TP PCA

The Principal Component Analysis is a tool used in visualization and compression. It has been implemented in Python with the module sklearn.decomposition.

1) Open the vehicules.data file. Check the characteristics of the data set (number of objects, number of attributes), normalized the data and check that the normalization did work.

2) Execute the function PCA available in the module sklearn.decomposition.

3) Get the eigenvalues obtained and plot the percentage of information against the new axes. Specify the number of factors you will retain.

For the rest of the TP, 2 factors are retained in order represent objects and variables in a 2D plan.

4) Plot objects in the first factorial plan.

5) Plot the correlation circle. Unfortunatly, sklearn does not provide the tool. Measure first the correlation (np.corrcoef) between the original variables and new ones. Then, plot in the same figure a circle and the correlations for the two first new axis.

6) What is observable ?