

Les forêts aléatoires avec R

Robin Genauer et Jean-Michel Poggi

2019

Presses universitaires de Rennes
www.pur-editions.fr

LES forêts aléatoires sont une méthode d'apprentissage statistique qui fait aujourd'hui partie des outils centraux des statisticiens ou autres data scientists. Introduites par Leo Breiman en 2001, elles sont depuis intensément utilisées dans de nombreux domaines d'application (comme l'écologie, la prévision de la pollution ou encore la santé), du fait des très bonnes performances de l'algorithme en prédiction, mais aussi de leur généralité, n'imposant que très peu de restrictions sur la nature des données. En effet, elles sont adaptées aussi bien à des problèmes de classification supervisée qu'à des problèmes de régression. De plus, elles permettent de prendre en compte un mélange de variables explicatives qualitatives et quantitatives. Enfin, elles sont capables de traiter des données standards pour lesquelles le nombre d'observations est plus élevé que le nombre de variables, mais se comportent également très bien dans le cas de données de grande dimension où le nombre de variables est très important.

Ce livre est une présentation statistique des forêts aléatoires, orientée vers les applications. Il s'adresse donc en premier lieu aux étudiants de filières comportant des enseignements de la statistique mais aussi bien entendu aux praticiens du domaine. Pour fixer les idées sur le plan pédagogique, un niveau de licence scientifique est tout à fait suffisant pour tirer profit des concepts, méthodes et outils introduits. Sur le plan informatique, les prérequis sont modestes mais une initiation au langage R est utile pour s'approprier pleinement l'usage des forêts aléatoires.