

Classification et Analyse des données (examen de rattrapage)

Le : 01/07/2013

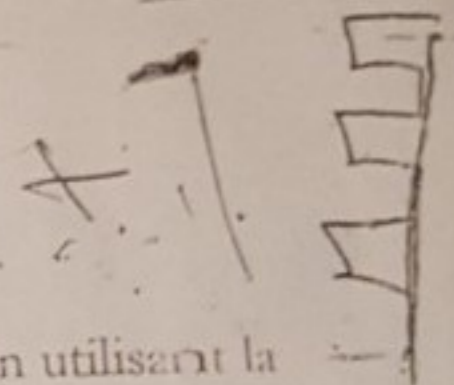
durée : 1 H : 30

Questions de cours (6 points) :

1. Quels sont les inconvénients majeurs de l'algorithme K-means ? *↳ - choix aléatoire n° de classe "la solution"*
2. Quelles sont les étapes de l'ACP ?
3. Dans les tests de classification, expliquez est-ce que on cherche à minimiser ou à maximiser la mesure FT.
4. Soit M la mesure suivante entre deux points a et b : $M(a, b) = \max |a_i - b_i|$. M Est une mesure de similarité ou de dissimilarité ? pourquoi ?

Exercice 1 (7 points) : ✓

A(0, 0, 1, 0) B(1, 1, 0, 1) C(1, 2, 1, 0) D(2, 0, 2, 1) E(3, 2, 4, 1)



Effectuer une classification hiérarchique ascendante de cet ensemble de données en utilisant la distance euclidienne. Il faut spécifier :

- Les tableaux de distances à chaque regroupement de points ou d'ensembles.
- Les données regroupées à chaque étape.
- Le dendrogramme obtenu.

Exercice 2 (7 points) :

Le tableau de contingence suivant indique la répartition en fonction des régions les résultats bac 2008 selon les filières :

	NORD	SUD	EST	OUEST
Scientifique	104 127	14 071	89 131	88 164
Technologique	91 567	9 651	75 042	73 417
Littéraire	19 237	6 483	9 247	8 632

Quels sont les régions les plus proches et les plus éloignées ?

max = pour éviter en trouver le ~~meilleur~~ élément dans une classe

min = pour regrouper les données qui a un ~~meilleur~~ résultat

Remarque : Présentez vos résultats en millièmes.
 La clarté des réponses et de la présentation sera prise en considération.