



# Méthodes de placement multidimensionnelles

Fabrice Rossi

Télécom ParisTech

Introduction

Analyse en composantes principales

Modèle

Qualité et interprétation

Autres méthodes

## Introduction

### Analyse en composantes principales

Modèle

Qualité et interprétation

### Autres méthodes



- intérêt des outils de visualisation
  - très efficaces en dimension 2
  - corrects en dimension 3
  - utilisables en dimensions 4 ou 5 (après un apprentissage)
- limites
  - dilemme lisibilité versus nombre de variables (et d'objets)
  - liens entre variables difficiles à comprendre
  - liens entre objets difficiles à comprendre



- intérêt des outils de visualisation
  - très efficaces en dimension 2
  - corrects en dimension 3
  - utilisables en dimensions 4 ou 5 (après un apprentissage)
- limites
  - dilemme lisibilité versus nombre de variables (et d'objets)
  - liens entre variables difficiles à comprendre
  - liens entre objets difficiles à comprendre
- solution
  - réduire **automatiquement** la dimension des données
  - en enlevant des variables
  - ou **en calculant de nouvelles coordonnées**

- formulation du problème :
  - données  $(X_i)_{1 \leq i \leq N}$  dans un espace  $\mathcal{X}$
  - placement  $(Y_i)_{1 \leq i \leq N}$  dans  $\mathbb{R}^Q$  avec  $Q$  petit (2 ou 3)
  - à chaque objet  $X_i$  est associé un vecteur en basse dimension  $Y_i$
  - problème : bien choisir les  $Y_i$
- deux sources de variabilité dans les solutions
  - représentation initiale (nature) des  $X_i$
  - critère de choix des  $Y_i$
- questions additionnelles
  - lien entre les  $X_i$  et les  $Y_i$
  - contrôle de la qualité du placement
  - interprétation

- $X$  est une matrice  $N \times P$ 
  - $N$  objets
  - $P$  variables numériques
- Critère d'approximation
  - $Y$  est de dimension  $Q < P$  : on considère les  $Y_i$  comme des éléments d'un sous-espace vectoriel de  $\mathbb{R}^P$
  - critère de qualité quadratique

$$\frac{1}{N} \sum_{i=1}^N \|X_i - Y_i\|^2$$

- **attention** : les  $Y_i$  sont décrits dans une base adaptée
- problème d'optimisation : choisir les meilleurs  $Y_i$  **sous une contrainte de rang**

- les  $X$  sont décrits seulement par une dissimilarité entre les objets,  $d_X(X_i, X_j)$
- Critères d'approximation
  - les  $Y$  sont dans  $\mathbb{R}^p$  :  $d(Y_i, Y_j) = \|Y_i - Y_j\|$
  - essayer d'avoir  $d(Y_i, Y_j) \simeq d_X(X_i, X_j)$  : principe de **préservation des distances**
  - critère générique

$$\frac{1}{N^2} \sum_{i=1}^N \sum_{j=1}^N (d_X(X_i, X_j) - \|Y_i - Y_j\|)^2 F(d_X(X_i, X_j), \|Y_i - Y_j\|)$$

- le rôle de  $F$  est de limiter l'influence des grandes distances (par exemple)
- d'autres variantes existent



Introduction

Analyse en composantes principales

Modèle

Qualité et interprétation

Autres méthodes



- méthode de placement pour données numériques inventée par Pearson (1901)
- approche par projection linéaire
- idée sous-jacente (modèle factoriel)
  - processus génératif

$$X = YW$$

- $W$  est une matrice  $Q \times P$  **orthogonale**  $WW^T = I$
- $Y$  est une représentation de dimension  $Q$
- but de l'analyse factorielle : retrouver  $W$  et  $Y$

- soit un système orthonormé  $(\phi_j)_{1 \leq j \leq Q}$  de  $\mathbb{R}^P$ , la projection d'un  $X_i$  sur le sous-e. v. associé est

$$P(X_i) = \sum_{j=1}^Q (X_i \phi_j) \phi_j^T$$

- erreur de projection

$$\frac{1}{N} \sum_{i=1}^N \left\| X_i - \sum_{j=1}^Q (X_i \phi_j) \phi_j^T \right\|^2$$

- meilleure projection : optimisation sur  $(\phi_j)_{1 \leq j \leq Q}$

- on suppose les données centrées :  $\sum_{i=1}^N X_i = 0$
- on note  $\Phi$  la matrice des  $\phi_j$  (en colonnes)
- l'erreur de projection devient

$$\frac{1}{N} \sum_{i=1}^N \left\| X_i - X_i \Phi \Phi^T \right\|^2$$

- minimiser l'erreur revient à résoudre

$$\begin{aligned} &\text{maximiser } \frac{1}{N} \text{Tr}(X \Phi \Phi^T X^T) \\ &\text{avec } \Phi^T \Phi = I \end{aligned}$$

- solution (admise) : les  $\phi_j$  sont les vecteurs propres de  $\frac{1}{N} X^T X$  associés aux  $P$  plus grandes valeurs propres

- on a donc

$$X \simeq YW$$

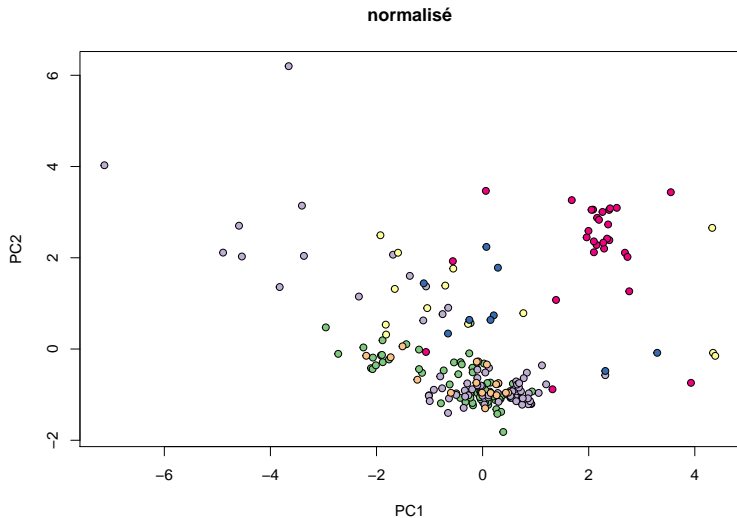
avec  $W = \Phi^T$  et  $Y = X\Phi$

- les colonnes de  $\Phi$  sont les axes principaux de l'ACP
- les colonnes de  $Y$  sont les composantes principales
- la variance de  $Y_{.i}$  est  $\lambda^i$ , la  $i$ -ème valeur propre de  $\frac{1}{N}X^T X$
- placement :
  - on conserve les deux premiers axes
  - on affiche les deux premières composantes



- autre vision de l'ACP : chercher une direction "intéressante" dans les données
- intéressant : avec une forte variabilité
- analyse similaire à la précédente :
  - axe de projection  $\phi$ , projection  $X\phi$
  - variance de la projection  $\frac{1}{N} X\phi\phi^T X^T$  (données centrées)
  - maximisation de la variance  $\Rightarrow$  même problème que précédemment !
- les axes principaux sont donc des axes de variance maximale

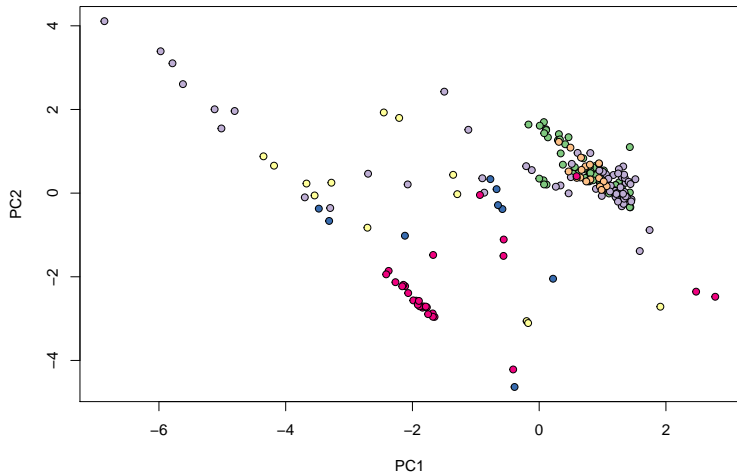
- données verre :
  - 213 observations sur 10 variables
  - 9 variables numériques : pas de prise en compte de la variable nominale
- données iris :
  - 150 observations sur 5 variables
  - 4 variables numériques (et une nominale)
- centrage et réduction :
  - centrage : obligatoire
  - réduction : au choix de l'analyste

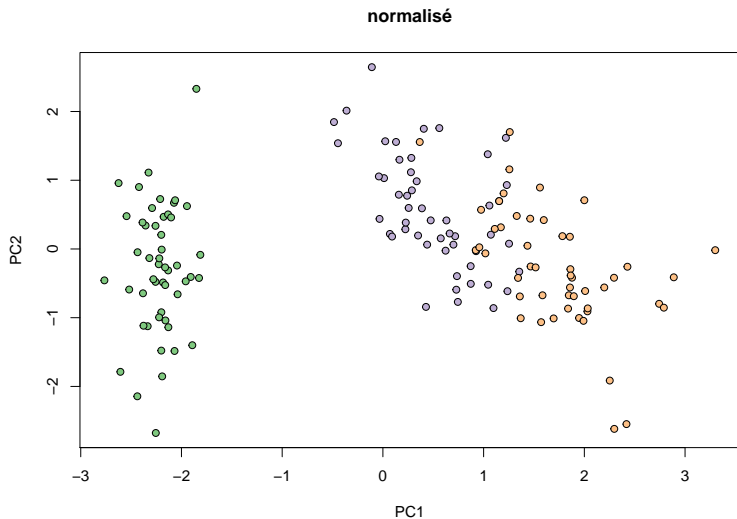


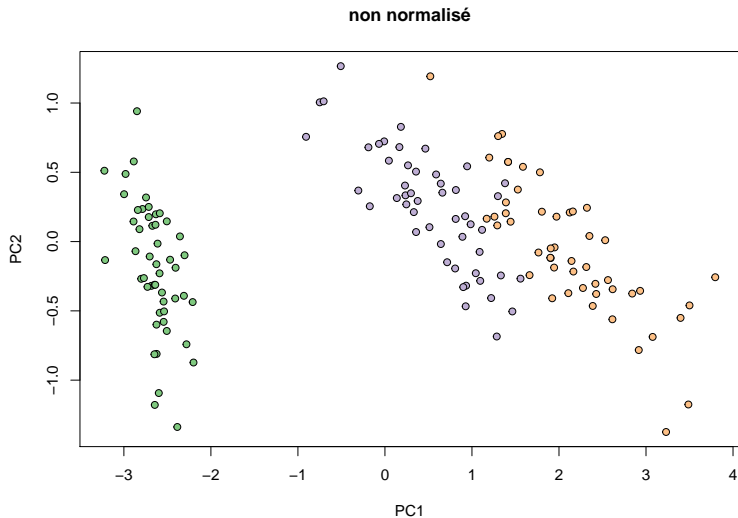




non normalisé







- l'ACP est une **projection linéaire**
- les distances sont donc réduites :
  - si deux points sont proches dans  $\mathbb{R}^P$  ils seront toujours proches dans  $\mathbb{R}^Q$  par ACP
  - **mais** deux points proches dans le placement ACP ne sont pas nécessairement proches dans l'espace d'origine
- les séparations sont “exactes”
- les regroupements peuvent être trompeurs



## Qualité de l'approximation

### ■ Erreur de reconstruction

$$\frac{1}{N} \text{Tr}(X^T X) - \frac{1}{N} \text{Tr}(X \Phi \Phi^T X^T)$$

### ■ or

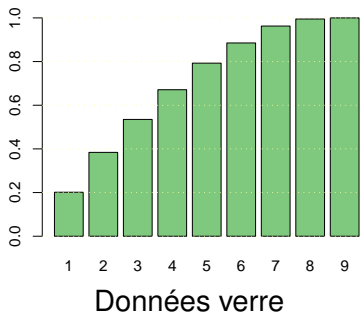
- $\frac{1}{N} \text{Tr}(X^T X) = \sum_{i=1}^P \lambda_i$  est la somme des variances des variables
- $\frac{1}{N} \text{Tr}(X \Phi \Phi^T X^T) = \frac{1}{N} \text{Tr}(\Phi^T X X^T \Phi) = \sum_{i=1}^Q \lambda_i$

### ■ Qualité : pourcentage de la variance expliquée

$$\frac{\sum_{i=1}^Q \lambda_i}{\sum_{i=1}^P \lambda_i}$$

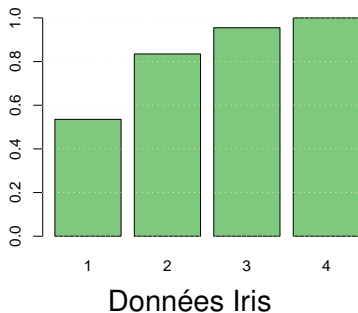
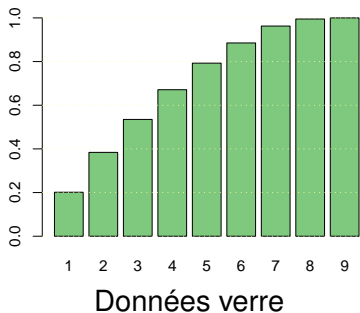


# Représentation graphique





## Représentation graphique



L'approximation est bien meilleure pour les données Iris en deux dimensions.

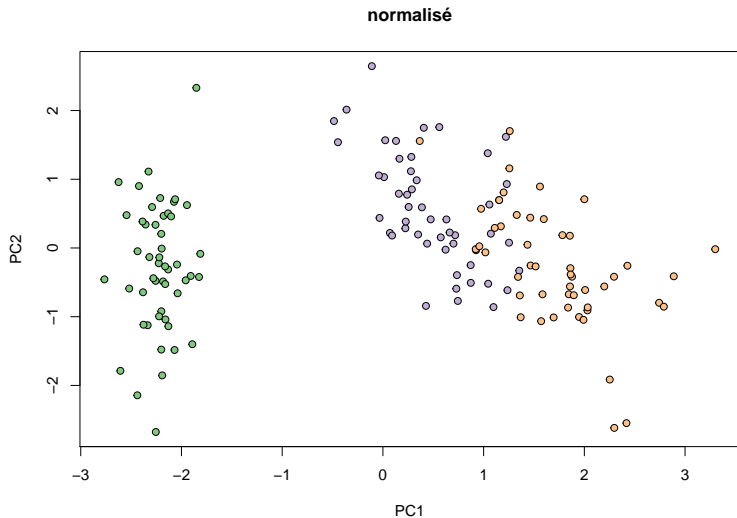
- aide à l'interprétation :
  - quel sens accorder aux axes principaux ?
  - lien entre les axes et les variables d'origine
- on considère la corrélation entre  $X_{.j}$  et  $Y_{.k}$  :
  - comme les variables sont centrées, la covariance de  $X_{.j}$  et  $Y_{.k}$  est  $\frac{1}{N}(X^T Y)_{jk}$
  - la corrélation entre  $X_{.j}$  et  $Y_{.k}$  est donc

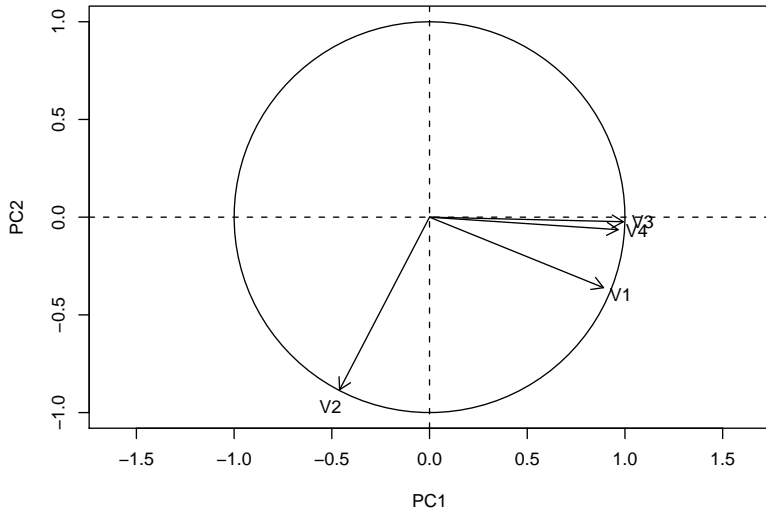
$$\frac{\sqrt{\lambda_k} \phi_{jk}}{\sigma_j},$$

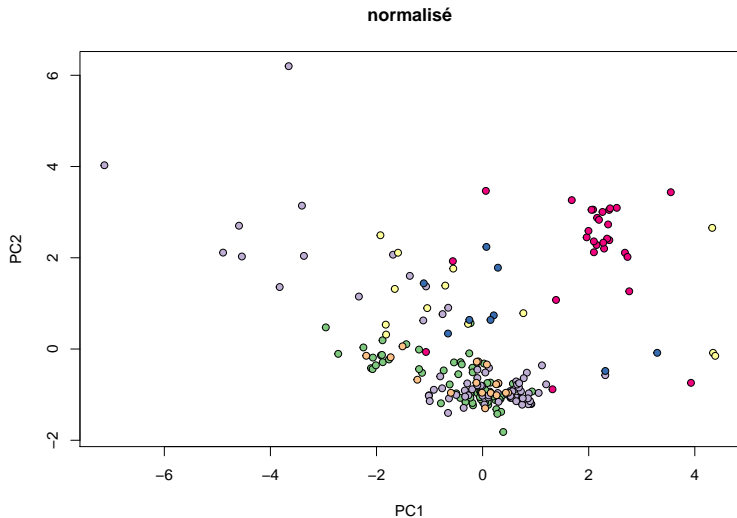
où  $\sigma_j$  est l'écart-type de  $X_{.j}$

- on peut dessiner les variables réduites : cercle des corrélations



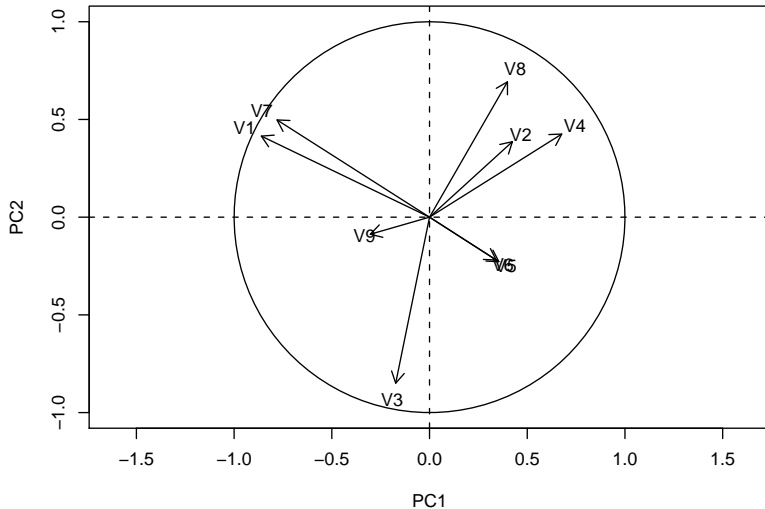








# Données glass



Introduction

Analyse en composantes principales

Modèle

Qualité et interprétation

Autres méthodes

- pas de prise en compte des dépendances non linéaires
- s'appuie sur les corrélations : pas de corrélation, pas de simplification !
- projection linéaire : écrase les données
- ne préserve pas les distances

- pas de prise en compte des dépendances non linéaires
- s'appuie sur les corrélations : pas de corrélation, pas de simplification !
- projection linéaire : écrase les données
- ne préserve pas les distances
- autres solutions :
  - méthodes de placement non linéaire : pas de lien linéaire entre  $X$  et  $Y$
  - optimisation d'une mesure de préservation des distances (dissimilarités)



# Multi Dimensional Scaling

- famille des algorithmes de *Multi Dimensional Scaling*
- méthode de Kruskal-Shepard : minimiser

$$\sum_{i \neq j} (d(x_i, x_j) - \|y_i - y_j\|)^2$$

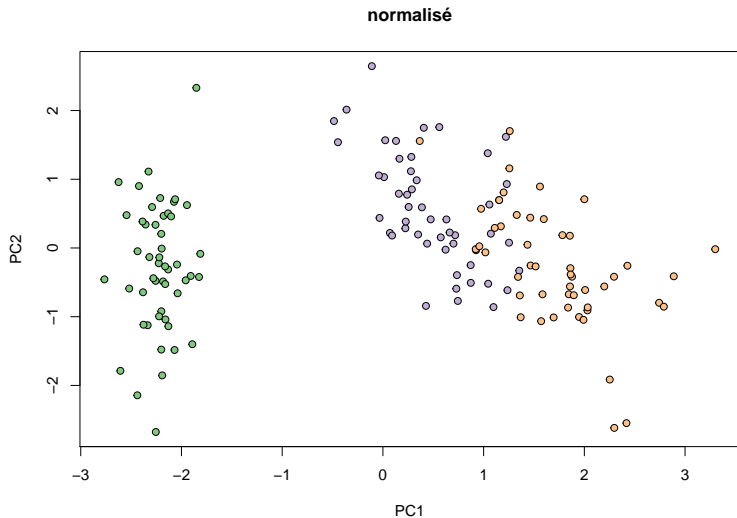
- méthode de Sammon : minimiser

$$\sum_{i \neq j} \frac{(d(x_i, x_j) - \|y_i - y_j\|)^2}{d(x_i, x_j)}$$

ce qui favorise les petites distances

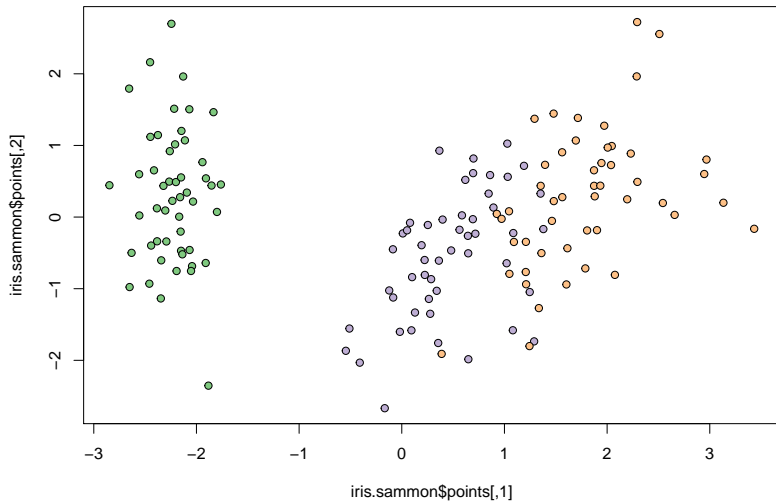
- nombreuses autres variantes

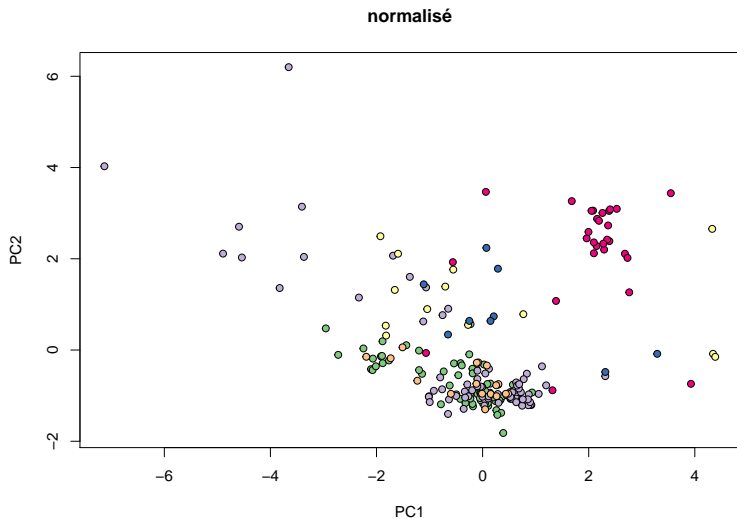






## Sammon







## Sammon

