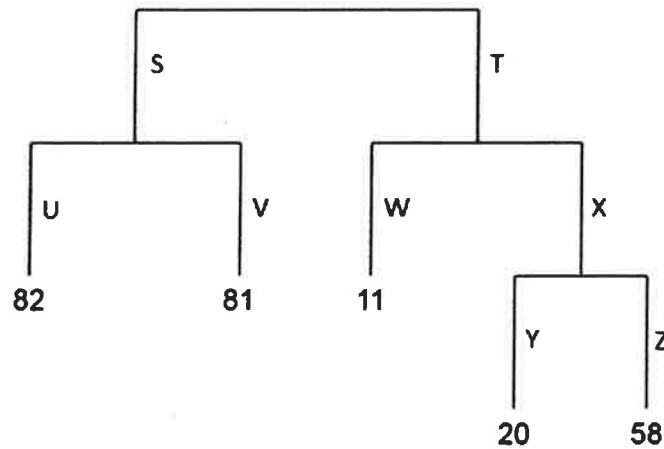


Exam MAS-II Fall 2018

38:

You are given the following unpruned decision tree:



The values at each terminal node are the residual sums of squares (RSS) at that node. The table below gives the RSS at nodes S, T, and X if the tree was pruned at those nodes:

Node	RSS
S	251
T	209
X	86

The RSS for the null model is 486. You use the cost complexity pruning algorithm with the tuning parameter, α , equal to 9 in order to evaluate the following pruning strategies.

- I. No nodes pruned
- II. Prune node S only
- III. Prune node T only
- IV. Prune node X only
- V. Prune both nodes S and X

Determine which pruning strategy is selected.

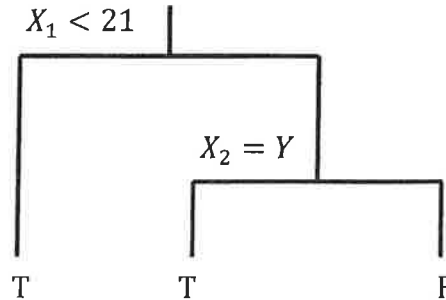
- A. I
- B. II
- C. III
- D. IV
- E. V

CONTINUED ON NEXT PAGE

Exam MAS-II Fall 2018

40.

You are given the following classification decision tree and data set:



i	X_1	X_2	Y
1	12	Y	T
2	23	N	F
3	4	Y	F
4	32	Y	F
5	22	N	T
6	30	Y	T
7	18	N	T

Determine the relationship between the classification error rate, the Gini index, and the cross-entropy, summed across all nodes.

- A. cross-entropy > Gini index > classification error rate
- B. cross-entropy > Gini index = classification error rate
- C. classification error rate > Gini index > cross-entropy
- D. Gini index > cross-entropy > classification error rate
- E. The answer is not given by (A), (B), (C), or (D).

CONTINUED ON NEXT PAGE

9. A classification tree is being constructed to predict if an insurance policy will lapse. A random sample of 100 policies contains 30 that lapsed. You are considering two splits:

Split 1: One node has 20 observations with 12 lapses and one node has 80 observations with 18 lapses.

Split 2: One node has 10 observations with 8 lapses and one node has 90 observations with 22 lapses.

The total Gini index after a split is the weighted average of the Gini index at each node, with the weights proportional to the number of observations in each node.

The total entropy after a split is the weighted average of the entropy at each node, with the weights proportional to the number of observations in each node.

Determine which of the following statements is/are true?

- I. Split 1 is preferred based on the total Gini index.
- II. Split 1 is preferred based on the total entropy.
- III. Split 1 is preferred based on having fewer classification errors.

- (A) I only
- (B) II only
- (C) III only
- (D) I, II, and III
- (E) The correct answer is not given by (A), (B), (C), or (D).