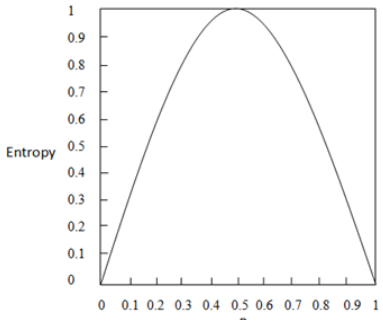
Indian Institute of Information Technology, Allahabad
Data Mining c2 Examination Quiz

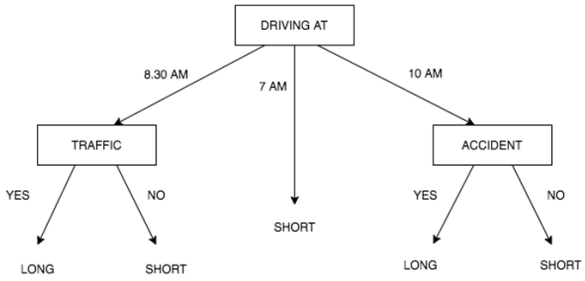
Name: **Date: 25.04.2024**
Roll No:
Section:

Instruction:

1. If the training data in an SVM problem is linearly separable, what is the purpose of introducing a slack variable in the optimization problem?
    a. To reduce overfitting
    b. To handle noise in the data
    c. To allow for a margin of error in classification
    d. To speed up the convergence of the optimization algorithm
2. What is the margin of a linear SVM classifier if the distance between the support vectors and the decision boundary is 3 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 1/50,00
    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. In Bayesian classification, the class of a test instance is determined by maximizing the _____. **posterior probability of each class given the features.**
5. What is the entropy at P = 0.3 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 A and B, respectively?



    a. LONG, LONG
    b. LONG, SHORT
    c. SHORT, LONG
    d. SHORT, SHORT
7. Which of the following is not a Pruning technique?

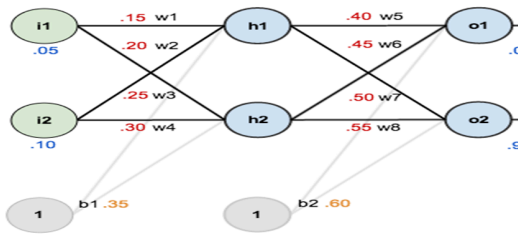
Indian Institute of Information Technology, Allahabad
Data Mining c2 Examination Quiz

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

Instruction:

a. Cost-based pruning
b. Cost complexity pruning
c. Minimum error pruning
d. Maximum error pruning

8. Why are activation functions important in neural networks using backpropagation?
   a. They normalize the input values.
   b. They introduce non-linearity into the model.
   c. They reduce the dimensionality of the input data.
   d. They increase the speed of learning.
9. Calculate the output for h2 (hidden layer) using the sigmoid activation function_____.



0.5962

10. Regression trees follow a _____greedy approach.
    top-down
11. You are analyzing a grocery store dataset using ARM. The minimum support threshold is set to 2%, and the minimum confidence threshold is set to 70%. If there are 10,000 transactions, what minimum number of times does a product combination (A, B) need to appear together to be considered a frequent itemset?
    a. 10
    b. 20
    c. 140
    d. 200
12. You are analyzing clickstream data to identify product recommendations. The average click-through rate (CTR) for all products is 2%. An association rule suggests that users who view product X are 4 times more likely to click on product Y than the average CTR. What is the estimated CTR for product Y given a user views product X?
    a. 4%
    b. 6%
    c. 8%
    d. Cannot be determined without additional information
13. When a classifier performs poorly on both the training data and unseen data, it is a sign of _____.
    underfitting
14. Which of the following is NOT a standard measure of interestingness in association rule mining?
    a. Support
    b. Confidence
    c. Precision
    d. Conviction
15. Naive Bayes classifiers are particularly effective when the feature dimensions are _____ and the training data is _____.
    high, sparse