

Student Number

Student Name

Date

CS5483 Lecture Review Question 11

- (i) Explain the term “Information Gain” in Decision Tree.
- (ii) What is the termination condition of Growing tree phase?
- (iii) Given a decision tree, which option do you prefer to prune the resulting rule and why?
 - (a) Converting the decision tree to rules and then prune the resulting rules.
 - (b) Pruning the decision tree and then converting the pruned tree to rules.

Model answer:

- (i) Information Gain means to minimize the information needed to classify the samples in the resultant partitions and reflects the least “impurity” in them.

- (ii) If the decision tree classified the instances correctly, the procedure terminates. If an instance is incorrectly classified, the instance is added to the selected subset of training instances and a new tree is constructed. This process continues until a tree that correctly classify all nonselected instances is created or the decision tree is built from the entire training set.

- (iii) Converting the decision tree to rules and then prune the resulting rules is a better choice.

The reason is the following:

If pruning a subtree, we would remove the subtree completely with method (b). However, with method (a), if pruning a tree, we may remove any precondition of it. The latter is less restrictive.

In other words, choice (a) can provide user more options of removing the whole (subtree) rule or eliminating a precondition of a rule.