DEPARTMENT OF MATHEMATICS
FACULTY OF SCIENCE
KU LEUVEN

KU LEUVEN

Report on project about clustering

Statistical Data Analysis

Project 2

Mitja Mandić

May 2022

1 Introduction

For the second project for the course Statistical Data Analysis we are once again worikng with the dataset of spectral data of four cultivars of canteloupe melons.

Groups we work in this report with were also again chosen randomly, with 50 observations drawn from each of the groups. Below in INSERT REFERENCE we see a figure of spectral plots of each of the groups. We see that groups 1, 2 and 4 differ only slightly in lower wavelength numbers with group 4 having more variability there, but are quite similar in their behaviour in higher frequencies. Group 3 is the one that clearly stands out; a part of it resembles the behaviour of other three groups, while some observations form a different pattern.

We predict that this part of group three will form a separate cluster. Another will possibly be formed by lower wavelengths of group 4 – other observations seem to be too similar to form different clusters.

2 K-medoids clustering

3 Hierarchical clustering

4 Conclusion