

Lab 4

Gustav Sternelöv

September 29, 2016

Assignment 1

1. Open *olive.csv* in GGobi and open Data Viewer. How many observations are present in the data?

By using the tool *Data viewer* in GGobi it is easy to look up that there are 572 observations in the data set.

2. Create a scatter plot matrix that shows how the contents of different acids are related to each other. Investigate the matrix to find plots where the clusters are present. Close the plot.

3. Create a scatter plot of the eicosenoic against linoleic. Based on section 2, comment why it can be interesting to investigate this pair of variables. You have probably found a group of observations having unusually low values of eicosenoic. Use identification tool to find out the exact values of eicosenoic for these observations.

They take the value 1, 2 or 3.

4. Create a histogram that shows how many observations fall within any given region. Use persistent brushing to identify the regions that correspond unusually low values of eicosenoic. Include the plots into your report and then remove the brushing (one way is to restart GGobi)

5. Create scatter plots eicosenoic against linoleic and arachidic against linolenic. Which outliers in (arachidic, linolenic) are also outliers in (eicosenoic, linoleic)? Are outliers grouped in some way?

Hard to say exactly what values are outliers. Guessing that the values coloured in yellow in the arachidic versus linolenic scatter plot are the outliers. All the outliers has low values for eicosenoic.

6. Use persistent brushing to paint by different colors the observations that fall into different regions. Keep these coloring during steps 7-9.

7. Create a parallel coordinate plot for the available eight acids. Select some proper subset of variables and define their order on the plot. Which variables can be taken for identifying clusters? (suggest at least three variables)

Oleic and Eicosenoic good for finding clusters. Linoleic and Palmitoleic can also be used for that purpose.

8. Create a 3D-rotation plot by using the variables found in step 7. Can you see clusters? Include proper screenshots motivating your answer.
9. Use all 8 acids and examine a 2D-tour. Try to find a projection with the best separation of the data into clusters. How the clusters detected are related to the regions the oils come from?
10. Based on the analysis above, try to suggest a strategy (or, maybe, several strategies) that would use information about the level of acids to discover which region the oil comes from.

Assignment 2