



# Licence d'Informatique 2

## Analyse de Données Utilisateur (C5-160412)

### TP 6 – Évaluation(s) de Méthodes

Carl FRÉLICOT – Dpt Info / Lab MIA

#### 1. Données des Exercices du TD 6

- 1-1) Chargez les données du fichier `terminale.xlsx`, puis mettez en œuvre la chaîne de traitements permettant d'afficher la matrice de confusion du reclassement par la règle des 2-PPV avec la distance euclidienne usuelle.
- 1-2) Comment connaître les élèves qui ont été mal reclassés ?
- 1-3) Qu'obtient-on en sélectionnant *Proportion of Actual ? Proportion of Predicted ?*
- 1-4) À l'aide du composant **Test & Score**, retrouvez les mesures de performance globales.
- 1-5) À partir des courbes *ROC*, pour quel groupe d'élèves la prédiction est la meilleure ?
- 1-6) Constatez que si on teste en *leave-one-out* alors les valeurs chutent.
- 1-7) Imaginez comment comparer plusieurs paramétrages de la règle de prédiction.
- 1-8) Amusez-vous à connecter des règles que vous ne connaissez pas : **Logistic Regression, Tree, SVM, Neural Network**, afin de les comparer.
- 1-9) Pourrait-on réaliser les mêmes expériences à partir de la projection des données dans l'espace des 3 premières composantes principales ? Dans celui des 2 premiers axes discriminants ?
- 1-10) Utilisez le composant **Rank** et interprétez les valeurs de la statistique **ANOVA**. Confirment-elles l'analyse discriminante réalisée au TP5 ?

#### 2. Données des Exercices du TD 6

Chargez les données en dimension 2 (fichier `TD6-Ex2.xlsx`) puis mettez en œuvre les traitements permettant de retrouver des résultats du TD.

#### 3. Données Réelles – Wine

- 3-1) Importez les données `wine.tab` puis réalisez l'analyse discriminante linéaire.
- 3-2) Évaluez les performances d'une ou plusieurs règles de classement dans l'espace discriminant, selon diverses stratégies
- 3-3) De même, à partir d'un nombre opportun de composantes principales.
- 3-4) Finalement quel méthode de prédiction suggerez-vous ?