

Decision Tree: nagdmc_prune_entropy_tree

Purpose

nagdmc_prune_entropy_tree prunes a decision tree computed by **nagdmc_entropy_tree** by using pessimistic error pruning.

Declaration

```
#include <nagdmc.h>
void nagdmc_prune_entropy_tree(long iproot);
```

Parameters

- 1: **iproot** – long *Input*
On entry: the integer value of the root node of a decision tree as returned by **nagdmc_entropy_tree**.

Notation

None.

Description

Leaf and internal nodes are often removed from a fitted decision tree in order to try to improve the accuracy of predictions on data records not used to fit the decision tree. This process is known as pruning.

A popular method used to prune decision trees based on an entropy criterion is known as pessimistic error pruning (Quinlan, 1987). This method works by testing the (penalised) accuracy of classifications of the lattice at a node against the lattice comprising of the descendents of the node.

Let T denote the lattice structure of a decision tree and T_i the lattice structure of a sub-tree formed by the node i and its descendents (to the level of leaf nodes). Furthermore, for a node i , let n_i and l_i be the number of data records and the number of data records that do not belong to the majority category, respectively.

According to Quinlan, l_i is penalised by 0.5, giving the error: $e(i) = l_i + 0.5$, and T_i is pruned if the following condition is true:

$$e(i) \leq e(T_i) + \left[e(T_i) \left(1 - \frac{e(i)}{n_i} \right) \right]^{\frac{1}{2}},$$

where $e(T_i)$ is the penalised error summed over the leaf nodes in T_i .

Nodes in a decision tree lattice are tested for pruning in a top-down fashion, i.e., by starting at the root node.

References and Further Reading

Quinlan J R (1987) Simplifying Decision Trees *J. Man-Machine Studies* **27** 221–234.

See Also

nagdmc_entropy_tree	computes an decision tree by using an entropy-based criterion.
nagdmc_free_entropy_tree	returns to the operating system memory used by an entropy tree.
nagdmc_load_entropy_tree	loads an entropy tree from a file.
nagdmc_predict_entropy_tree	computes predictions given an entropy tree.
nagdmc_save_entropy_tree	writes an entropy tree to a file.
entropy_tree_ex.c	the example calling program.