

Traitement statistique des données

TP n^03

Discrimination

Les fichiers mis à votre disposition se trouvent à l'adresse :
http ://www.esiee.fr/~decambro

Le Tp porte sur l'utilisation de différentes méthodes de discrimination.

On travaillera sur les données de **Pollco2_disc.mat**. Ce sont les données sur la pollution par le CO2 étudiées dans les 2 TP précédents auxquelles on a ajouté des étiquettes de classes et d'autres pays à classer.

- A - Représentation des données

- **pays** Nom du pays
- **X** Matrice des données
- **L** Classes d'appartenance des pays (3 classes)
- **paysc** Nom des pays à classer
- **Xc** Coordonnées des nouveaux pays à classer

1. *Normaliser et visualiser les données de grâce à la fonction **classr** du TP2.*
2. *Normaliser les données à classer contenues dans la matrice **Xc** et déterminer leur projection sur le plan principal. Représenter les sur la même figure.*

Dans la suite on travaille avec les données normalisées.

- B - Discrimination par la règle des k-ppv

On utilise la fonction Matlab **knnclassify** qui permet de réaliser la discrimination par la méthode des plus proches voisins.

1. *Expliquer la différence entre les 3 options proposées : 'nearest', 'random', 'consensus'.*
2. *Appliquer cette fonction sur les données normalisées et comparer les résultats obtenus pour la classification avec cette fonction pour k allant de 1 à 4, et pour les 3 options proposées et commenter ces résultats.*

- C - Discrimination bayésienne

On utilise la fonction Matlab **classify** qui permet de réaliser la discrimination par mesure de voisinage de Mahalanobis et la discrimination bayésienne sous différents modèles. Cette fonction comporte différents types. On s'intéresse aux types : 'linear' , 'quadratic', 'mahalanobis'.

1. *Expliquer comment les fonctions discriminantes sont calculées dans les 3 types précédents.*
2. *Appliquer cette fonction sur les données normalisées et comparer les résultats obtenus par les 3 types. On donnera le classement, le taux d'erreur de chaque méthode et la probabilité a posteriori d'affectation quand elle existe.*
3. On travaille pour la suite avec les données normalisées projetées dans le plan principal.
*Appliquer la fonction **classify** sur les données normalisées projetées et comparer les résultats obtenus avec ceux obtenus en - 1 -.*
4. On souhaite déterminer les courbes séparatrices des classes dans les 3 types pour les données projetées. Dans l'argument de sortie **coeff** de **classify** on a accès à la fonction discriminante.
*Déterminer et tracer pour les 3 types les courbes séparatrices obtenues et comparer les résultats obtenus (utiliser la fonction **courbesep.m** fournie).*

Ce TP est à rendre au plus tard le Mercredi 2 Novembre 2011.