

Classification

Examen partiel du 15/11/2016

Préalables : Ouvrir Rstudio, puis entrer les commandes suivantes :

```
install.packages("ade4")  
library(ade4)
```

On écrira toutes les commandes demandées dans l'éditeur en prenant soin d'enregistrer le fichier à la fin de la séance (peu importe le nom et l'emplacement du fichier).

Exercice 1

Soit le tableau de données suivant (1 variable, 5 individus) :

	e_1	e_2	e_3	e_4	e_5
X	-1	0	2	3	5

1. Appliquer à la main l'algorithme des k -moyennes sur les données, avec $k = 3$ et en choisissant les individus e_3 , e_4 et e_5 comme centres initiaux. On décrira précisément chacune des étapes et les groupes obtenus à la fin.
2. Entrer les données dans R, puis effectuer plusieurs fois avec R l'algorithme des k -moyennes sur les données, avec $k = 3$ et une initialisation aléatoire. Obtient-on toujours le même partitionnement à la fin, et ce partitionnement correspond-il à celui obtenu à la question 1 ? Expliquer pourquoi.

Exercice 2

Le tableau de données `escopage` contient 27 caractéristiques quantitatives observées sur 21 vins différents. Les caractéristiques sont regroupées suivant leurs types : "repos" (variables dont le nom commence par la lettre R), "visuel" (variables commençant par V), "olfactif" (variables commençant par A), "général" (variables commençant par G).

1. Charger les données avec la commande `data(escopage)`, puis les observer pour comprendre comment elles s'organisent.
2. Effectuer, sur le tableau complet des données, plusieurs classifications ascendantes hiérarchiques, en choisissant différents critères de dissimilarités et différentes stratégies d'agrégation. Afficher à chaque fois les dendrogrammes obtenus.
3. En observant les résultats, choisir un critère de dissimilarité et une stratégie d'agrégation qui semblent convenir, en justifiant votre choix.
4. Choisir un nombre de classes optimal (justifier également ce choix), puis donner la classification des différents vins obtenue pour ce choix.
5. En gardant la même dissimilarité et le même critère d'agrégation choisis à la question 3, effectuer des classifications ascendantes hiérarchiques en ne gardant que les variables de type "repos", puis seulement celles de type "visuel", puis celles de type "olfactif", puis "général". Calculer les classifications obtenues à chaque fois.
6. Comparer, à l'aide de tables de contingence, les différents classements obtenus à la question précédente avec le classement global (celui effectué sur toutes les variables). Pour quel type de variables obtient-on la meilleure corrélation avec le classement global ?