Université Paris Dauphine PSL Research University

Département MIDO
Module Analyse de Données
2016 / 2017

Travaux Dirigés n° 6 : ACP

Objectifs : comprendre l'ACP.

1 Exercice 1

Démontrer l'égalité suivante (avec g le centre de gravité et l'opérateur $\tilde{x}$ correspondant à la projection orthogonale de x sur un espace vectoriel) :

$$\begin{aligned} D_V &= \sum_{i \neq j} p_i p_j (d^2(x_i, x_j) - d^2(\tilde{x}_i, \tilde{x}_j)) \\ &= \sum_i p_i (d^2(x_i, g) - d^2(\tilde{x}_i, \tilde{g})) \end{aligned}$$

2 Exercice 2

Considérons le tableau de données suivant :

	Math	Info	Gestion
Noam	0	4	0
Jean	1	3	1
Li	2	2	3
Lisa	3	1	1
Mina	4	0	0

où les lignes représentent les individus (noms de quelques étudiants de L3 informatique) et les colonnes les variables (notes en mathématiques, informatique et gestion). Ce tableau de données peut être représenté par la matrice X de données brutes :

$$X = \begin{pmatrix} 0 & 4 & 0 \\ 1 & 3 & 1 \\ 2 & 2 & 3 \\ 3 & 1 & 1 \\ 4 & 0 & 0 \end{pmatrix}$$

1. Calculer la matrice Y des données centrées et la matrice Z de données centrées et réduites.
2. Calculer la matrice R_X des corrélations de X et la matrice V_Z des variances et covariances de Z. Commentez.

3. Vérifier que les vecteurs unitaires

$$\begin{aligned}\mathbf{u}_1^t &= \frac{1}{\sqrt{2}}[-1, 1, 0] \\ \mathbf{u}_2^t &= [0, 0, 1]\end{aligned}$$

sont les vecteurs propres de R_X (ou, équiv., de R_Z) associés à des valeurs propres non nulles. Trouvez-les.

4. Peut-on représenter parfaitement le nuage des individus à 2 dimensions ? Si oui, représentez les individus de Z sur le plan défini par la base $S = \{\mathbf{u}_1, \mathbf{u}_2\}$ (Indication : calculer la projection de chacun des points - par rapport au produit scalaire usuel - dans le plan engendré par les deux vecteurs propres).

3 Exercice 3

On considère la matrice de données suivante :

$$A = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 2 \\ 2 & 2 & 1 \\ 1 & 0 & 0 \\ 2 & 3 & 2 \end{pmatrix}$$

1. Calculer g (le centre de gravité), Y (la matrice des données centrées) et V (la matrice de covariances).
2. Vérifier que $\begin{pmatrix} 1 \\ -1 \\ 1 \end{pmatrix}$ et $\begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}$ sont vecteurs propres de V, en indiquant les valeurs propres associées. En déduire le troisième vecteur propre de V et la valeur propre correspondante.
3. À partir des résultats précédents, déterminer les deux axes factoriels non triviaux de l'ACP du nuage $N(I)$ des individus associé au tableau X. Pour chacun de ces axes, préciser l'inertie du nuage projeté sur l'axe considéré, et la part d'inertie qu'il explique.
4. Calculer les composantes principales pour chaque individu.
5. Représenter graphiquement le nuage $N(I)$ sur le plan factoriel défini par les deux premiers axes factoriels. Que peut-on dire de cette représentation graphique ?
6. Quel est l'individu qui contribue le plus à l'inertie du premier axe factoriel ? Calculer sa qualité de représentation sur chacun des deux axes factoriels non triviaux.
7. Représenter graphiquement le nuage des variables $N(V)$ sur le plan factoriel défini par les deux premiers axes factoriels. Que peut-on dire de cette représentation graphique ?
8. Quelle est la variable qui contribue le plus à l'inertie du premier axe factoriel ? Calculer sa qualité de représentation sur chacun des deux axes factoriels non triviaux.