

Register No.: Name:

SAINTGITS COLLEGE OF ENGINEERING (AUTONOMOUS)

(AFFILIATED TO APJ ABDUL KALAM TECHNOLOGICAL UNIVERSITY, THIRUVANANTHAPURAM)

SEVENTH SEMESTER B.TECH DEGREE EXAMINATION (R), DECEMBER 2023**COMPUTER SCIENCE AND ENGINEERING****(2020 SCHEME)****Course Code : 20CST411****Course Name: Machine Learning****Max. Marks : 100****Duration: 3 Hours****PART A*****(Answer all questions. Each question carries 3 marks)***

1. Differentiate between unsupervised learning and reinforcement learning.
2. What is a categorical feature? Which of the following is a categorical feature?
 - i) Your weekly expenditure in rupees.
 - ii) Area (in sq. centimeter) of your laptop screen.
 - iii) The color of the curtains in your room.
3. Define the terms a) Entropy b) Information Gain.
4. What is overfitting?
5. Calculate the output y of a three-input neuron with bias. The input feature vector is $(x_1, x_2, x_3) = (0.8, 0.6, 0.4)$ and weight values are $(w_1, w_2, w_3, b) = (0.2, 0.6, -0.3, 0.35)$. Use binary sigmoid, bipolar Sigmoid function and ReLU activation functions.
6. Distinguish between linearly separable and linearly inseparable problems.
7. Consider two data points in three dimension, A(12,3,9) and B(6,5,15). Calculate a) City block distance (b) Chessboard distance (c) Euclidean distance between A and B.
8. Compare and contrast clustering and classification.
9. Define the terms Precision, Recall and Accuracy for a classification problem.
10. Define cross validation. Briefly explain different cross validation techniques.

PART B***(Answer one full question from each module, each question carries 14 marks)*****MODULE I**

11. a) Three factories manufacture TV sets. The share of manufacture of the three factories A, B and C are 20%, 30% and 50%, respectively. The above factories produce 3%, 4% and 5% defective TVs. For an arbitrarily picked defective TV, what is the probability that it is manufactured by A, B or C. (8)
- b) Explain maximum a posteriori estimation. (6)

OR

12. a) Identify suitable learning method in each case. Justify your answer.
- Predicting if a particular route between two points has traffic jam or not based on the travel time of vehicles.
 - Predicting the monthly sales of a cloth store in rupees.
 - Predicting if a new edible item is sweet or spicy based on the information of the ingredients, their quantities, and labels (sweet or spicy) for many other similar dishes. (8)
 - Grouping related documents from an unannotated corpus.
- b) Explain maximum likelihood estimation. A coin is flipped 100 times. Given that there were 55 heads, find the maximum likelihood estimate for the probability p of heads on a single toss. (6)

MODULE II

13. a) Identify the first splitting attribute for decision tree by using ID3 algorithm with the following dataset

Major	Experience	Tie	Hired?
CS	Programming	pretty	No
CS	Programming	pretty	No
CS	Management	pretty	yes
CS	Management	ugly	yes
Business	Programming	pretty	yes
Business	Programming	ugly	yes
Business	Management	pretty	No
Business	Management	pretty	No

- b) Explain Logistic regression and Linear regression with example. (8)

OR

14. a) The following table shows the data collected from the touristic center in a city during one month in the summer, counting the number of people that arrive at the square at the same time every day. The relationship between the weather and tourism levels is shown below. (9)

Temperature	Number of Visitors
12	87
21	150
20	110
25	90
17	85
15	70
13	88

Find regression function to predict number of Visitors from temperature. Also predict the number of visitors if the temperature is 27.

- b) What is Bayes theorem? Write Naive Bayes algorithm. (5)

MODULE III

15. a) Explain how Support Vector Machine can be used for classification of linearly inseparable data. (7)
- b) What is a Perceptron? Explain the working of a perceptron with a neat diagram (7)

OR

16. a) With suitable example explain how back propagation algorithm works (8)
- b) Why do you use the kernel trick? Explain different kernel functions. (6)

MODULE IV

17. a) Write basic algorithm of agglomerative clustering. Show the final result of hierarchical clustering with single link by drawing a dendrogram.

	A	B	C	D	E	F
A	0					
B	0.71	0				
C	5.66	4.95	0			
D	3.61	2.92	2.24	0		
E	4.24	3.54	1.411	1	0	
F	3.2	2.5	2.5	0.5	1.12	0

- b) Discuss about the significance of dimensionality reduction technique. (6)

OR

18. a) Find the 2 cluster centers after second epoch for the following 6 samples using the K-means algorithm and Euclidean distance. (8)

$A1=(3,6)$, $A2=(4,9)$, $A(5,8)$, $A4=(2,7)$, $A5=(9,7)$ and $A6=(1,7)$ Suppose that the initial seeds are $A1$ and $A3$.

- b) Explain expectation maximization (EM) for soft clustering. (6)

MODULE V

19. a) Explain receiver operating characteristic space in machine learning. (7)
b) What is Bootstrap Sampling in Machine Learning and Why is it Important? How to test machine learning models using bootstrapping? (7)

OR

20. a) Explain the terms i) bagging ii) boosting and iii) voting. (6)
b) Suppose the dataset contains 80 records on cancer diagnosis of which 55 are relevant to the investigation. A search was conducted on the cancer diagnosis and 50 records were retrieved. Of the 50 records retrieved, 40 are relevant. Construct the confusion matrix for the search and calculate the precision and recall scores for the search. (8)
