Résumé : Analyse en Composantes Principales (ACP) : réduction des données, inertie, orthogonalisation, axe factoriel, composante principale, indicateurs QLT et CTR, simulation

L' ´enoncé de ce TP se trouve sur l'intranet : \ISTIA\Public\Depots_Ens\Jean-Claude_Jolly

Exercice 1

Le but est de retrouver et programmer avec Matlab les résultats du polycopié *Introduction à l'ACP par la notion d'inertie*. Le fil conducteur est celui de l'exemple des bénéfices hebdomadaires (en centaines d'euros) sur trois produits pour six magasins ($m_i, i = 1, \ldots, 6$) d'une même enseigne : lait frais (x), eau (y) et huile (z). Ils sont donnés par le tableau X_0 suivant :

X_0	x	y	z
m_1	15	7	5
m_2	8	12	1
m_3	11	7	7
m_4	10	8	13
m_5	6	14	10
m_6	10	6	12

Les poids des magasins sont uniformes (1/6). Des indications (Indic. :) sur les fonctions utiles de Matlab sont données au fil des questions. Il pourra être utile de consulter l'aide de Matlab (helpdesk ou taper *help mot* dans la fenêtre de commande). Pour les questions théoriques, les réponses sont dans le polycopié !

1. Calculez une matrice diagonale D dont les éléments diagonaux sont les poids (1/6) des magasins. *Indic.* : eye
2. Calculez l'individu moyen G sous forme d'un vecteur en ligne. *Indic.* : mean
3. Calculez une matrice S dont les éléments diagonaux sont les écarts types des variables x, y, z. *Indic.* : std (option 1), diag
4. Calculez le tableau X des données réduites.
5. Calculez la matrice $\Phi_{X,X} = X^T \cdot D \cdot X$. Pourquoi ses éléments diagonaux sont-ils égaux à 1 ? Quelle est l'interprétation statistique d'un élément de $\Phi_{X,X}$ en général ?
6. Diagonalisez $\Phi_{X,X}$ pour obtenir une matrice ligne L constituée des valeurs propres λ_j de $\Phi_{X,X}$ en ordre décroissant ainsi qu'une matrice carrée V dont les colonnes sont les vecteurs propres associés. Vérifiez et justifiez l'identite $\lambda_1 + \lambda_2 + \lambda_3 = 3$. *Indic.* : eigs
7. Pourquoi était-il prévisible que les valeurs propres de Φ_{XX} soient positives ?
8. Que représente la diagonale de la matrice $V^T \cdot V$? En déduire une matrice carrée U d'ordre trois dont les vecteurs colonnes U_j sont les vecteurs colonnes de V mais normés (i.e. unitaires). On pose $u_j = (U_j)_\varepsilon$ le vecteur correspondant de l'espace $\mathbb{R}^3$ des magasins muni de sa base canonique ε. On note $\mu = (u_1, u_2, u_3)$ la nouvelle base orthonormée ainsi construite.
9. Calulez une matrice $Fu \in \mathcal{M}_{6,3}(\mathbb{R})$ dont les vecteurs colonnes Fu_j sont les composantes principales (dites aussi variables synthétiques) associées à X.
10. Calculez les moyennes des composantes principales Fu_j. Pourquoi le résultat était-il prévisible ?
11. Calculez la matrice $Fu^T \cdot D \cdot Fu$. Quelle est l'interprétation statistique d'un élément de cette matrice ?

12. Soit In_O l'inertie par rapport à $O = G$ du nuage N_6 associé à X et soit $In_O(u_j)$ son inertie projetée sur l'axe $\Delta_{u_j} = (O, u_j)$. Démontrez la relation $In_O = \sum_{j=1}^{3} In_O(u_j) = 3$.
13. Retrouvez dans le polycopié la démonstration de $In_O(u_j) = \lambda_j$. En revenant à la définition de $In_O(u_j)$, en déduire que $Var(Fu_j) = \lambda_j$.
14. Calculez les approximations F_1 (avec un degré de liberté associé au 1er axe factoriel) et $F_{1,2}$ (avec deux degrés de liberté associé au 1er plan factoriel) de X comme indiqué dans l'exemple 4 p.7 du poly.
15. L'enseigne envisage l'ouverture d'un septième magasin m_7. Le gérant de celui-ci mise sur un stock de lait frais correspondant à un bénéfice de $x_{7,1} = 11$ centaines d'euros la première semaine. Quels bénéfices $x_{7,2}$ et $x_{7,3}$ sur l'eau et l'huile peut-il espérer selon l'approximation F_1 ?
16. Calculez une matrice $QLT \in \mathcal{M}_{6,3}(\mathbb{R})$ dont les colonnes sont les indicateurs QLT_{u_1}, QLT_{u_2} et $QLT_{u_1+u_2}$.
17. Calculez une matrice $CTR \in \mathcal{M}_{6,3}(\mathbb{R})$ dont les colonnes sont les indicateurs CTR_{u_1}, CTR_{u_2} et $CTR_{u_1+u_2}$.
18. Donnez une figure avec le nuage X_0, le magasin moyen G, l'approximation plane $F_{1,2}$ (1er plan factoriel) et l'approximation affine F_1 (1er axe factoriel). *Indic. :* voici un programme qui trace la surface $z = -1 + 2 * x^2 + y^2$ dans le rectangle $R = [-2, 2] \times [-3, 3]$, le segment $[-1, 20]$ surl'axe (Oz) et deux points $G_1 = (1, 2, 3)$ et $G_2 = (-1, 2, 3)$.

```
xi = linspace(-2,2,11) ; % 11 points répartis entre -2 et 2
yj = linspace(-3,3,15) ; % 15 points répartis entre -3 et 3
[Xi,Yj] = meshgrid(xi,yj) ; % lignes (resp. colonnes) de Xi (resp Yj) = copies de xi (resp
yj)
G = [1,2,3 ; -1,2,3] ; % points
Zk = -1+2*Xi.^2+Yj.^2 ; % équation de la surface
t = linspace(-1,20,100) ; % paramètre
xt = 0*t ; yt = 0*t ; zt = t ; % droite paramétrée
figure(1)
clf % efface ce qui précède
hold on % superpose les tracés
grid on % grilles
surf(Xi,Yj,Zk) % surface
alpha(0.5) % transparence
plot3(G( :,1),G( :,2),G( :,3),'o','MarkerEdgeColor','k','MarkerFaceColor','r','MarkerSize',8)
% points
plot3(xt,yt,zt,'b-','LineWidth',2) % segment
title ('Un paraboloïde, un segment et deux points')
xlabel('x'), ylabel('y'), zlabel('z')
```