

Le : 24/05/2014

(Contrôle finale) Classification et Analyse des données

durée : 1H : 30

Question de cours : (3 points)

1. Quelle est la différence entre l'objectif des deux algorithmes de classification BIRCH et K-PP ?
2. Quelle est la différence fondamentale entre la classification et l'analyse discriminante linéaire (LDA) ?
3. Quelle est l'intérêt du centrage de la matrice des données dans l'analyse ACP ?

Exercice 1 : (6 points)

X	33	60	43	71	35	86	95
Y	5.5	21.3	15.1	39	6.7	43.5	52.4

1. Calculez la variance et l'écart-type de la variable X et de la variable Y.
2. Calculez la covariance entre X et Y.
3. Déterminez, par la méthode des moindres carrés, l'équation de la droite de régression de Y.
4. Calculez le coefficient de corrélation linéaire. Commenter.

Exercice 2 : (5 points)

Considérons un ensemble d'individus divisé en deux classes, A et B, chaque individu est décrit par deux variables numériques. La classe A contient 10 individus et 20 pour la classe B. Les centres de gravité des classes sont  $g_A = (0, -1)$  et  $g_B = (0, 2)$  et les matrices de covariance intra-classe sont :

$$S_A = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}, \quad S_B = \begin{pmatrix} 3 & 0 \\ 0 & 3 \end{pmatrix}$$

Déterminez la matrice de covariance totale S.

Exercice 3 : (6 points)

Nous avons utilisé l'algorithme k-means ( $k=3$ ) pour classer un ensemble de données A.

A =

Vect 0	0.7	0.2	0.5
Vect 1	0.6	0.3	0.4
Vect 2	0.7	0.4	0.6
Vect 3	0.8	0.2	0.6
Vect 4	0.4	0.8	0.5
Vect 5	0.7	0.5	0.6
Vect 6	0.2	0.3	0.5

Le tableau suivant représente les résultats obtenus de deux exécutions de l'algorithme :

1<sup>ère</sup> classification

C1	0	2
C2	1	4 5
C3	3	6

2<sup>ème</sup> classification

C1	1	4
C2	0	2 5
C3	3	6

1. Pourquoi on n'a pas obtenu le même résultat ?
2. Quelle est la meilleure classification ?

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