Practicals single block analyses

Nutrimouse dataset

```
library(CCA)
data("nutrimouse")
```

The data sets come from a nutrigenomic study in the mouse (*Martin et al., 2007*, https://doi.org/10.1002/hep.21510) in which the effects of five regimens with contrasted fatty acid compositions on liver lipids and hepatic gene expression in mice were considered.

Two sets of variables were acquired on forty mice

- **genes**: expressions of **120 genes** measured in liver cells, selected (among about 30,000) as potentially relevant in the context of the nutrition study. These expressions come from a nylon macroarray with radioactive labelling
- **lipids**: concentrations (in percentages) of **21 hepatic fatty acids** measured by gas chromatography

Biological units (mice) were cross-classified according to two factors experimental design (4 replicates)

- **genotype**: **2-levels factor**, wild-type (WT) and PPARα -/- (PPAR)
- **diet**: **5-levels factor**. Oils used for experimental diets preparation were corn and colza oils (50/50) for a reference diet (REF), hydrogenated coconut oil for a saturated fatty acid diet (COC), sunflower oil for an Omega6 fatty acid-rich diet (SUN), linseed oil for an Omega3-rich diet (LIN) and corn/colza/enriched fish oils for the FISH diet (43/43/14)

PCA analysis for genes and lipids datasets

Question 1: Perform PCA and investigate variances, sample distribution and variable relationship with plots.

```
library(ropls)
opls()
```

- Investigate distribution of data
- perform PCA

Question 2: Plot scree plot

```
PCA_res@pcaVarVn
```

- plot explained variance

Question 3: observe the samples distributions in the space of the dimensions, what are the main sources of variation?

```
PCA_res@scoreMN
```

- plot scores on Dim.1 vs Dim.2 and Dim.3 vs Dim.4 with percentages of explained variance on axes

Question 4: which variables are responsible of the samples distribution?

```
PCA_res@loadingsMN
```

- plot loadings on Dim.1 vs Dim.2 and Dim.3 vs Dim.4 with percentages of explained variance on axes

PLS between two matrices

Question 1: perform PLS canonical analysis

- compute pls model

```
library(mixOmics)
pls(…,mode=”canonical")
```

Question 2: Samples distribution in the new reference (rotated axes) for each of the two blocks

- plot local scores based on genes and lipids

```
PLS_res$variates$X and PLS_res$variates$Y
```

Question 3: Variables contribution in each data block to each dimension

- plot variables loadings

```
PLS_res$loadings$X and PLS_res$loadings$Y
```

Question 4: perform PLS regression analysis

- compute pls model

```
pls(…,mode="regression")
```

Question 5: observe the differences between the two modes

- plot local scores

```
PLS_res_reg$variates$X and
PLS_res_reg$variates$Y
```

PLS discriminant analysis

Question 1: based on genes data, can we discriminate wt vs ppar samples ?

- perform PLS-DA using ropls package

```
library(ropls)
opls(x = nutrimouse$gene,
     y = metadata$genotype,
     predI = NA,
     permI = 100)
```

Question 2: Is the model statistically significant?

- observe the results of permutations
- observe Q^2, and R^2Y values

```
PLS_res@modelDF
```

Question 3: Show the distribution of samples in the space of the latent variables

- plot scores

```
PLS_res@scoreMN
```

Question 4: observe the variables distributions in the space of the latent variables

- plot loadings of the optimal model

```
PLS_res@loadingsMN
```

Question 5: which variables are discriminant for genotype?

- plot Variables Importance in Projection of the model

```
PLS_res@vipVn
```

OPLS discriminant analysis

Question 1: based on genes data, can we discriminate wt vs ppar samples ?

- perform OPLS-DA using ropls package

```
library(ropls)
opls(x = nutrimouse$gene,
     y = metadata$genotype,
     predI = 1,
     orthoI = 1,
     permI = 100)
```

Question 2: Is the model statistically significant?

- observe the results of permutations
- observe Q^2, and R^2Y values

```
OPLS_res@modelDF
```

Question 3: Show the distribution of samples in the space of the latent variables

- plot scores

```
OPLS_res@scoreMN
OPLS_res@orthoScoreMN
```

Question 4: observe the variables distributions in the space of the latent variables

- plot loadings of the optimal model

```
OPLS_res@loadingsMN
OPLS_res@orthoLoadingsMN
```

Question 5: which variables are discriminant for genotype?

- plot Variables Importance in Projection of the model

```
OPLS_res@vipVn
```

CCA

1. Perform CCA (mixOmics::rcc) between 20 genes and all lipids. Investigate correlations, variable relationship and sample distribution with plots.

- Plot the scores
- Plot the loadings

2. Perform CCA with scaled datasets and observe the difference with previous result

- Plot the scores
- Plot the loadings

3. Perform regularized CCA with all genes and lipids, discuss results

- Plot the scores
- Plot the loadings