

Roll No.: _____

Amrita Vishwa Vidyapeetham
Amrita School of Computing, Coimbatore

B.Tech. Second Assessment Examinations – December 2022

Fifth Semester

Computer Science and Engineering

19CSE305 Machine Learning

Duration: Two hours

Maximum: 50 Marks

CO#	Course Outcomes
CO01	Understand issues and challenges of machine learning: data, model selection, model complexity
CO02	Design and implement various machine learning algorithms in a range of real-world applications
CO03	Understand strengths and weaknesses of many popular machine learning approaches
CO04	Analyze the underlying mathematical relationships within and across Machine Learning algorithms
CO05	Apply the paradigms of supervised and un-supervised learning

Answer all questions in all 3 parts

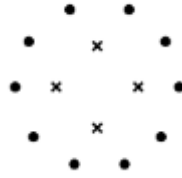
Part A

(5 X 3 = 15 marks)

1. Assume that you are given a dataset with the total score of a team batting first, along with whether they won the match (binary variable). If a linear regression model ($y = mx + c$) estimate of a set of data points is derived as $m = 0.09$ and $c = -11$, use logit to derive a function that computes the probability that the team may win the match. Use the function to compute the probability that the team may win if they have scored 120 runs. [3] [CO4] [BTL3]
2. a. Explain Naïve bayes theorem using equation. [1.5] [CO4] [BTL1]
b. Consider a school with a total population of 100 persons. These 100 persons can be seen either as 'Students' and 'Teachers' or as a population of 'Males' and 'Females'. With below tabulation of the 100 people, what is the conditional probability that a certain member of the school is a 'Teacher' given that he is a 'Man'? [1.5] [CO4] [BTL2]

	Female	Male
Teacher	8	12
Student	32	48

3. For the sample data given below 'x' represents a sample belonging to a class and black dot represents sample not belonging to the class. What will be the transfer function used by SVM to convert the data into linearly separable data? Show the graph after transforming the points. What will the sign of data 'x' after transformation. [3] [CO5] [BTL2]



4. Elaborate on any two advantages of K-Nearest neighbour algorithm. [3] [CO3] [BTL1]
5. How does bagging ensemble aggregate the predictions made by different base models/weak learners for classification and regression problems? Give the mathematical notations of aggregation strategies for regression problems. [3] [CO5] [BTL2]

Part B

(3 X 5 = 15 Marks)

6. What is Random Forest Bagging Ensemble? Write down the details about how the base learners in a Random Forest ensemble are getting added, how diversity is introduced among base models, how the overall prediction is made, and how the performance evaluation is done with a neat sketch that shows a typical bagging ensemble. [5] [CO5] [BTL2]
7. The height, age and weight chart of babies is given. Apply k-nearest neighbour (k=3, Euclidean distance measure) to obtain the weight value of the test case with height = 1.24 m and age = 10 years. [5] [CO4] [BTL3]

Height (in m)	Age (in years)	Weight (in kg)
0.81	2	10.1
0.88	3	11.8
0.96	4	13.5
1.02	5	14.8
1.08	6	16.3
1.19	8	19.7

8. From the table given below classify a Red Domestic SUV using naive bayes equation. [5] [CO4] [BTL3]

Example No.	Color	Type	Origin	Stolen?
1	Red	Sports	Domestic	Yes
2	Red	Sports	Domestic	No
3	Red	Sports	Domestic	Yes
4	Yellow	Sports	Domestic	No
5	Yellow	Sports	Imported	Yes
6	Yellow	SUV	Imported	No
7	Yellow	SUV	Imported	Yes
8	Yellow	SUV	Domestic	No
9	Red	SUV	Imported	No
10	Red	Sports	Imported	Yes

Part C

(2 X 10 = 20 Marks)

9. Given (-1,3), (-1,-1), (3,-1), and (3,3) represents positive data samples; (0,2), (0,0), (2,2) and (2,0) represents negative data samples. If SVM applies the following transform function on positive samples and not on negative samples what should be the condition used for that?

$$\phi \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{cases} 6 - x_2 + |x_1 - x_2| \\ 6 - x_1 + |x_1 - x_2| \end{cases}$$

where $|x|$ indicate the absolute value of x . Transform the points and plot the data points before and after transformation. Indicate the support vectors in the graph. Using the support vectors frame the equation and solve for decision plane. Based on the decision plane calculate and show the class of (2,-2). Draw the decision surface on the graph. [10] [CO4] [BTL3]

10. Assume that a dataset is built to explain the difficulty level of a question (Yes/No) based on the course, length of the question, and whether the question involves math. Build a decision tree model for the sample dataset given below. Draw the tree model with the corresponding 'information gain' calculations at each depth level. [10] [CO4] [BTL3]

Course	Length	Math	Difficult
19CSE305	Short	Yes	No
19CSE302	Medium	Yes	Yes
19CSE301	Long	No	No
19CSE305	Medium	Yes	Yes
19CSE302	Short	No	No
19CSE301	Medium	No	No
19CSE305	Long	Yes	Yes
19CSE302	Long	Yes	Yes
19CSE301	Short	Yes	Yes
19CSE305	Short	No	No

Course Outcome /Bloom's Taxonomy Level (BTL) Mark Distribution Table)

CO	Marks	BTL	Marks
CO01		BTL 1	4.5
CO02		BTL 2	12.5
CO03	3	BTL 3	33
CO04	36	BTL 4	
CO05	11	BTL 5	
CO06		BTL 6	