

Dans cet atelier, on va appliquer l'analyse en composante principale sur deux problèmes. Le premier, une application from scratch de l'ACP sur la base de données Iris. Le deuxième problème traite le problème de réduction de dimensions sur les images.

### Problème 1 : ACP from scratch sur Iris

Le but de l'étude est de comparer les différents types des fleurs iris. L'analyse en composantes principales (ACP) sur ce tableau devra préciser les points suivants. Il s'agit en outre de

- réaliser une typologie des fleurs iris, c'est-à-dire celles qui se ressemblent et celles qui diffèrent ;
- proposer un bilan des liaisons entre les variables et, dans la mesure du possible, résumer approximativement l'ensemble des variables par un petit nombre de variables synthétiques, c'est-à-dire non pas extraites du tableau mais les combinant ;

#### Question de cours :

1. Comment mesure-t-on la dissimilitude entre les individus ?
2. Le nuage est toujours centré. Pourquoi ? Quel est l'effet du centrage sur l'analyse du nuage des individus ? Quel type de liaisons évalue-t-on dans le nuage des variables ?
3. Montrer que si on ne réduit pas les variables, on accorde aux variables un poids égal à leur écart-type. La liste des écarts-types permet-elle, a priori, de prévoir une différence significative entre une ACP normée et une ACP non-normée ?

#### Analyse en composantes principales :

1. Importer les données à partir de la bibliothèque seaborn or pandas.
2. Créer une fonction centrer qui centre les données. tester la fonction sur le data
3. Créer la fonction réduite qui retourne la matrice des données réduites.
4. Calculer et interpréter la matrice des corrélations. Quelles sont les variables les plus corrélées ?
5. Analyser les valeurs propres (np.linalg.eig). Quel est le pourcentage d'inertie expliquée par chaque axe ? Combien d'axes peut-on conserver ?
6. Tracer les valeurs propres principales et les interpréter.
7. Quelles sont les composantes principales ? Étudier la représentation des individus sur le premier axe et le second axe . Quels sont les points dont la contribution est la plus grande ? Quelles sont leurs coordonnées ?
8. Visualiser les fleurs sur le plan 2d contenant les deux axes principaux.

### Problème 2 : EigenFaces

L'objectif de cette partie est d'explorer l'utilisation de l'analyse en composantes principales pour la reconnaissance faciale.

utiliser la base de données *fetch\_lfw\_people* qui se trouve dans le package sklearn.datasets.