

## Analyse en Composantes Principales (ACP)

### Objectifs du TP

Ce travail pratique vise à vous familiariser avec l'Analyse en Composantes Principales (ACP), une méthode statistique de réduction de dimension. À la fin du TP, vous serez capables de :

- Comprendre les principes fondamentaux de l'ACP ;
  - Appliquer l'ACP à un jeu de données réelles ;
  - Interpréter les résultats sous forme graphique et numérique ;
  - Identifier les structures sous-jacentes dans les données.
- 

### Contexte

Vous travaillez avec un jeu de données multivariées contenant des observations (individus) décrites par plusieurs variables quantitatives. L'objectif est de résumer l'information contenue dans ces variables de manière visuelle et interprétable, tout en réduisant la complexité du jeu de données. Le dataset est celui de decathlon.csv partagé par email.

---

### Travail à réaliser

#### 1. Préparation des données

- Identifier les variables quantitatives à inclure dans l'analyse.
- Vérifier si les données nécessitent un centrage et une réduction.
- Justifier les choix faits à cette étape.

#### 2. Lancer l'ACP

- Réaliser une ACP sur les données préparées.
- Choisir un nombre de dimensions à conserver et motiver ce choix.

#### 3. Analyse de l'inertie

- Analyser l'inertie expliquée par chaque axe principal.
- Présenter le pourcentage de variance cumulée et justifier le nombre d'axes retenus pour l'interprétation.

#### 4. Étude des variables

- Interpréter la contribution des variables aux axes principaux.
- Représenter les variables dans un cercle de corrélation.
- Identifier les groupes de variables corrélées entre elles.
- Repérer les variables les plus représentatives de chaque axe.

#### 5. Étude des individus

- Analyser la position des individus dans les plans factoriels.
- Identifier des regroupements ou oppositions entre individus.
- Déterminer les individus qui contribuent le plus à chaque axe.

#### 6. Qualité de représentation

- Étudier la qualité de représentation des variables et des individus dans le plan factoriel.
- Identifier ceux qui sont bien ou mal représentés, et discuter les limites de l'interprétation.

#### 7. Synthèse des résultats

- Proposer une interprétation globale des axes.
- Faire un lien entre les résultats obtenus et les caractéristiques observées dans le jeu de données.
- Résumer les principaux enseignements de l'analyse.

## Analyse en Composantes Principales (ACP) et classification

### Objectifs:

- Comprendre les principes de l'Analyse en Composantes Principales (ACP)
  - Appliquer l'ACP à un jeu de données haute dimension
  - Visualiser les données projetées dans l'espace des composantes principales
  - Utiliser un algorithme de classification (k-NN) avant et après ACP
  - Comparer les performances avec et sans réduction de dimension
- 

### Données utilisées :

- **Jeu de données** : digits (fourni par sklearn.datasets)
  - Chaque observation représente une image 8×8 d'un chiffre manuscrit (0–9)
  - 64 variables par observation (8×8 pixels à niveaux de gris)
- 

### Consignes :

#### 1. Chargement et visualisation des données

- Charger le jeu de données digits
- Afficher quelques images avec leur étiquette
- Afficher la distribution des classes

#### 2. Prétraitement

- Normaliser les données
- Séparer en ensembles d'entraînement (70%) et de test (30%)

#### 3. Classification sans ACP

- Entraîner un classifieur **de votre choix** sur les données originales
- Évaluer la précision sur le test

#### 4. Réduction de dimension avec ACP

- Appliquer l'ACP sur les données d'entraînement
- Choisir un nombre de composantes **k** qui explique au moins **90% de la variance**

- Transformer les données d'entraînement et de test

## 5. Classification après ACP

- Réentraîner le même classifieur sur les données réduites
- Évaluer la précision et comparer aux résultats précédents

## 6. Visualisation

- Tracer un nuage de points (scatter plot) des **2 premières composantes** principales
- Colorer les points selon la classe réelle pour observer les regroupements

---

**Livrables attendus :** À soumettre via [o.banouar@uca.ac.ma](mailto:o.banouar@uca.ac.ma)

- Un notebook googlecolab ou un fichier Python bien commenté
- Rapport intitulé `votre_nom_prénom_lab2.pdf` contenant
  - Le lien vers le code source commenté
  - Les prises d'écran des résultats obtenus avec leurs codes correspondant et commentaires
  - Graphiques et tableaux des résultats
  - Analyse comparative des algorithmes