Tree-Based Methods

Lester Cajegas

March 16, 2018

Regression Tree

The goal of regression tree is to find boxes/rectagles $R_1,\ldots,R_j$ that minimize the RSS, given by

$$\sum_{j=1}^{J}\sum_{i\in R_j}(y_i-\hat{y}_{R_j})^2$$

Tree Pruning

Training a regression tree may produce complex structures that can overfit the training data, leading to a poor test set performance. To overcome the said issue, one can build a tree so long as it decreases in the RSS due to each split exceeding some (high) threshold.

Cost Complexity Pruning

Also known as *weakest link pruning*, the *cost complexity pruning* considers a sequence of trees indexed by a nonnegative tuning parameter α .

For each value of α, there corresponds a subtree $T \subset T_0$, such that:

$$\sum_{m=1}^{|T|}\sum_{i:x_i\in R_m}(y_i-\hat{y}_{R_m})^2+\alpha|T|$$