The aim of this homework is to perform an **unsupervised** analysis of the *Breast Cancer Wisconsin (Diagnostic)* DataSet. You have to analyze it, try and combine several transformations and dimension reduction techniques before applying clustering, evaluate the results.

The data set can be obtained using

```
from sklearn.datasets import load_breast_cancer
cancer = load_breast_cancer()
```

and more informations about it can be obtained at the kaggle site. You can also find on this site a lot of Notebooks related with this data set, but mostly for *supervised* learning (which is not asked).

The tasks you have to perform are (not limitative list...)

1. find meaningful representations of the original data set allowing to better understand it,
2. find meaningful representations of the transformed and reduced data set,
3. apply clustering on the original data set and on the transformed data set,
4. evaluate the clustering obtained by comparaison with the `Diagnosis` feature. Are you able to discover if the cancer is `malign` or `benign` ?

Observe that you should not used the `Diagnosis` feature when performing unsupervised learning but rather try to recover it using the other features !

The results of your analysis have to be deposited on moodle as a **Notebook** before December 12.

Serge Iovleff