# Predictive Analytics

## TP 1 & 2

The purpose of this practical lecture is to create a Matlab function able to analyze a set of multivariate data grouped into classes.

For all of the data and for each group, this function will be able to provide a description of each variable and to detect the links between variables. It will also be able to represent data in two or three dimensions retaining the maximum information on the differences between groups.

### A. Individual Analysis of variables:

1. The function must provide and display the number of variables for all data and for each group, the number of data and the corresponding histograms for each variable.
2. For each group, the function should provide the position and dispersion criteria.

### B. Analysis of linear correlations between variables:

Note: This analysis must be done for all data and for each group.

1. The function will display for each pair of variable the plot (cloud of points) corresponding to the data in this space. Each point must to be displayed in a color that is characteristic of the label.
2. The function must to provide the correlations between the variables (r2) for each group of data.
3. For each significant correlation, the function must draw the corresponding regression line (another color) on the scatterplot (use a3). It should also display the value of r ² on the regression line.

### C. Linear Discriminant Analysis

1. The function must display on a new figure a projection, in two dimensions, of the dataset using Linear Discriminant Analysis. Each point have to be displayed in a color that is characteristic of group membership (label).
2. The function must also display on a new figure, the "circle of correlation" between the old and new variables. Both axes are scaled between -1 and 1.

### D. Use of the function:

**1.** Import the **hepta** data from Hepta.mat (http://lipn.univ-paris13.fr/~grozavu/DataMining/datasets/). This dataset contains 7 groups in three dimensions. Use these data to test your function.

**2.** Apply the function on the **iris** and **wine** data. Deduce the characteristics of each group and differences between the groups. Compare projection obtained with the PCA and LDA, explain why there are differences.