STAT 503 Assignment 2: Music classification

Due: in class March 9

This data was collected by Dr Cook from her own CDs. Using a Mac she read the track into the music editing software Amadeus II, snipped and saved the first 40 seconds as a WAV file. (WAV is an audio format developed by Microsoft, commonly used on Windows but it is getting less popular.) These files were read into R using the package tuneR. This converts the audio file into numeric data. All of the CDs contained left and right channels, and variables were calculated on both channels. The resulting data has 57 rows (cases) and 72 columns (variables).

- LVar, LAve, LMax, RVar, RAve, RMax: average, variance, maximum of the frequencies of the left and right channels, respectively.
- LPer1-LPer15, LFreq1-LFreq15, RPer1-RPer15, RFreq1-RFreq15: height and frequency of the highest peak in the periodogram.
- LFEn, RFEn: an indicator of the amplitude or loudness of the sound.
- LFVar, RFVar: variance in the frequencies as computed by the periodogram function.

There are 30 tracks by Abba, the Beatles and the Eels, which would be considered to be Rock, and 24 tracks by Vivaldi, Mozart and Beethoven, considered to be Classical. Your main job is to build a classifier to accurately predict tracks as Rock or Classical. One of the evaluation methods will be how accurately your classifier can predict 5 new tracks, from the same CDs that I have omitted from this data. Here are some additional points to think about:

1. There are more variables than cases, so a good classifier will not use all the variables. You’ll need to find a subset of the most important variables.
2. There are a few missing values. You’ll need to think about how to handle these.
3. There are 3 tracks from an Enya CD, which would be considered to be new wave music, which you should not use to build your classifier. I’m curious how these tracks compare with rock and classical tracks. Are they more like one or the other, or do they have entirely different characteristics?

This assignment requires you to construct your own report on this data with the members of your working group. Follow the format of the case studies. Write down a plan for the analysis ahead of time before, doing any plots or summaries. Along with building a good classifier you should report any other interesting patterns and unusual or unexpected details about the data. You will need to present your results in tabular and graphical form, and concisely summarize the major findings. In keeping with doing reproducible research, you should also submit your R code and modified data to the assignment page in WebCT, so that I can re-do your analysis. Your report needs to be handed in on paper, though. The grade for the assignment will depend on neatness of the report, how comprehensive is the analysis, clarity of results presentation, and preciseness of final conclusions. The assignment grading policy on the web is a guide for how the assignment will be graded.