

Multivariate Statistics: Exercise 10

December 18, 2014

Discriminant analysis (>2 groups):

Consider the data `olives` from the package `classifly`, see last exercise. We discriminate the data based on the fatty acid concentrations (do not use `eicosenoic`) into the three regions (variable `Region`).

- (a) Write a function for Fisher's discriminant analysis for the case of more than two groups. Return the matrix \mathbf{V} and Fisher's discriminant values.
- (b) Use all available observations for the computation of the discriminant rule, and assign afterwards the same observations to the groups according to their discriminant values. Take as estimations of the prior probabilities the relative frequencies of the groups. Compute the misclassification rate.
- (c) Project the data on the first two columns of \mathbf{V} and plot the result. Use different colors for the true group memberships, and different symbols for the predicted memberships.
- (d) Discriminate now according to the 9 groups of variable `Area` and proceed according to (b). Use the function `matchClasses()` from the `library(e1071)` to compute the misclassification error.
- (e) Project the data as in (c). How does the misclassification error change if the Fisher discriminant values are only computed with the first two columns of \mathbf{V} ?

Save your (successful) R code together with short documentations and interpretations of results in a text file, named as *Familyname12.R*. Send the file as an email attachment to *mehmet.mert@tuwien.ac.at*, at latest Tuesday (16.12)