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

CLASS: IMSC | **SEMESTER :VIII**
BRANCH: MATHS AND COMPUTING | **SESSION : SP/2025**

SUBJECT: CA640 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)	Let X be a discrete random variable with the probability mass function: $P(X=x) = k \times 0.5^x$, for x=1,2, 3…,5. Find the value of k that makes this a valid probability distribution. Find P(X=2), P(X<2) and P(X>2).	[5]	2	2
Q.1(b)	Define linear regression. Imagine a classification task where the dataset has binary features and class labels. Under the Naïve Bayes assumption of feature independence, the probability P(y=1|x) can be represented by a logistic function. Assume that the data approximately satisfies the independence conditions required by Naïve Bayes. Now, consider training both a Naïve Bayes classifier and a logistic regression model on this dataset. Which of these two models would you choose when only a small amount of training data is available? Justify your answer. How does your choice change as the size of the training data increases, eventually becoming very large?	[5]	3	3

Q.2(a) Define mutual information between two random variables X and Y? How information gain is important in decision tree learning model? [5] CO 3 BL 3

For a binary classification problem, consider the training examples shown in the following table. The attributes, A_1 and A_2 can take either True or False values, whereas A_3 is a continuous attribute. The Target Class can be either + (positive) or − (negative). Answer the following.

a) What is the entropy of this collection of training examples?
b) What are the information gains of A_1 and A_2 relative to these training examples?

Instance	A_1	A_2	A_3	Target Class
1	*True*	*True*	1.0	+
2	*True*	*True*	6.0	+
3	*True*	*False*	5.0	−
4	*False*	*False*	4.0	+
5	*False*	*True*	7.0	−
6	*False*	*True*	3.0	−
7	*False*	*False*	8.0	−
8	*True*	*False*	7.0	+
9	*False*	*True*	5.0	−

Q.2(b) Using the SVM algorithm, find the SVM classifier for the following data. [5] BL 3

Example	X_1	X_2	Class
1	2	1	+1
2	4	3	-1

PTO

Q.3(a) Look at the dataset presented in the table below. The features A, B, and C are binary attributes, meaning they can each have a value of either 0 or 1. The target class label can be either + (positive) or − (negative).	[5]	5	3

Instance	A	B	C	Class
1	0	0	1	−
2	1	0	1	+
3	0	1	0	−
4	1	0	0	−
5	1	0	1	+
6	0	0	1	+
7	1	1	0	−
8	0	0	0	−
9	0	1	0	+
10	1	1	1	+

Answer the following questions.

a) Estimate the conditional probabilities for the following: $P(A=0|+)$, $P(B=1|+)$, $P(B=1|-)$, $P(C=1|+)$, $P(C=1|-)$, $P(A=0|-)$

b) Use the conditional probabilities in part (a) to predict the class label for a given test sample, (A = 0, B = 1, C = 1), using the Naive Bayes approach.

c) Are the variables, A and B, independent with values, A = 1 and B = 1?

d) Are these variables, A and B, conditionally independent with values, A = 1 and B = 1, given the class '+'?

Q.3(b) Design a perceptron to implement the NAND logical Boolean function. Show that XOR logical function cannot be represented using a perceptron. What is an activation function and give examples of different types of activation functions.	[5]	1	2
Q.4(a) Given the (x, y)-coordinates of four data points in two-dimensional space: (4, 1), (2, 3), (5, 4) and (1, 0), calculate the first principal component. Show your calculations in details.	[5]	5	3
Q.4(b) Consider a dataset with the following points in a two-dimensional space: A (2, 4), B (5, 9), C (1, 2), D (7, 7), E (9, 6). Use Agglomerative Clustering with the Euclidean distance as the similarity measure. Begin by considering each data point as an individual cluster. Then, step-by-step, merge the two clusters that are closest together based on the single-linkage criterion (minimum distance between any two points in different clusters). Continue this process until only two clusters remain.	[5]	4	3
Q.5(a) What are ensemble methods? Explain the difference between Bagging and Boosting. Does Random Forest needs pruning?	[5]	3	2
Q.5(b) Write short notes on	[5]	5	2

a) Precision Recall Tradeoff

b) Ada Boosting

:::::28/04/2025:::::E