Feuille d'exercices ACP et k-moyennes

Chloé Clavel

Exercice 1 : les k- moyennes

Soient 8 observations : $A1 = (2, 10)$, $A2 = (2, 5)$, $A3 = (8, 4)$, $A4 = (5, 8)$, $A5 = (7, 5)$, $A6 = (6, 4)$, $A7 = (1, 2)$, $A8 = (4, 9)$.

Q. 1 Utiliser l'algorithme des k-moyennes avec la distance Euclidienne pour regrouper les 8 exemples précédents en trois groupes (ou *clusters*). Vous prendrez comme centres pour l'initialisation A1, A4, et A7. Faire tourner l'algorithme sur une époque, dans un premier temps. A la fin de cette époque, vous préciserez :
- les exemples appartenant à chacun des trois groupes,
- les nouveaux centres des groupes
- la représentation en deux dimension des exemples et vous entourerez les exemples associés à chaque groupe en fonction du regroupement obtenu à la fin de cette époque

Q. 2 Continuer de dérouler l'algorithme des k-moyennes jusqu'à ce que l'algorithme converge. Comme dans la question précédente, représenter les regroupements obtenus et les nouveaux centres obtenus à la fin de chaque époque.

Exercice 2 : Analyse en composantes principales

Un expert a noté quatre restaurants R1, R2, R3 et R4 en fonction des trois critères suivants : service, qualité et prix. La note donnée est comprise entre -3 et 3.

Les notes des 4 restaurants selon ces trois critères sont stockées dans la matrice X suivante :

Restaurant	Service	Qualité	Prix
R1	-2	3	-1
R2	-1	1	0
R3	2	-1	-1
R4	1	-3	2

Le calcul de la matrice de covariance V de X donne :

$$V = \begin{pmatrix} \frac{5}{2} & -3 & \frac{1}{2} \\ -3 & 5 & -2 \\ \frac{1}{2} & -2 & \frac{3}{2} \end{pmatrix}$$

Q. 1 Montrer que les variables de la matrice X sont centrées.

Q. 2 Etude des valeurs propres de V
- Vérifier que V admet une valeur propre $\lambda_3 = 0$
- On donne $\lambda_1 = \frac{30.5}{4}$, en déduire λ_2
- Calculer les pourcentages d'inertie associés à chacune des valeurs propres. Quel est le sous-espace optimal de dimension 1 ?

Q. 3 On donne les vecteurs propres (aux erreurs d'arrondi près)

$$v_1 = \begin{pmatrix} 0.5 \\ -0.8 \\ 0.3 \end{pmatrix} \text{ et } v_2 = \begin{pmatrix} 0.65 \\ 0.11 \\ -0.75 \end{pmatrix}.$$

Calculer les composantes principales et représenter les restaurants dans l'espace de projection optimal de dimension 2 et interpréter les résultats.