ETC3250 Lab 8

Di Cook

15 September 2015

Supervised classification, and data visualisation

Task 1

Read in the music data, from the ggobi web site:

1. Subset the data to drop the "Enya" class.

2. Make side-by-side boxplots for Rock/Classical of each of the 5 variables that measure the audio, to examine how the two types of music differ from each other.

3. Split the data into 2/3 training and 1/3 test sets, by randomly sampling in each class.

Task 2

1. Compute a logistic regression classifier to distinguish "Rock" from "Classical" music.

2. Predict the error for the training set, and also the test set.

3. Make a plot that shows the predicted values from the model fit, facetted by music type.

4. Highlight the missclassified cases on a parallel coordinate plot of the 5 variables.