Lab dimension reduction n°1

Exercice 1 (The iris dataset)

1. Load the iris dataset (from **sklearn.decomposition**). This dataset contains two important values:
    a. The data: The description of the iris flower (4 columns).
    b. The target: The iris species which each iris flower belongs to.

2. Plot in a 3D scatter figure the first 3 columns of the dataset colored with the species of each individual (Contained into dataset.target).
   In order to plot in 3D figure you can use the **Axes3D** of the **mpl_toolkits.mplot3d**.

3. Apply the PCA decomposition to the data.

4. Plot the same 3D figure as done before using the pca-transformed data. What do you see ?

Exercice 2

1. Download the dataset “crowdedness-at-the-campus-gym” from github.

2. Compute the correlation matrix between the features. Which covariance is the most important? Is it logic?

3. After applying PCA and computing the individual explained variances, how many dimensions will we hold on in order to explain at least 85% of the dataset variance.

4. **Bonus question:**
   Using can you create a method that can allow you to guess how many of this gym's members are regular and how many aren't ?