Analyse de Données – Développeur
C5-160512-INFO
2. Tableaux et Espaces

Pr. Carl FRÉLICOT

FST/PAS221 – carl.frelicot@univ-lr.fr

Licence d'Informatique 3ème année

Automne 2021 - La Rochelle

1. Préambule

2. Tableaux et Espaces
   2.1 Types de Tableaux
   2.2 Espace des Individus – $\mathbb{R}^p$
   2.3 Espace des Variables – $\mathbb{R}^n$

3. Réduction de la Dimensionnalité

4. Classification Non Supervisée

5. Classification Supervisée

- individus *vs* variables quantitatives
  données dites *vectorielles*

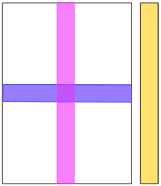
  $X = [X_{ki}]_{k=1,n;i=1,p}$ nuage de n points en dim. p

  Principal Component Analysis

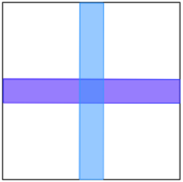
  $+$ éventuellement une variable catégorielle (supervisé)

  Linear Discriminant Analysis

$X \ (n \times p)$ $Y \ (n \times 1)$



$D \ (n \times n)$



- individus *vs* individus
  données non vectorielles

  ex. : tableau de distances, proximités, etc

  $D = [d_{kl}]_{k,l=1,n}$ où $d_{kl} = d(x_k, x_l)$

  K-Means, Hierarchical Clustering

- 1 variable qualitatives *vs* 1 variables qualitative

  ex : tableau de contingence

  Correspondance Analysis

  extensible à plus de deux variables qualitatives

modalités x modalités

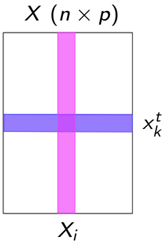
1. Préambule

2. Tableaux et Espaces
   2.1 Types de Tableaux
   2.2 Espace des Individus – $\mathbb{R}^p$
   2.3 Espace des Variables – $\mathbb{R}^n$

3. Réduction de la Dimensionnalité

4. Classification Non Supervisée

5. Classification Supervisée



- individus *vs* variables quantitatives
  données dites *vectorielles*
  $X = [X_{ki}]_{k=1,n;i=1,p}$ nuage de n points en dim. p
- $x_k = [x_{ki}]_{i=1,p}$ où $x_{ki} = < x_k, e_i > = {}^t x_k \, e_i$ e_i unitaires

Produit Scalaire

- $<x, y> = {}^t x \, y = \sum_{i=1}^{p} x_i \times y_i$ projection
- norme $||x|| = \sqrt{<x, x>}$ dist. à l'origine
- angle $cos(x, y) = \frac{<x,y>}{||x|| \, ||y||}$ proximité
- distance $d(x, y) = ||x - y||$ euclidienne usuelle

M-Produit Scalaire

matrice de produit scalaire

- $<x, y>_M = {}^t x \, M \, y$
  en général M définit une métrique *si M symétrique, définie positive*
- cas particuliers usuels : M diagonale, $M = I_p$
- M-norme (carré) $||x||_M^2 = {}^t x \, M \, x$, ... pondération des variables

Comment est définie une variable X_i ? $\boxed{X_i = X \, e_i}$

Une (nouvelle) variable C_j de vecteur directeur u_j unitaire ? $\boxed{C_j = X \, u_j}$

Projeter un nuage X dans espace défini par $U = [u_1 \, u_2 \ldots u_q]$? $\boxed{C = X \, U}$

1. Préambule

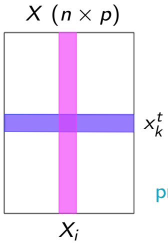
2. Tableaux et Espaces
   - 2.1 Types de Tableaux
   - 2.2 Espace des Individus – $\mathbb{R}^p$
   - 2.3 Espace des Variables – $\mathbb{R}^n$

3. Réduction de la Dimensionnalité

4. Classification Non Supervisée

5. Classification Supervisée

- individus *vs* variables quantitatives
  données dites *vectorielles*
  $X = [X_{ki}]_{k=1,n;i=1,p}$ nuage de n points en dim. p
- $x_k = [x_{ki}]_{i=1,p}$ où $x_{ki} = < x_k, e_i >= {}^t x_k \, e_i$ e_i unitaires



Produit Scalaire (variables centrées) X'

- $<X_i', X_j'>_{D_n} = {}^t X_i' \, D_n \, X_j' = \boxed{Cov(X_i, X_j)}$ projection
  avec la métrique $D_n = \frac{1}{n} I_n$ pondération des individus
- norme (carré) $||X_i'||^2_{D_n} = {}^t X_i' \, D_n \, X_i' = \boxed{Var(X_i)}$ longueur
- angle $cos(X_i', X_j') = \frac{<X_i', X_j'>_{D_n}}{||X_i'||_D \, ||X_j'||_{D_n}} = < X_i'', X_j'' >_{D_n} = \boxed{Corr(X_i, X_j)}$ proximité

M-Produit Scalaire

matrice de produit scalaire

- $<x, y>_M = {}^t x \, M \, y$
  en général M définit une métrique *si M symétrique, définie positive*
- cas particuliers usuels : $M = I$, M diagonale
- M-norme (carré) $||x||^2_M = {}^t x \, M \, x$, ... pondération des variables

Comment est définie une variable X_i ?

$\boxed{X_i = X \, e_i}$

Une (nouvelle) variable C_j de vecteur directeur u_j unitaire ?

$\boxed{C_j = X \, u_j}$

Projeter un nuage X dans espace défini par $U = [u_1 \, u_2 \ldots u_q]$?

$\boxed{C = X \, U}$

1. Préambule

2. Tableaux et Espaces
   - 2.1 Types de Tableaux
   - 2.2 Espace des Individus – $\mathbb{R}^p$
   - 2.3 Espace des Variables – $\mathbb{R}^n$

3. Réduction de la Dimensionnalité

4. Classification Non Supervisée

5. Classification Supervisée