

TP1
Rappels ACP - ACI

Exercice 1[Python]

On va comparer deux méthodes d'Analyse factorielles l'ACP et l'ACI. On travaille en Python et on utilisera les fonctions pré-programmées de `scikit learn` :

<https://scikit-learn.org/stable/modules/classes.html#module-sklearn.decomposition>

Pour illustrer ces méthodes on considère le jeu de données `load_digits` disponible dans `sklearn.datasets`.

1. Importer le jeu de données. Visualiser rapidement les variables disponibles. Isoler une matrice contenant les différentes images (stockées sous forme de vecteurs de pixels) et la tracer pour quelques individus (voir `plt.matshow`).
2. Regarder l'aide de la fonction `PCA()` et l'appliquer aux données. Tracer le graphe des valeurs propres, combien d'axes sélectionner ? Quelle part de variance expliquent-ils ?
3. Tracer dans le premier plan factoriel le nuage des individus, on utilisera la fonction `plt.scatter` pour colorier les individus en fonction des différentes classes. Commenter les premiers axes.
4. Tracer les images des deux premières colonnes de B dans la décomposition $X = FB$. Que représentent ces images et qu'observe-t-on ?
5. Tracer pour un individu l'image filtrée par ACP et la comparer à l'image initiale en faisant varier le nombre d'axes utilisés. Commenter.
6. On va maintenant, pour le même jeu de données, effectuer une ACI. Regarder l'aide de `FastICA` et l'appliquer aux données.
7. Proposer une méthode pour déterminer un nombre adéquat de facteurs.
8. Tracer les individus dans les premiers plans factoriels et commenter.
9. (*Facultatif*) Reprendre les questions précédentes en appliquant une NMF.

Exercice 2[R]

Dans cet exercice on va voir un exemple où l'ACI donne des résultats meilleurs que l'ACP. Pour faire une ACP et une ACI en R on utilise les librairies `FactoMineR` et `fastICA`. On va considérer le jeu de données suivant : deux signaux provenant de 2 sources différentes qui ont été mélangés. Le signal initial est

```
S <- cbind(sin((1:1000)/20), rep((((1:200)-100)/100), 5))
```

que l'on recentre et re-normalise et qu'on mélange via la matrice

$$A = \begin{pmatrix} 0.99 & -1 \\ -1 & 1 \end{pmatrix}.$$

On travaille avec le signal $X = SA$ à partir duquel on veut récupérer les deux sources S . Comparer les reconstructions issues de l'ACP et de l'ACI. Commenter.