Statistics 503 Homework 3
Due: in class on Wednesday, Feb 23 2005

Background: The next data set will be on classifying music clips as Rock or Classical based on a variety of statistics calculated on 40 second clips of songs. To do this we need to break the data into training and test samples, and learn how to compute classifiers. The software you write will be helpful for the next data assignment. The data is available on the public web site, music sample. There are 54 songs from three CDs of each type, and we’ve calculated these variables using the tuneR package in R:

- LVar, LAve, LMax: average, variance, maximum of the frequencies of the left channel
- LPer1, LFreq1: height and frequency of the highest peak in the periodogram. (Note that there is one missing value in LFreq1. Ignore this variable for this homework.)
- LFener: an indicator of the amplitude or loudness of the sound.
- LFVar: variance in the frequencies as computed by the periodogram function.

Purpose: To learn how to break data intro training and test samples, and to compute lda, qda, and logistic regression classifiers.

Exercises:

1. What is the sample size, $N$?
2. How many predictors are there, $p$?
3. Break the sample into 2/3 training, 1/3 test data sets. The code I used to do this with the olive oils data is
   \[
   \text{indx1} \leftarrow \text{sample}(\text{c}(1:nrow(\text{d.olive.Sth})), 2/3 \cdot nrow(\text{d.olive.Sth}))
   \]
   Report the indices for your training and test samples, and your R code.
4. Build an LDA rule for classifying the training data. Calculate the test error for your rule. Report your code, rule and your test error.
5. Are any of the variables not important in classifying music type? Explain your answer.
6. Build an QDA rule for classifying the training data. Calculate the test error for your rule. Report your code, and your test error.
7. Build a logistic regression rule for classifying the training data. Calculate the test error for your rule. Report your code, rule and your test error.