

## TP 2 : CART

**Vous aurez besoin dans ce TP d'utiliser les librairies R suivantes :**

*rpart* et *rpart.plot*

### Classification

1. Charger le jeu de donnée *iris* dans R.
2. Visualiser la base : combien d'espèce, quelle proportions, statistiques descriptives, visualisation graphique, etc.
3. Implémenter l'algorithme : expliquer l'espèce à partir des mesures sur les pétales et sépales
  - a. Construire l'arbre maximal ;
  - b. Afficher les résultats obtenus en tapant le nom de l'objet créé ;
  - c. Afficher graphiquement l'arbre avec *rpart.plot*. Quelle variable explicative semble la plus importante ?
  - d. Accéder aux attributs de l'arbre. Quelle variable explicative est finalement la plus importante ?
  - e. Afficher l'attribut *cptable* avec *plotcp*. Analyser.
  - f. Élaguer l'arbre grâce à la commande *prune* au meilleur coefficient de complexité.
  - g. Afficher et analyser l'arbre élagué.
  - h. Effectuer des prévisions pour un jeu de données quelconques.
4. Retirer aléatoirement 5 observations, et refaire les analyses. Comparer.

### Régression

1. Charger le jeu de données *cu.summary* dans R
2. Visualiser et comprendre la base. Nous tentons d'expliquer le prix des véhicules en fonction de l'ensemble des autres caractéristiques fournies.
3. Construire l'arbre maximal. Quel paramètre supplémentaire faut-il renseigner par rapport à la première question ? Retrouver la valeur de l'homogénéité de la racine.
4. Afficher graphiquement l'arbre.
5. Mettre en avant l'effet de surapprentissage, puis élaguer l'arbre.
6. Quelle est le pourcentage de la variance non-expliquée par l'estimateur.
7. Retirer aléatoirement 10 observations et refaire les analyses. Comparer.