HW4, ECE7252, Assigned 2/18, due 2/25, Spring 2008

1. HTF Exercise 4.2.

2. HTF Exercise 4.6, part (a).

3. Consider two sets of samples, hw4-train1.txt and hw4-train2.txt (2000 each, available from the class website), drawn form two source distributions. Design two 2-class classifiers, with LDA and logistic regression using these two set of training samples, and test the classifiers on a collection of 200 test samples, hw4-test.txt, in which the data is arranged in 10 consecutive batches, with 10 samples from class 1 followed by 10 samples from class 2 in each batch, until all samples are exhausted. It should be clear soon that both classes are not Gaussian sources. However we can still use LDA and logistic regression classifiers to classify the test data. Your job is to do the following: (a) plot the histograms of the two sets of data, one histogram for each set, on the same figure using different colors to make it easy to visualize the separation between the two classes; (b) design the LDA for the two classes based on the LDF in Eq. (4.10) in HTF, estimate the unknown means, μ_1 and μ_2 , and the common but unknown variance, σ^2, from the given two sets of training data, classify the test data, tabulate the results in a contingency table, and estimate the classification error; (c) do the same as above with the logistic regression classifier that maximizes the posterior probabilities shown in Eq. (4.19) by estimating the unknown regression parameters that maximizes the likelihood function of Eq. (4.20) which can be solved iteratively with a Newton-Raphson algorithm shown in Eq. (4. 24) following the description in Pages 98 and 99 in HTF; (d) compare the error rates on the results obtained from the two classifiers; (e) so far the classification decision process involves only one test sample per test. If we modify the process by making decisions based on multiple test samples knowing they are drawn from the same class, then can you design a different classification procedure? State your reasons why this may produce better results than those obtained when performing single-sample test; (f) try your multiple-sample decision procedure with both LDA and logistic regression, and compare the two strategies for the two classifiers.