

Nous travaillons sous R (RStudio).

1. Données

Nous traitons les données IRIS : [https://fr.wikipedia.org/wiki/Iris_\(jeu_de_données\)](https://fr.wikipedia.org/wiki/Iris_(jeu_de_donn%C3%A9es))

Elles sont disponibles le package « dataset » automatiquement associé à R. Nous pouvons y accéder avec la commande `data(iris)`.

2. Arbre de décision sur les données iris

Nous nous intéressons à l'interprétation des frontières dans l'espace de représentation dans cet exercice. Il n'est donc pas nécessaire de partitionner les données en échantillons d'apprentissage et de test.

A rendre : un rapport .docx issu d'un projet R Markdown

1. Chargez en mémoire les données IRIS avec la commande `data(iris)`
2. Quel est le nombre d'observations ? Le nombre de variables ? Enumérez les.
3. SPECIES est la variable cible. Listez ses modalités et leurs fréquences absolues.
4. Construisez un arbre de décision – avec les paramètres par défaut – permettant de prédire SPECIES à partir des autres variables disponibles.
5. Affichez l'arbre. Utilisez la commande `rpart.plot()` du package éponyme. Quelles sont les variables sélectionnées dans l'arbre ?
6. Affichez la matrice de confusion calculée sur les données ayant permis d'élaborer le modèle. Combien d'observations sont mal classées ? Quelles en sont les caractéristiques (Classe d'appartenance vs. Classe prédite) ?
7. Appuyez-vous sur ces variables pour projeter les observations dans le plan (*on ne devrait avoir que 2 variables dans l'arbre*). Distinguez les points selon leur classe d'appartenance en utilisant des couleurs différentes (vert pour « setosa », rouge pour « versicolor », bleu pour « virginica »).
8. Tracez alors dans ce graphique les frontières induites par l'arbre de décision ([abline](#)).
9. Rapprochez ce que vous observez avec ce que vous avez constaté avec la matrice de confusion (question n°6). Quel commentaire peut-on faire ?