Quiz

Easy

Q1. What are tree based classifiers?
   a. Classifiers which form a tree with each attribute at one level
   b. Classifiers which perform series of condition checking with one attribute at a time
   c. Both options except none
   d. None of the options
   Ans: c

Q2. What is gini index?
   a. It is a type of index structure
   b. It is a measure of purity
   c. Both options except none
   d. None of the options
   Ans: b

Q3. Tree/Rule based classification algorithms generate ... rule to perform the classification.
   a. if-then.
   b. while.
   c. do while.
   d. switch.
   Ans: a

Q4. Which of the following sentences are correct in reference to Information gain?
   a. It is biased towards single-valued attributes
   b. It is biased towards multi-valued attributes
   c. ID3 makes use of information gain
   d. The approach used by ID3 is greedy
   Ans: b, c, d

Q5. Cost complexity pruning algorithm is used in?
   a. CART
   b. C4.5
   c. ID3
d. All
Ans: a

Q6. Cost complexity pruning algorithm is used in?
a. CART
b. C4.5
c. ID3
d. All
Ans: a

For the below questions answer as true / false
Q6. Multivariate split is where the partitioning of tuples is based on a combination of attributes rather than on a single attribute.
Ans: True

Q7. CART system cannot find multivariate splits.
Ans: False

Q8. Gain ratio tends to prefer unbalanced splits in which one partition is much smaller than the other.
Ans: True

Q9. The gini index is not biased towards multivalued attributes.
Ans: False

Q10. Gini index does not favour equal sized partitions.
Ans: False

Q11. When the number of classes is large Gini index is not a good choice.
Ans: True

Q12. Attribute selection measures are also known as splitting rules.
Ans: True

Medium

Q1. Which one of these is not a tree based learner?
a. CART
b. ID3
c. Bayesian classifier
d. Random Forest
Ans: c
Q2. Which one of these is a tree based learner?
   a. Rule based
   b. Bayesian Belief Network
   c. Bayesian classifier
   d. Random Forest
   Ans: d

Q3. What is the approach of basic algorithm for decision tree induction?
   a. Greedy
   b. Top Down
   c. Procedural
   d. Step by Step
   Ans: a

Q4. Which of the following classifications would best suit the student performance classification systems?
   a. If...then... analysis
   b. Market-basket analysis
   c. Regression analysis
   d. Cluster analysis
   Ans: a

Consider the following dataset in table 1 for question 5-10 where each record represents the age, income and is a student or not. And we need to classify it as buyers/non-buyers of computer.

Q5. How many records will be in the root initially?
   a. 14  
   b. 9  
   c. 5  
   d. 10  
   Ans: 14

Q6. In the root node how many classes will be there?
   a. 3  
   b. 2  
   c. 4
Q7. The records at the root node will be divided into how many nodes in the next level based on age?
   a. 4
   b. 3
   c. 14
   d. 2
   Ans: 3

Q8. Using ID3 algorithm calculate the information gain for each possible split if you split based on age?
   a. 0.30
   b. 0.40
   c. 0.246
   d. 0.1
   Ans. c

Q9. Calculate the information gain(income)?
   a. 0.1
   b. 0.3
   c. 0.029
   d. 0.35
   Ans. c

Q10. Calculate the information gain(student)?
    a. 0.35
    b. 0.248
    c. 0.151
    d. 0.43
    Ans. c

Q11. Calculate the gain ratio(income)?
    a. 0.024
    b. 0.019
    c. 0.34
    d. 0.112
    Ans. b
Q12. Calculate the gini(income)?
   a. 0.34
   b. 0.443
   c. 0.56
   d. 0.123
   Ans. b

Hard

Q1. How will you counter over-fitting in decision tree?
   a. By pruning the longer rules
   b. By creating new rules
   c. Both By pruning the longer rules’ and ‘By creating new rules’
   d. None of the options
   Ans: a

Q2. What are two steps of tree pruning work?
   a. Pessimistic pruning and Optimistic pruning
   b. Postpruning and Prepruning
   c. Cost complexity pruning and time complexity pruning
   d. None of the options
   Ans: b

Q3. Which of the following sentences are true?
   a. In pre-pruning a tree is 'pruned' by halting its construction early.
   b. A pruning set of class labelled tuples is used to estimate cost complexity.
   c. The best pruned tree is the one that minimizes the number of encoding bits.
   d. All of the above
   Ans: d