Tree-based Methods-Summary Sheet

ML@LSE

2018/2019

Abstract

Objectives: Understand what tree-based methods are and how they are buildt. Understand how Cross-Validation can be applied to decision trees. Understand how ensemble methods can improve the power of our techniques.

Requirements: Introductory Bootcamp (you can read the slides if you didn't attend). Familiarity with the notions of independence and correlation may be useful.

Keywords: Decision tree, Nodes, Recursive binary splitting, Pruning, Cost complexity, Bootstrapping, Bagging, Random forest, Boosting.

A *DECISION AND CLASSIFICATION TREES*

A.1 Decision trees

a Intuition: what is a decision tree?

b How to build a decision tree?

A.2 Why using decision trees?

a Decision trees vs. Linear models

b Advantages and Disavantages

A.3 An example of Cross-Validation: Pruning

a Pruning

b How to prune a tree using cost complexity?

B *ENSEMBLE METHODS*

B.1 Bagging

a How to reduce the variance of our method?

b Bagging in practice

B.2 Random forest

a Can we do even better?

b Implementing random forests

B.3 Boosting

a Boosting a regression tree

b Why does this work well?