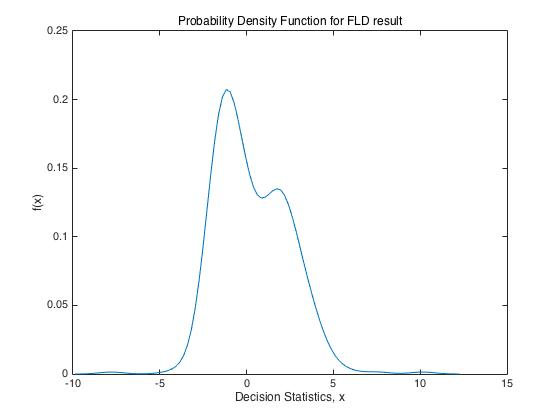
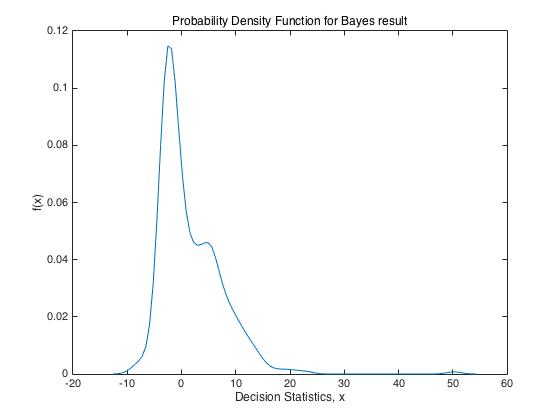
**Figure.10**

**Figure.11**

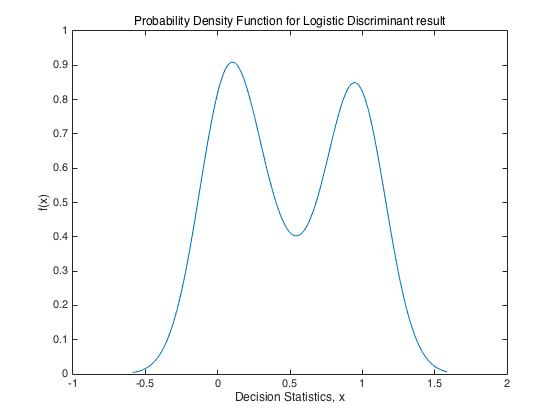
**Figure.12**

**Figure.13**

As we can see in Figure.10-13, when we use DLRT to deal with GMM data, the boundary on decision surface could successfully separate two classes. This result does make sense. Because DLRT classifier determines the decision statistics based its neighbor points’ location and density. As long as there is a boundary which could separate different classes, DLRT classifier could separate them no matter whether two classes are linearly separable or not. DLRT classifier is more versatile than Bayes and FLD classifiers. However, DLRT may underfit when k increases and overfit when k is low. More than that, the computational complexity is higher than SVM and RVM even the performance is similar.

1. ******Blind Tests**

**Figure.14** **Figure.15**



**Figure.16**

For blind “Gaussian” data set, I would choose Logistic Discriminant Classifier. As we can see in the graphs above, Figure.14 shows the probability density curve of FLD, Figure.15 shows the pdf of Bayes and Figure.16 shows the pdf of Logistic Discriminant Classifier. Clearly, Logistic Discriminant Classifier could make decision statistics have the best separability. More than that, after training process, the computational cost of Logistic Discriminant is pretty low. I did not choose Bayes Classifier because our training set has outliers which would significantly change the performance of classifier and Figure.15 supports my assumption. I did not choose FLD Classifier also because its probability density curve did not show a good separability.

For blind “Gaussian Mixture” data set, I would choose RVM classifier. I chose RVM because it performed pretty well on GMM data and the computational cost is low. I did not choose the Bayes classifier because the assumed unimodal Gaussian distribution may not handle Gaussian Mixture data very well. I did not choose SVM classifier because its computational complexity is high than RVM and the performance is similar. I did not choose DLRT because its computational complexity is too high.