Flag for Review

Record Notes

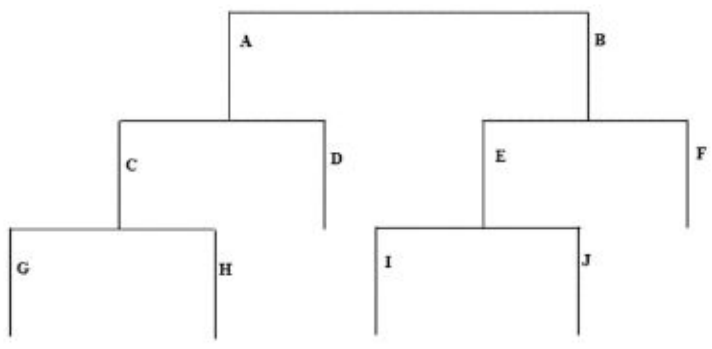
Email the Professor

Question

Difficulty: Mastery

Monitor Difficulty Level

The following diagram illustrates an unpruned decision tree:



The table below gives the residual sum of squares (RSS) at each node:

Pruned decision tree		Unpruned decision tree	
Node	RSS	Node	RSS
A	75	D	32
B	50	F	28
C	55	G and H	19
E	40	I and J	22

You are to use the cost-complexity pruning method to prune the tree, and the value of α of the cost-complexity function is 3. Which of the following is preferred?

A Prune the tree at node A only.

B Prune the tree at node B only.

C Prune the tree at node C only.

D Prune the tree at node E only.

E Do not prune the tree.

Helpful Strategies To Get You Started

Comprehensive Solutions with Alternative Solutions When Available

Help Me Start

The cost-complexity function is $f = \sum_{m=1}^{|T|} \sum_{i:y_i \in R_m} \left(y_i - \hat{y}_{R_m}\right)^2 + \alpha |T|$ where y_i is an observation of node R_m and $|T|$ is the number of nodes.

Solution

If the tree is pruned at node A only:

The number of nodes is $|T| = 4$.

The value of the function is $f = 75 + 22 + 28 + 3\,(4) = 137$.

Graphs and Other Solution Techniques Demonstrated When Available

If the tree is pruned at node B only:

The number of nodes is $|T| = 4$.

The value of the function is $f = 50 + 19 + 32 + 3\,(4) = 113$.

If the tree is pruned at node C only:

The number of nodes is $|T| = 5$.

The value of the function is $f = 55 + 32 + 22 + 28 + 3\,(5) = 152$

If the tree is pruned at node E only:

The number of nodes is $|T| = 5$.

The value of the function is $f = 40 + 19 + 32 + 28 + 3\,(5) = 134$.

If the tree is not pruned:

The number of nodes is $|T| = 6$.

The value of the function is $f = 19 + 22 + 28 + 32 + 3\,(6) = 119$.

We can see that the value of the function is the smallest when the tree is pruned at node B only.

Leave Feedback By Rating Problems

Rate this problem Excellent Needs Improvement Inadequate