

**BIRLA INSTITUTE OF TECHNOLOGY, MESRA, RANCHI
(END SEMESTER EXAMINATION)**

CLASS: BCA
BRANCH: BCA

SEMESTER : VI
SESSION : SP/2025

SUBJECT: CA373 INTRODUCTION TO OF MACHINE LEARNING

TIME: 3 Hours

FULL MARKS: 50

INSTRUCTIONS:

1. The question paper contains 5 questions each of 10 marks and total 50 marks.
2. Attempt all questions.
3. The missing data, if any, may be assumed suitably.
4. Before attempting the question paper, be sure that you have got the correct question paper.
5. Tables/Data hand book/Graph paper etc. to be supplied to the candidates in the examination hall.

- | | | CO | BL | | | | | | | | | | | | |
|--|-----|-----|----|------------|----|---|---|---|---|----------|----|----|----|----|----|
| Q.1(a) Describe two different real-world problems where supervised learning would be applicable. What algorithm would you choose for each of the problem and why? | [5] | CO1 | 3 | | | | | | | | | | | | |
| Q.1(b) A small company makes two products, A and B. They organize the data in a 2x2 matrix that shows how the sales of one product relate to the other:
$S = \begin{bmatrix} 5 & 2 \\ 2 & 3 \end{bmatrix}$
Find the eigenvalues and eigenvectors of the matrix S. Which eigenvector represents the direction of highest variation in sales? If the company wants to reduce the data to one main trend, which eigenvector should they keep? | [5] | CO2 | 4 | | | | | | | | | | | | |
| Q.2(a) Describe the parameters computation method for single linear regression. Write the assumptions for linear regression model and method to evaluate performance of the model. | [5] | CO2 | 2 | | | | | | | | | | | | |
| Q.2(b) A teacher keeps track of whether students pass or fail the exam based on how many hours they study. Use logistic regression to model the chance of passing based on study hours. Using your model, predict the probability that a student who studies for "3" hours will pass. | [5] | CO4 | 4 | | | | | | | | | | | | |
| Q.3(a) Describe the Support Vector Machine (SVM) and its kernel trick using suitable examples and diagrams. Discuss the merits and demerits of SVM. | [5] | CO3 | 3 | | | | | | | | | | | | |
| Q.3(b) A company wants to build a system that automatically filters spam emails. They collect a dataset where each email is labelled as spam (1) or not spam (0) based on words like "free", "win", or "urgent".
They plan to use a classification algorithm like Naive Bayes, or Decision Tree. How Do You Design an Email Spam Filter for them? Explain steps to implement using a suitable algorithm of your choice. How will you evaluate the performance of your model? | [5] | CO4 | 5 | | | | | | | | | | | | |
| Q.4(a) What is PCA and why it is used? Explain each of the steps involved in PCA using a suitable example. | [5] | CO2 | 2 | | | | | | | | | | | | |
| Q.4(b) Describe various distance measures used for hierarchical clustering. Explain the method of agglomerative clustering using following data: | [5] | CO3 | 4 | | | | | | | | | | | | |
| <table border="1" style="margin-left: auto; margin-right: auto; border-collapse: collapse;"> <thead> <tr> <th style="padding: 2px;">Process ID</th> <th style="padding: 2px;">1</th> <th style="padding: 2px;">2</th> <th style="padding: 2px;">3</th> <th style="padding: 2px;">4</th> <th style="padding: 2px;">5</th> </tr> </thead> <tbody> <tr> <td style="padding: 2px;">CPU Time</td> <td style="padding: 2px;">12</td> <td style="padding: 2px;">31</td> <td style="padding: 2px;">27</td> <td style="padding: 2px;">18</td> <td style="padding: 2px;">21</td> </tr> </tbody> </table> | | | | Process ID | 1 | 2 | 3 | 4 | 5 | CPU Time | 12 | 31 | 27 | 18 | 21 |
| Process ID | 1 | 2 | 3 | 4 | 5 | | | | | | | | | | |
| CPU Time | 12 | 31 | 27 | 18 | 21 | | | | | | | | | | |
| Q.5(a) Explain the followings:
i. Backpropagation Learning
ii. Regularization | [5] | CO2 | 2 | | | | | | | | | | | | |
| Q.5(b) Describe the perceptron model of ANN using a suitable diagram. Prove that a single perceptron is not able to solve the nonlinear classification problem. | [5] | CO4 | 4 | | | | | | | | | | | | |