

## Analyse discriminante : les iris de Fisher

Didier MAQUIN

L'enjeu de cette séance de TD est de comprendre les principes de base de l'Analyse Discriminante Linéaire à travers le traitement, en utilisant Matlab®, du célèbre jeu de données des iris de Fisher.

Le fichier de données est disponible sous Matlab®. Il contient deux matrices : **meas** de dimension  $150 \times 4$  contenant 150 observations de 4 variables caractéristiques des iris (longueur sépale, largeur sépale, longueur pétale et largeur pétale) et un vecteur **species** de dimension 150 et contenant le descripteur d'espèce de chaque observation.

1. Charger le fichier en mémoire : **load fisheriris**;
2. Reproduire la figure 1 (page 7) du document de cours (instruction **gplotmatrix**);
3. Calculer la table 1 (page 7) d'indicateurs statistiques (instructions **min**, **mean**, **max**, **quantile**);
4. Reproduire la figure 2 (page 8) décrivant les boîtes à moustaches des différentes variables pour les différentes espèces (instruction **boxplot**);
5. Calculer la table 2 (page 8) (instruction **cov** ou **std**);
6. Calculer **T**, la matrice de covariance empirique totale des données (instruction **cov**);
7. Calculer **W** =  $\text{Esp}(\text{Var}(X | Y))$  : la matrice de covariance intra-classes (moyenne pondérée par leurs probabilités des matrices de covariance de chaque groupe);
8. En déduire la valeur de **B** la matrice de variance inter-classes;
9. Calculer les vecteurs propres et valeurs propres de la matrice **T**<sup>-1</sup>**B** (instruction **eig**);
10. Calculer les valeurs de la table 4 (page 13);
11. Reproduire la figure 3, projection des observations dans le plan factoriel des deux premières variables discriminantes;
12. Etendre l'analyse à l'analyse discriminante quadratique en augmentant artificiellement la dimension de la matrice de données (page 13) et obtenir les résultats de la figure 4.