

JAYPEE UNIVERSITY OF INFORMATION TECHNOLOGY, WAKNAGHAT

TEST -1 EXAMINATION Jan-Jun 2024

B.Tech-VI Semester (Bio-Informatics)

COURSE CODE (CREDITS): 18B11BI611 (2)

MAX. MARKS: 15

COURSE NAME: Machine Learning for Bio-Informatics

COURSE INSTRUCTORS: Praveen Modi, Dr. Aman Sharma, Dr. Shubham Goel MAX. TIME: 1:00Hr

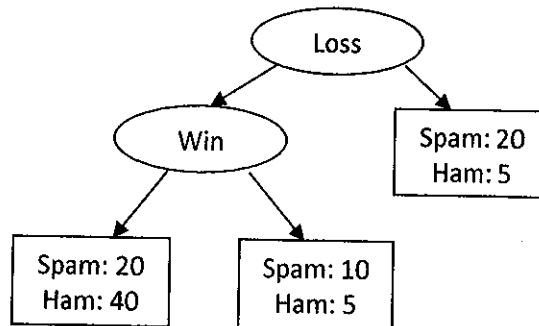
Note: All questions are compulsory. Marks are indicated against each question in square brackets. Write the answer of the question belonging to the same part in the same order.

Q1. What are the various methods to find the missing values for an attribute? [2] (CO1)

Q2. Which statements are TRUE or False? [2] (CO2)

- (a) The i -th principal component is taken as the direction that is orthogonal to the $(i+1)$ -th principal component and maximize the remaining variability.
- (b) Different individual principal component are linearly uncorrelated.
- (c) The principal component with the largest eigen value maximizes the reconstruction error.
- (d) Ignoring the small eigen value component will not loss the information

Q3. In an Email- Spam detection classifier system, if the email contains the word "Loss" it is classified as spam otherwise, the word "Win" decides whether it gets labelled as spam or ham. The following decision is constructed. Construct the confusion matrix and calculate the accuracy and F1-Score performance measurement? [5] (CO2)



Q4 Explain the basic concept, strength and weakness of ID-3 algorithm to construct the decision tree? [3] (CO2)

Q5. What do you understand by the sensitivity and specificity performance measurement? Give their significance with example? [2+1] (CO1)