Pattern Classification and Recognition
ECE 681

Spring 2016
Homework #3: Bayes Classifier, FLD, and DLRT

Due: 4:30 PM, Thursday, March 31, 2016[1]

This homework assignment is worth **280 points**. (280/280 = 97%)
Up to **10 points** extra credit may be earned by going above and beyond the given problem statements (*e.g.*, performing additional analyses, or providing additional insightful interpretation of the results; simply doing "more work" does not necessarily justify extra credit).

Your homework is not considered submitted until all three components (hard-copy, Matlab `.m` code, and Blind Test Results Matlab `.mat` file) have been submitted.

Submit a **hard-copy** with your plots and commentary/interpretations to the homework box in Teer.
Submit your **Matlab `.m` code** as an Attachment to the Assignment in Sakai.
Submit your **Blind Test Results** in a Matlab `.mat` file as an Attachment to the Assignment in Sakai.

Gaussian Data

Implement your own Bayes Classifier function.
Implement your own Fisher Linear Discriminant (FLD) Classifier function.
Implement your own Distance Likelihood Ratio Test (DLRT) Classifier function.
Implement your own decision surface visualization function.

Generate the following data: (HINT: The Matlab function `mvnrnd` may be helpful.)

H_0 data are 500 samples drawn from a normal (Gaussian) distribution with mean vector $\mu = [-1\ -1]^T$ and covariance matrix Σ_0 given below.
H_1 data are 500 samples drawn from a normal (Gaussian) distribution with mean vector $\mu = [2\ 1]^T$ and covariance matrix Σ_1 given below.
The covariance matrices for the H_0 and H_1 data are given by

$$\Sigma_0 = \begin{bmatrix} 1 & -0.4 \\ -0.4 & 1.5 \end{bmatrix} \qquad \Sigma_1 = \begin{bmatrix} 2 & 0.9 \\ 0.9 & 1 \end{bmatrix}$$

Generate a single realization of the data and use that ***same*** realization of the data for cross-validated training/testing for each of the following questions (*i.e.*, do not generate new data for each classifier, and do not generate new data for testing the classifier).

[1] I (Dr. Tantum) will collect homework from the homework box after 4:30PM, allowing for an appropriate grace period.
DO NOT submit late work to the locked homework box in Teer.
Late work is to be submitted to me (Dr. Tantum) in person, to my mailbox in Hudson 130, or slid under my office door (Hudson 114).
Submitted by 4:30PM, Friday, April 1, 2016 = 1 day late.
Submitted by 4:30PM, Monday, April 4, 2016 = 2 days late.
Submitted by 4:30PM, Tuesday, April 5, 2016 = 3 days late.
Late submissions not accepted after 4:30PM, Tuesday, April 5, 2016.
Work submitted in person to me (Dr. Tantum), to my mailbox in Hudson 130, or slid under my office door (Hudson 114) after the submission deadline but prior to my collecting homework from the homework box will be treated as if it were submitted on time.
Work submitted to the homework box after I have collected homework from the box will receive zero credit.

(10) 1. Visualize the decision surface for a Bayes Classifier trained on this data, and find the cross-validated ROC.

(10) 2. Visualize the decision surface for a Fisher Linear Discriminant (FLD) classifier trained on this data, and find the cross-validated ROC.

(10) 3. Visualize the decision surfaces for a Distance Likelihood Ratio Test (DLRT) classifier trained on this data for $k = \{3, 5, 9, 15\}$, and find the cross-validated ROCs.

(20) 4. Comment on whether or nor the ROCs you obtained in Questions 1, 2, and 3 make sense, considering the nature of the data and the nature of the classifiers.

(20) 5. Comment on the relative strengths and weaknesses of the Bayes Classifier, the FLD Classifier, and the DLRT when applied to this "Gaussian" data set.

(30) 6. Submit the Matlab code that produced the above results as an Attachment to the Assignment in Sakai. (We should be able to run this code to replicate your results.)

Spiral Data

Use your own implementation of the Bayes Classifier, Fisher Linear Discriminant (FLD) Classifier, and Distance Likelihood Ratio Test (DLRT) Classifier, as well as your own implementation of decision surface visualization.

Use the training data provided in the Matlab .mat file `HW03spiralDataSetFeatures.mat` for cross-validated training/testing for each of the following questions.

(10) 1. Visualize the decision surface for a Bayes Classifier trained on this data, and find the cross-validated ROC.

(10) 2. Visualize the decision surface for a Fisher Linear Discriminant (FLD) classifier trained on this data, and find the cross-validated ROC.

(10) 3. Visualize the decision surfaces for a Distance Likelihood Ratio Test (DLRT) classifier trained on this data for $k = \{3, 5, 9, 15\}$, and find the cross-validated ROCs.

(20) 4. Comment on whether or nor the ROCs you obtained in Questions 1, 2, and 3 make sense, considering the nature of the data and the nature of the classifiers..

(20) 5. Comment on the relative strengths and weaknesses of the Bayes Classifier, the FLD Classifier, and the DLRT when applied to this "Spiral" data set.

(30) 6. Submit the Matlab code that produced the above results as an Attachment to the Assignment in Sakai. (We should be able to run this code to replicate your results.)

Blind Tests

(20) 1. Select a classifier for the blind "Gaussian" data set based on your analysis of the classifiers applied to the Gaussian training data. Explain and justify why you chose the classifier (and associated classifier parameter(s)) you selected, being sure to comment on the trade-offs you considered when selecting the classifier (and any associated parameter(s)).

(20) 2. Generate decision statistics for the features provided in the Matlab .mat file:
`HW03gaussianBlindTestFeatures.mat`

Save the decision statistics to a Matlab `.mat` file, with the decision statistics stored in the vector `decStat` and saved in the same order as the blind test features. Submit the Matlab `.mat` file containing your decision statistics as an Attachment to the Assignment in Sakai.

(We know the corresponding targets for the blind test data, and will score your decision statistics to generate an ROC curve.)

(20) 3. Select a classifier for the blind "Spiral" data set based on your analysis of the classifiers applied to the Spiral training data. Explain and justify why you chose the classifier (and associated classifier parameter(s)) you selected, being sure to comment on the trade-offs you considered when selecting the classifier (and any associated parameter(s)).

(20) 4. Generate decision statistics for the features provided in the Matlab .mat file:
`HW03spiralBlindTestFeatures.mat`

Save the decision statistics to a Matlab `.mat` file, with the decision statistics stored in the vector `decStat` and saved in the same order as the blind test features. Submit the Matlab `.mat` file containing your decision statistics as an Attachment to the Assignment in Sakai.

(We know the corresponding targets for the blind test data, and will score your decision statistics to generate an ROC curve.)