ahmedrahmoune.umbb@gmail.com

Année universitaire 2021-2022

Umbb/FS/Dépt Maths/Stat.MSS1/Analyse des Données

Dec2021

Notion d'Inertie et Théorème de Huygens.

Décompositon en valeurs singuliéres (SVD)

Analyse en composantes principales (ACP)

Feuilles de Travaux Dirigés n° 2

Exercice 1

Appliquer le principe de reconstitution des données pour les tableaux suivants avec la métrique usuelle (distance euclédienne) et les points non pondérés(points matériels) méthode SVD

$$X=\begin{bmatrix} 1 & 1 & 1 \\ 1 & -1 & 0 \\ 0 & 1 & -1 \\ -1 & 0 & 1 \\ -1 & -1 & -1 \end{bmatrix} \qquad Y=\begin{bmatrix} 2 & -3 & -5 \\ -2 & 2 & 0 \\ 0 & 4 & 2 \\ 4 & 0 & 2 \\ 2 & -2 & 0 \\ -6 & -1 & 1 \end{bmatrix} \qquad Z=\begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & -1 \\ -1 & 1 & 0 \\ 0 & -1 & 1 \\ -1 & 0 & 1 \\ 1 & -1 & 0 \end{bmatrix}$$

Exercice 2

Questions de cours

Donner la définition des notions suivantes

(i) Le nuage des points lignes ou nuage des individus $\mathcal{N}$(I). Centre de gravité du nuage des individus N(I) noté g

(ii) L'inertie du nuage par rapport à un point x$\in$ $\boldsymbol{R}^p$ noté $I_x$et l'inertie de $\mathcal{N}$(I) contenant n points.

(iii) Enoncer le Théorème de Huygens.

(iv) Indiquer le point où I_x atteint son minimum

(v) Que signifie I_G

(vi) Quelle relation existe entre I_G et V (Matrice variance covariance)

(vii) L'axe principal, Plan principal

(viii) Taux d'inertie.

Exercice 3

Soit le tableau des données $X_{n,p} = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \\ 2 & 0 \\ 2 & 1 \\ 2 & 2 \\ 3 & 0 \\ 3 & 1 \\ 3 & 2 \end{bmatrix}$ croisant 9 individus de **même** poids et 2 variables X^1, X^2

(i) Représenter géométriquement le nuage N($X_i, i = 1, ..., 9$) des individus dans l'espaces $\mathbb{R}^2$ muni de la base canonique usuelle.

(ii) Calculer le centre de gravité g de N(X_i), représenter le graphiquement.

(iii) Calculer I_0l'inertie du nuage N(X_i) par rapport à l'origine de $\mathbb{R}^2$, muni de la métrique usuelle en appliquant la définition (donner la définition théorique et appliquer)

(iv)Déterminer XC la matrice des donnés centrées, représenter graphiquement le nuage des individus centrés N(X_i^c) Quel est son centre de gravité?

(v)Déterminer V la matrice variance covariance de $X_{n,p}$ en déduire tr(V).

(vi) Déterminer I_g en appliquant le théorème de Huyghens.Comparer I_g et tr(V).Généraliser le résultat lorsque $\mathbb{R}^2$ est muni d'une métrique euclédienne M quelconque.

(vii) Donner le spectre de V (valeurs propres et vecteurs propres)

(viii) Donner Les facteurs principaux F_1, F_2

Exercice n°4

Soit le corpus statistique suivant: Deux variables X^1, X^2 décrivant 10 individus de même poids.

Les résultats sont résumés ci-dessous:

$$X_{10,2} = \begin{array}{c} \begin{array}{cc} X^1 & X^2 \end{array} \\ \begin{bmatrix} 0.5 & 0 \\ -0.1 & 1.2 \\ -0.5 & 0.5 \\ -0.3 & 0.1 \\ 0 & 2.5 \\ 1.6 & -0.7 \\ 2.0 & 2.0 \\ 2.4 & 1.2 \\ 0.5 & 3.5 \\ 2.7 & -0.9 \end{bmatrix} \end{array}$$

1: Représenter graphiquement le nuage des individus dans le plan (Chaque individu est plongé dans l'espace $\mathbb{R}^2$)

2. Calculer le centre de gravité g^t=($\overline{X_1}$,$\overline{X_2}$) et figurer le dans le même graphe.

3. Centrer les données (Trouver Y=X-1g ^{t} où 1est un vecteur de 10 lignes avec la valeur 1

sur chaque composante)

4. Déterminer V la matrice variance-covariance de X en fonction de Y. et déterminer son spectre (valeurs propres et vecteurs propres associées)

5. Soit une nouvelle variable F (combinaison linéaire de X^1, X^2)

$$F = \alpha_1 X^1 + \alpha_2 X^2$$

Trouver les coefficients α_1, α_2 telle que var (F) soit maximale avec la contrainte $\alpha_1^2 + \alpha_2^2 = 1 (var(f)$ désigne la variance de la variable F)

Indication: Utiliser la question 4 et un résultat concernant la -Maximisation d'une forme quadratique sous contrainte qudratique (On ne demande pas la démonstration du théorème)

6. Comment peut-on représenter graphiquement F à l'aide d'une droite passant par g?

7. Dire en quoi le nuage des points -des individus- projeté sur F est la meilleure représentation linéaire à une dimension du nuage initial?

Exercice n°5

Reprendre les mêms questions de l'exercice précédent 4 avec 3 variables X^1, X^2, X^3

ainsi F= $\alpha_1 X^1 + \alpha_2 X^2 + \alpha_3 X^3$

$$X_{5,3} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & -1 & 0 \\ 0 & 1 & -1 \\ -1 & 0 & 1 \\ -1 & -1 & -1 \end{bmatrix}$$

Exercice 6

Soit X la matrice des rangs (classements) de 9 étudiants dans 3 matières différentes.

$$X=\begin{bmatrix} 1 & 2 & 9 \\ 2 & 3 & 7 \\ 3 & 5 & 8 \\ 4 & 9 & 6 \\ 5 & 6 & 5 \\ 6 & 7 & 4 \\ 7 & 4 & 3 \\ 8 & 1 & 2 \\ 9 & 8 & 1 \end{bmatrix}$$

(i)Donner g le centre de gravité de $\mathcal{N}(\mathcal{I})$. Donner g en générale pour le classement de n étudiants dans p matiéres.

(ii) Donner Var(X^j) pour j=1,2,3. généraliser pour n étudiants avec p matiéres.

(iii) Donner XC la matrice centrée.

(iv) Donner V la matrice variance covariance.

Exercice 7

Soit $X_{n,p}$ le tableau des données croisant n=8 individus et p=4 variables.

$$X_{n,p}=\begin{bmatrix} 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 \end{bmatrix}$$

(i) En considérant les individus de même poids, calculer l'individu moyen G.

(ii)Déduire la matrice X^* des données centrées.

(iii) $\mathbb{R}^4$ étant munis de la métrique euclédienne usuelle , calculer V (Matrice variance co-variance)

(iv) Montrer que le polynôme caractéristique de V est : $p(\lambda) = \lambda^2(\frac{1}{2} - \lambda)^2$

(v)En déduire les valeurs propres λ de V avec leurs multiplicités.et calculer Trace(V)

(vi) Montrer que les vecteurs $u_1 = \begin{bmatrix} 1 \\ -1 \\ 0 \\ 0 \end{bmatrix}$ et $u_2 = \begin{bmatrix} 0 \\ 0 \\ 1 \\ -1 \end{bmatrix}$ sont des vecteurs propres de V associées à la même valeur propre à déterminer..

(vii) Déterminer $u'_1 = \dfrac{u_1}{\|u_1\|}$ et $u'_2 = \dfrac{u_2}{\|u_2\|}$

Calculer $c_{n,1} = X^*\ u'_1$ et $c_{n,2} = X^*\ u'_2$

(viii) Faite la représentation sur les axes et sur le plan principal (relatif à l'origine, X^*a pris la place de X) en mentionant les taux d'inerties.

<u>Exercice n°8</u>

Le but de cet exercice: Analyse d'opinions.

Méthode utilisée: ACP

Soit le tableau suivant:$R_{n,p}$: $\begin{bmatrix} 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & -1 & 0 \\ 0 & -1 & 0 \\ 0 & -1 & 0 \\ 1 & 0 & -1 \\ -1 & 0 & 1 \end{bmatrix}$

résumant les réponses codées comme suit: $\begin{cases} 1 : \text{ Pour} \\ 0 : \text{Pas d'opinion} \\ -1 : \text{Contre} \end{cases}$

de n=10 individus concernant p=3 avis X^1, X^2, X^3.

(i) En considérant les individus de même poids, calculer g l'individu moyen ainsi que la matrice R^*matrice des données centrées.

(ii) $\mathbb{R}^3$ étant muni de la métrique euclédienne usuelle, calculer V matrice variance covriance.

(iii) a)Vérifier que $\lambda_1 = \frac{19}{25}$ est une valeur propre de V de vecteur propre $v_1 \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$

b)Quelle est la valeur propre $\lambda_2$correspondant au vecteur propre $v_2 \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}$

c)Quel est le vecteur propre v_3 correspondant à la valeur propre $\lambda_3 = 0$

(iv)Calculer l'inertie totale du nuage des individus, normaliser les vecteurs propres v_1 et v_2.

(v) En déduire les composantes principales des deux premiers axes (Axes d'inerties maximales)

(vi) Faite la représentation dans cet espace en mentionnant le taux d'inertie de chaque axe. Interpréter le résultat.