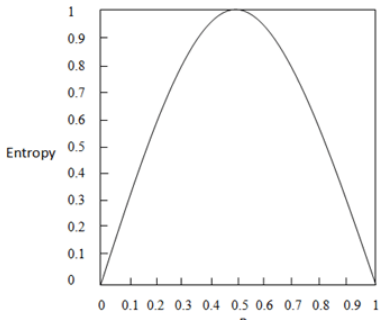
Indian Institute of Information Technology, Allahabad
Data Mining C2 Examination Quiz

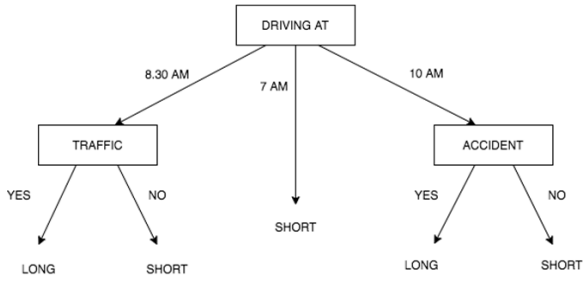
Name: **Date: 25.04.2024**
Roll No:
Section:

Instruction:

1. In a binary classification problem using SVM, if a data point lies on the wrong side of the decision boundary and violates the margin, it is known as a:
    a. Support vector
    b. Margin violator
    c. Outlier
    d. Prototype
2. What is the margin of a linear SVM classifier if the distance between the support vectors and the decision boundary is 2 units, and the margin is symmetric around the decision boundary?
    a. 1
    b. 2
    c. 4
    d. 6
3. Given:
    A doctor knows that Cold causes fever 50% of the time
    The prior probability of any patient having cold is 1/50,000
    The prior probability of any patient having fever is 1/20
    If a patient has fever, what's the probability he/she has cold via naive bayes?
    a. 0.0002
    b. 0.005
    c. 0.002
    d. 0.0005
4. Naive Bayes classifiers are particularly effective when the feature dimensions are _____, and the training data is _____.
    high, sparse
5. What is the entropy at P = 0.5 from the given figure?



    a. 0.5
    b. 0.88
    c. 1
    d. 0.9
6. Consider the figure. If person A starts driving at 8:30 AM and there are no other vehicles on the road, and another person B starts driving at 10 AM and there is an accident on the road, what will be the commute time of B and A, respectively?



    a. LONG, LONG
    b. LONG, SHORT
    c. SHORT, LONG
    d. SHORT, SHORT
7. Which of the following statements is not true about the pruning in the decision tree?
    a. When the decision tree is created, many of the branches

Indian Institute of Information Technology, Allahabad
Data Mining C2 Examination Quiz

Name: **Date: 25.04.2024**
Roll No:
Section:

Instruction:

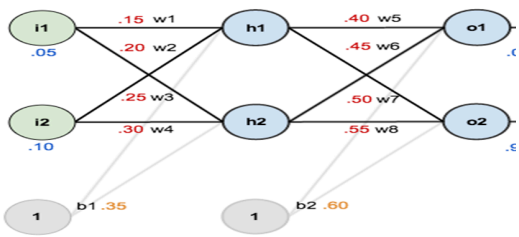
will reflect anomalies in the training data due to noise

b. The overfitting happens when the learning algorithm continues to develop hypothesis that reduce training set error at the cost of an increased test set errors

c. It optimizes the computational efficiency

d. It reduces the classification accuracy

8. Which of the following best describes the role of the learning rate in the backpropagation algorithm?
   a. It defines the maximum number of iterations for weight updates.
   b. It determines how much the weights are adjusted during each step.
   c. It increases the computation speed of the algorithm.
   d. It specifies the initial values of the weights.
9. Calculate the output for h1 (hidden layer) using the sigmoid activation function_____.



0.5944

10. A Regression tree is built through a process known as_____.
    binary recursive partitioning
11. You have an association rule "Bread -> Butter (80% confidence, 3% support)". If 100 customers buy bread, how many would be expected to also buy butter based on this rule?
    a. 20
    b. 30
    c. 80
    d. 240
12. You are evaluating two classifiers on a dataset with 1000 data points. Classifier A has a True Positive Rate (TPR) of 0.8 and a False Positive Rate (FPR) of 0.1. Classifier B has a TPR of 0.75 and an FPR of 0.05. Which classifier has a higher lift?
    a. **Classifier A**
    b. Classifier B
    c. They have the same lift.
    d. Impossible to determine without precision values.
13. When a classifier performs well on the training data but poorly on unseen data, it is a sign of _____.
    overfitting
14. How can classifier accuracy be misleading in some situations?
    a. It doesn't consider the cost of misclassification.
    b. It assumes a balanced class distribution.
    c. It doesn't account for overfitting to the training data.
    d. All of the above.
15. The Naive Bayes classifier assumes that the features are _____ given the class.
    independent