MSc course at FHWN, January-February 2023
Exam project (50% of final grade)

Purpose

To demonstrate proficiency in performing PCA and PLS-DA on metabolomics data, handling outliers and interpreting results.

Problem

To test if it is possible to develop models capable of differentiating between milk samples from dairy cows in traditional and organic farming or from farms delivering milk to specific dairies. Milk samples are collected using automated milking in the normal routine at the farm.
Milk samples were analysed for metabolites using ^{1}H NMR spectroscopy.

Data

The dimension of the data structure is 89x34.
The first row is metabolite id.
The first column is a sample id, 2nd dairy and 3rd organic/conventional farming. Columns 4-34 are metabolites.

File name: dataforexamproject.xlsx

The following questions need to be answered:

1. PCA (40%)
    a. Describe steps taken to reach the final PCA model
    b. Interpret the final PCA model
    c. Can we use the model to classify dairies or organic/conventional farming?
    d. Can we identify differences in milk metabolome due to dairies or farming method?

2. PLS-DA (40%)
    a. Describe steps taken to reach the final PLS-DA models
    b. Interpret the final PLS-DA models
    c. Can we use the model to classify dairies or organic/conventional farming?
    d. Can we identify differences in milk metabolome due to dairies or farming method?
    e. Is the interpretation the same comparing with the PCA?
    f. What is the advantage of using PLS-DA and what do we need to ensure when using PLS-DA?

3. OPLS-DA on organic/conventional farming (20%)
    a. Describe steps taken to reach the final OPLS-DA model
    b. Interpret the final PLS-DA model
    c. Can we identify differences in milk metabolome due to farming method?
    d. Is the interpretation the same comparing with the PCA and PLS-DA?
    e. What is the advantage of using OPLS-DA over PLS-DA?

Remember to include where relevant which samples and variables are used, how data is scaled, and show relevant scores and loadings plots or other diagnostic plots.

What to hand in:

A word file with relevant plots of figures and short descriptions of each item 1-3 and .R file(s) for 1-3.