

PROCEDURE MANUELLE POUR LA CONSTRUCTION D'ARBRES DE REGRESSION SOUS S+

Badih GHATTAS

GREQAM - Université de la Méditerranée

Ghattas@lumimath.univ-mrs.fr

1. Introduction

Les techniques de régression et de classification par arbres (CART, Breiman *et al.* (1984)) constituent un outil de modélisation et de prévision simple et de plus en plus répandu par la diversité des applications qui lui font appel. Comme beaucoup de méthodes statistiques, ces techniques sont basées sur l'estimation d'une espérance conditionnelle (dans le cas des arbres de régression) en minimisant une erreur quadratique, éventuellement pénalisée. Le logiciel Splus, par exemple, dispose des outils nécessaires à la mise en œuvre de ces techniques.

Le modèle obtenu avec ces méthodes peut être visualisé par un arbre (cf figure 1). Selon les applications considérées, l'arbre *optimal* n'est pas satisfaisant : sa taille peut être soit trop petite soit trop grande, ou les règles binaires figurant aux nœuds de l'arbre peuvent être incohérentes à la réalité physique dont sont issues les données.

Dans un contexte particulier, qui est celui de données issues de l'environnement (la pollution atmosphérique), nous proposons une procédure dite *manuelle* de construction d'arbres de régression et de classification.

Cette procédure permet de construire des arbres en combinant le critère quadratique usuel, et d'autres critères qui sont souvent subjectifs et qui dépendent du problème considéré: la structure de l'arbre, le contenu des feuilles et l'objectif de la modélisation.

Après un bref rappel de ces techniques, nous décrivons cette procédure en l'illustrant par la prévision de concentrations maximales quotidiennes de l'ozone. Nous montrons d'autre part qu'une telle procédure "subjective" peut s'avérer plus performante que la procédure classique.