

Homework 4: Practicals in R

L'apprentissage automatique consiste à résoudre un problème à partir d'expériences. Il se classe en deux catégories :

- L'apprentissage supervisé : A chaque donnée est associée la sortie observée. L'apprentissage revient à trouver la relation entre l'entrée et la sortie. La sortie peut être de type numérique, dans ce cas nous parlerons de régression, ou de type catégorique, nous parlerons de classification.
- L'apprentissage non-supervisé: Seules les données sont connues. Le système est amené à découvrir leur structure et à les classer.

Dans ce travail, nous allons nous familiariser à R par un cas simple d'apprentissage supervisé (régression) et nous nous pencherons plus particulièrement sur le sur-apprentissage (overfitting).

Partie 1: Effectuer une prédiction

- Importez les données du fichier « friedman1.att » joint à ce travail. Celui-ci contient un header: les variables d'entrées sont notées x_1 à x_{10} et la variable de sortie est notée y . Chaque ligne représente un échantillon.
- Séparez vos données en deux groupes. Le premier, que nous appellerons ensemble d'apprentissage, contiendra 90% des échantillons, et le second, que nous appellerons ensemble de test, contiendra les 10% restant.
- Construisez un modèle linéaire à l'aide de la fonction $lm()$ sur l'ensemble d'apprentissage. N'oubliez pas de spécifier vos variables d'entrée et votre variable de sortie.
- En utilisant votre modèle, effectuez une prédiction sur votre ensemble de test et calculez l'erreur quadratique moyenne de celle-ci. Soit y_1, y_2, \dots, y_k les valeurs réelles des variables de sortie d'un ensemble de taille k et y'_1, y'_2, \dots, y'_k les valeurs prédites pour ce même ensemble, alors l'erreur quadratique moyenne e se calcule selon la formule suivante:

$$e = \sum (y_i - y'_i)^2$$

Partie 2: Sur-apprentissage

- Exécutez le code R du fichier « overfitting.R » joint à ce travail soit en copiant le code dans une console R, soit à l'aide de la commande `source(« filename »)`.
- Commentez ce que vous observez.