

# Fiche 10 - TIAD - M1 IM

## Analyse discriminante

Tran Viet Chi, [chi.tran@math.univ-lille1.fr](mailto:chi.tran@math.univ-lille1.fr), bureau 316 (bâtiment M3).

### 1 Analyse discriminante sous R

#### Exercice 1 (Iris de Fisher)

On reprend les données sur les iris de Fisher. Sous R, on peut charger les données avec l'instruction `data(iris)`.

1. Une analyse de la variance permet de tester l'effet de la variable discrète `Species` sur les variables continues. Quelles sont les variables les plus discriminantes ?
2. Donner, pour chacune des espèces d'iris l'histogramme et la densité de la longueur des pétales. Superposer les densités gaussiennes qui approchent les lois correspondantes.

Dans la suite de cet exercice, on s'intéresse plus spécifiquement à discriminer les espèces `Versicolor` et `Virginica`.

3. Si on considère les variables relatives aux pétales, peut-on supposer que les matrices de variances-covariances pour les espèces `Versicolor` et `Virginica` sont égales ?
4. Visualiser la loi jointe en utilisant la fonction `bkde2D` du package `KernSmooth`.
5. Réaliser une analyse discriminante pour séparer `Versicolor` et `Virginica`, en utilisant les variables relatives aux pétales, qu'on supposera gaussiennes avec mêmes matrices de variances-covariances par espèce. Représenter sur un graphique les iris, en différenciant les espèces et en indiquant comment elles sont classées.
6. Comment peut-on s'affranchir de l'hypothèse gaussienne et du choix particulier de 2 variables parmi 4, tout en conservant la possibilité de tout représenter sur un plan ?
7. En reprenant les programmes d'estimation non-paramétrique de la densité de la fiche 8, faire une analyse discriminante fondée sur un échantillon d'apprentissage de 70 individus, qu'on isolera.

#### Exercice 2 (Iris, sous SAS)

On peut télécharger sur ma page web la base `iris.sas7bdat`.

1. Faire une analyse discriminante dans le cas homoscédastique gaussien.
2. Faire l'analyse discriminante en utilisant des estimateurs à noyaux.
3. Faire l'analyse discriminante en utilisant la méthode des plus proches voisins.