

Classification et Analyse des données

durée : 1H : 30

Le : 08/06/2013

Question de cours (8 points) :

1. Quelle est la différence fondamentale entre l'Analyse en Composantes Principales (ACP) et l'Analyse Discriminante Linéaire (ADL).
2. Démontrez que les valeurs propres  $\lambda_i$  (dans l'espace des variables) =  $\beta_i$  (valeurs propres dans l'espace des individus) d'une ACP.
3. Expliquez brièvement l'algorithme de classification des K-PP et comment l'algorithme prend sa décision pour classer un nouveau point  $x$ .
4. Quelle est la différence entre des techniques de classification et des techniques de réduction de dimensions ?

Exercice 1 (6 points) :

On cherche à établir une règle de discrimination à partir des données  $R^2$  de deux classes suivant :

$$x^+ = \begin{pmatrix} 1 & 1 \\ 1.5 & 0.5 \\ 0.8 & 1.2 \\ 1.5 & -0.2 \\ 0.5 & -0.3 \end{pmatrix} \quad x^- = \begin{pmatrix} -1 & -0.8 \\ -0.2 & 0.5 \\ 0.2 & -1 \\ -0.5 & -0.2 \\ 0.3 & -0.5 \end{pmatrix} \quad x^T = \begin{pmatrix} 0.5 & 0.4 \\ 0.2 & -0.3 \\ -0.3 & 0.1 \end{pmatrix}$$

*0,5 0,4  
0,2  
0,5 0,5  
0,65*

1. Classer les données  $x^T$  suivant l'algorithme K-PP pour ( $k=1$  et  $3$ ) en utilisant la distance euclidienne.

Exercice 2 (6 points) :

On observe les séries statistiques (accédants routière) suivantes :

Année	1985	1990	1995	2000	2005	2010
routes nationales	652	773	829	886	907	1045
autoroutes	450	676	816	969	1026	1324

1. Calculer les droites d'ajustement par la méthode des moindres carrés en fonction de l'année pour les deux séries considérées.
2. Calculer et commenter les coefficients  $R^2$  pour chacune de ces régressions linéaires.
3. Estimer le nombre d'accédants sur routes nationales en 2015. Même question pour les autoroutes.
4. En quelle année peut-on penser que les accédants sur les autoroutes vont dépasser ceux sur les routes nationales ?

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