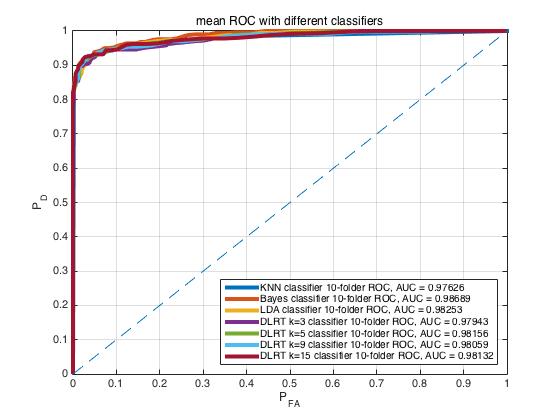
**ECE681 HW3 REPORT**

SHENGXIN QIAN

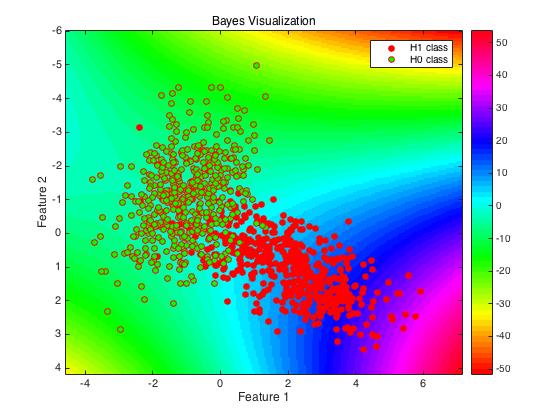
1. **Gaussian Data**

**Figure.1**

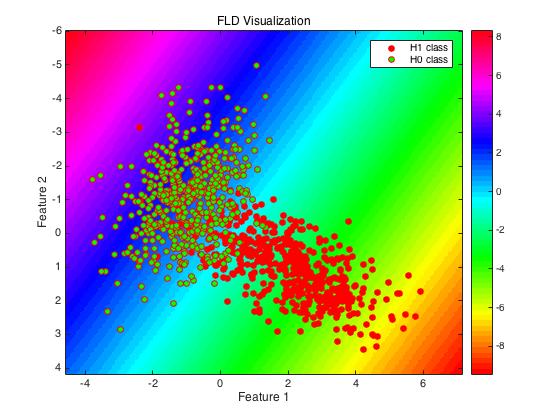
As we can see in Figure.1, it shows the cross-validated ROC curve and AUC of KNN with k=9, Bayes, FLD, DLRT with k=3,5,9,15. All tests used 10-folder cross validation. All ROC curves are pretty close to each other which means that the performance of each classifier is pretty close when dealing with 2-D Gaussian data. However, there is still a small difference when we look at the values of AUC.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Classifier | KNN k=9 | Bayes | FLD | DLRT k=3 | DLRT k=5 | DLRT k=9 | DLRT k=15 |
| AUC | 0.9763 | 0.9869 | 0.9825 | 0.9794 | 0.9816 | 0.9806 | 0.9813 |

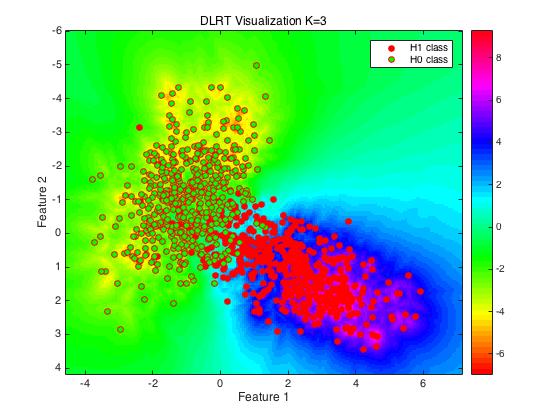
**Table.1**

As we can see in Table.1, the Bayes classifier has highest AUC. The AUC of DLRT relates to the number of neighbor points it chooses. We can not say there is a major performance difference between KNN with k=9, FLD, DLRT with k=3, 5, 9, 15 because the difference is less than the error of cross validation itself.

**Figure.2**

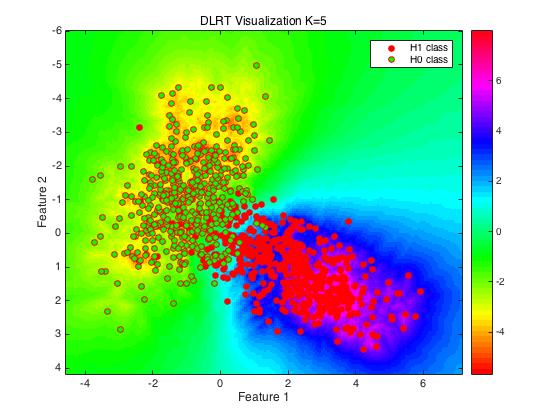
As we can see in Figure.2, the visualization result of Bayes classifier does make sense. In this case, we assume the probability of each class is 1/2 and probability density function of test data is 2-D Gaussian distribution which is consistent with how we generated samples. So, the decision statistics 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. The Bayes classifier has the best performance in this case is because the distribution of samples is Gaussian which is same as we assumed. 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. The weak part is that if the distribution of samples is not Gaussian or not consistent with the distribution we assumed, we need to change our assumed pdf. Otherwise, the performance would ****drop a lot.

**Figure.3**

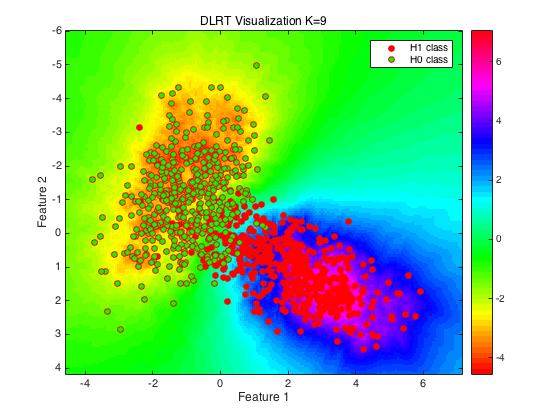
As we can see in Figure.3, the visualization of FLD classifier does make sense. The FLD classifier transforms 2-D data into one dimension and makes two classes have the best separability after transformation. It is obvious that the diagonal direction is the best choice. So, the corresponding optional decision boundaries are perpendicular to the diagonal. The FLD classifier has good performance is because the samples could 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.

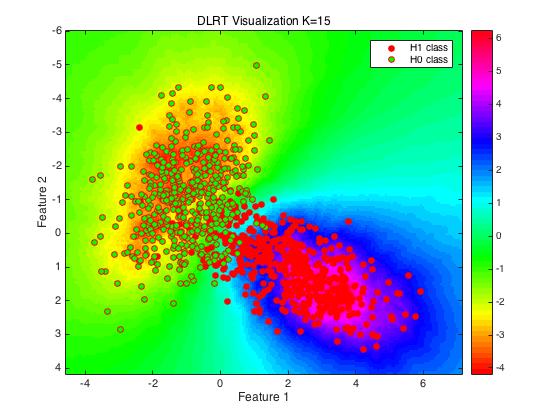
**Figure.4**

As we can see in Figure.4, the lower right is H1 class and the upper left is H0 class. Figure.4 clearly shows the difference between DLRT classifier and KNN classifier. When we use KNN classifier, the decision statistics only relate to the portion of each class in k nearest neighbors set which means the values of decision statistics are discrete and k nearest neighbors have the same weight. However, when we use DLRT classifier, the decision statistics are determined by the ratio of the kth furthest neighbor points distance of each class which makes decision statistics become continuous. More than that, the area with higher density would have the higher absolute value of decision statistics. The visualizations with k=5,9,15 were shown in Figure.5, Figure.6, Figure.7. DLRT classifier is a kind of low bias/high variance classifier. So, it may overfit especially when k is low. The strengths of DLRT is that it is simple and versatile to many distributions of samples. DLRT does not require linearly separable samples and assumed pdf. The weakness is that the computational complexity is higher than others.



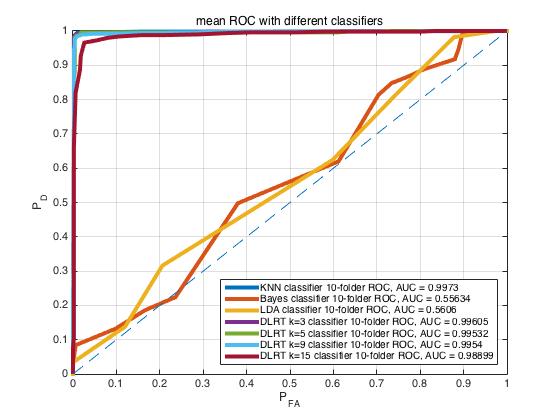
**Figure.5**



**Figure.6**

**Figure.7**

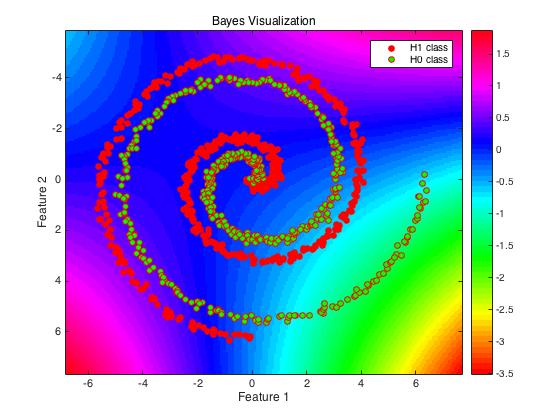
1. **Spiral Data**

**Figure.8**

As we can see in Figure.8, it shows the cross-validated ROC curve and AUC of KNN with k=9, Bayes, FLD, DLRT with k=3,5,9,15. All tests used 10-folder cross validation. As we can see in Table.2, Bayes Classifier with Gaussian pdf assumption and LDA Classifier have really poor performance in this case. KNN and DLRT classifiers have excellent performance when dealing with spiral data. DLRT classifiers with k=3,5,9 have similar performance but when k=15, the performance drops a little.

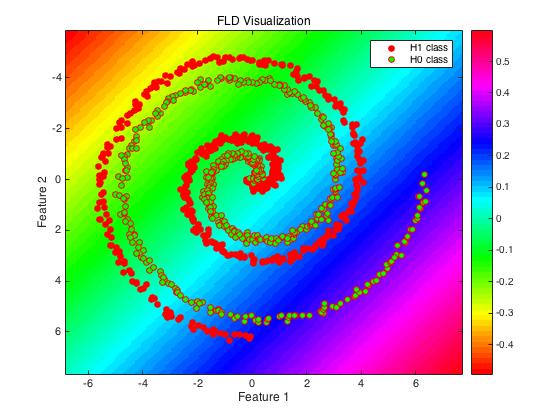
|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Classifier | KNN k=9 | Bayes | FLD | DLRT k=3 | DLRT k=5 | DLRT k=9 | DLRT k=15 |
| AUC | 0.9973 | 0.5563 | 0.5606 | 0.9961 | 0.9953 | 0.9954 | 0.9890 |

**Table.2**

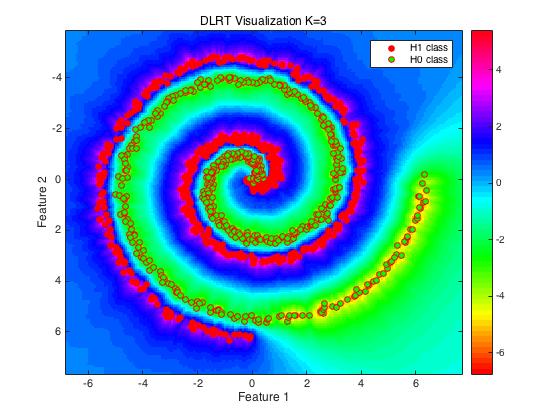
****

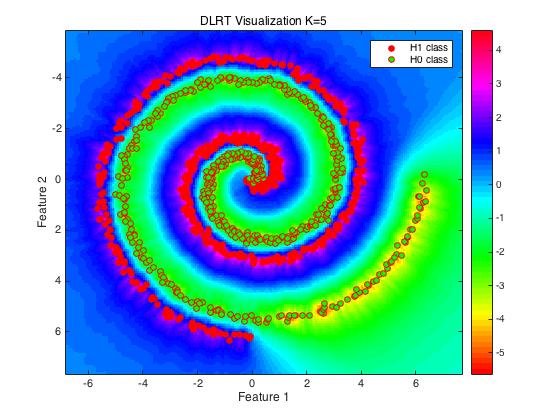
**Figure.9**

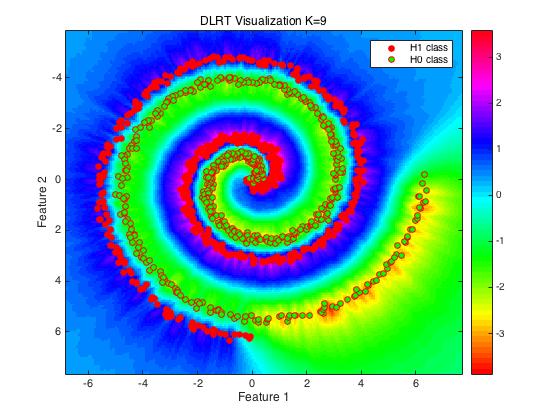
As we can see in Figure.9, the contour line does not separate two class in this case. That is also why the ROC of Bayes Classifier is close to the diagonal in Figure.8. The result does make sense because the probability distribution of test samples we assumed is Gaussian which is inconsistent with the probability distribution of spiral data. This result also supports our conclusion before: if the assumed probability distribution is not consistent with the actual probability distribution, the performance of Bayes Classifier would drop a lot. We should not use Bayes Classifier when we are not sure about the probability distribution of training and testing data.

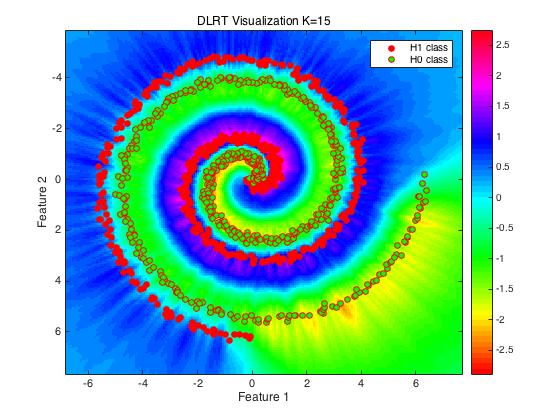
**Figure.10**

As we can see in Figure.10, none of those boundaries on the decision surface successfully separate two classes. That is why the AUC of FLD classifier is close to 0.5 in Figure.8 which means two class could not be classified by this classifier. This result does make sense because the spiral data is not linearly separable. The boundary surface between two classes is in 3-D space. So, if the training data and testing data are not linearly separable, we should not choose FLD classifier.

**Figure.11**

**Figure.12**

**Figure.13**



**Figure.14**

As we can see in Figure.11-14, when we use DLRT to deal with spiral 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, when k=15, the performance of DLRT drop a little. That is because DLRT may underfit when k increases. Even though the computational complexity of DLRT is higher, it is suitable for more kinds of data than Bayes and FLD classifiers.

1. **Blind Tests**

For blind “Gaussian” data set, I would choose Bayes classifier. Because we already know the blind data set is Gaussian, we have confidence to use Gaussian probability density function to calculate decision statistics. More than that, the computational complexity of Bayes classifier is lower than other classifiers. However, if the blind data set is not so “Gaussian” or there are some noises in data sets, the performance may drop a little. I didn’t choose FLD classifier because Gaussian data set may not be linearly separable if two classes overlap a lot. I didn’t choose DLRT classifier because the computational complexity is much higher than Bayes classifier and the performance is similar.

For blind “Spiral” data set, I would choose DLRT classifier with k=5. I chose DLRT because it is the only classifier which successfully separate two classes. The other two could not work. I chose k=5 because DLRT classifier tends to be overfitting when k is low and underfitting when k is high. More than that, higher k means higher computational complexity. So, I prefer to choose k=5.