Register No.: ....   Name: ....

SAINTGITS COLLEGE OF ENGINEERING (AUTONOMOUS)

(AFFILIATED TO APJ ABDUL KALAM TECHNOLOGICAL UNIVERSITY, THIRUVANANTHAPURAM)

SEVENTH SEMESTER B. TECH DEGREE EXAMINATION (S), FEBRUARY 2024
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. What is reinforcement learning in machine learning? List any two applications.
2. Differentiate between Maximum Likelihood estimation (MLE) and Maximum a Posteriori (MAP) estimation.
3. How does logistic regression classify data? Why is it called logistic regression if it is used for classification?
4. What is overfitting? How can we avoid overfitting?
5. What is an activation function and why do we need Non-linear activation function?
6. Differentiate between Hard Margin SVM and Soft Margin SVM.
7. What is the significance of expectation-maximization algorithm in machine learning?
8. Write the basic difference between the agglomerative and divisive hierarchical clustering algorithms and mention which type of hierarchical clustering algorithm is more commonly used.
9. List and explain the different ways to reduce high bias in Machine Learning.
10. How does AUC-ROC work? When should we use the AUC-ROC evaluation metric?

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

MODULE I

11. a) Write short notes on supervised learning, unsupervised learning and semi-supervised learning with examples.   (6)

     b) Let X be a continuous random variable with following PDF: (4)

P(X) =   2x   if 0≤x≤1

           0 otherwise

Suppose $P(Y/X) = x(1-x)^{y-1}$ for y=1, 2... Then find MAP estimate of X given y=3.

c) List any two real-life problems where clustering algorithms are preferred to get the results. Why do you think that Classification algorithms are not suitable for such problems?   (4)

OR

12. a) Illustrate with an example maximum likelihood method for predicting probabilities in Bayesian learning.   (7)

    b) 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.   (7)

MODULE II

13. a) Find which attribute can be chosen as the root for the decision tree classification, given dataset of 'Emotions' with features; crying and smiling. Justify your answer.

Sl. No	Crying	Smiling	Emotion
1	T	T	Happy
2	T	T	Happy
3	T	F	Sad
4	F	F	Happy
5	F	T	Sad
6	F	T	Sad

(8)

    b) How the performance of decision trees can be improved using pruning methods.   (6)

OR

14. a) Determine the least square regression line using matrix method and also find the sum of squared error for the following data.

Price(x)	Demand(y)
49	124
69	95
89	71
99	45
109	18

(8)

    b) Compare batch stochastic gradient descent and mini-batch stochastic gradient descent.   (6)

MODULE III

15. a) A neuron with 4 inputs has the weights 1; 2; 3; 4 and bias 0. The activation function is linear, say the function f(x) = 2x. If the inputs are 4; 8; 5; 6, compute the output. Draw a diagram representing the neural network. Also compute the output for the sigmoid activation function. (7)

    b) Explain back propagation algorithm and derive expressions for weight update relations with an example. (7)

OR

16. a) Demonstrate an algorithm to find the SVM classifier and describe the mathematical formulation of the SVM problem. (8)

    b) What is the purpose of kernel functions used in Support Vector Machines? Explain the common types of kernel functions used in SVM. (6)

MODULE IV

17. a) Suppose that we have the following data. Use single linkage Agglomerative clustering to identify the clusters and draw its dendrogram. (7)

Data Item	1	2	3	4	5
Values	(4,4)	(8,4)	(15,8)	(24,4)	(24,12)

    b) Cluster the following eight points (with (x, y) representing locations) into three clusters: A1(2, 10), A2(2, 5), A3(8, 4), A4(5, 8), A5(7, 5), A6(6, 4), A7(1, 2), A8(4, 9) using K-Means algorithm. (7)

OR

18. a) Explain the procedure for performing a Principal Component Analysis (PCA) on a given data set. (7)

    b) Given the following data, compute the principal component vectors and the first principal components (7)

x1	4	8	13	7
x2	11	4	5	14

MODULE V

19. a) 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. (10)

    b) In ROC space, which points correspond to perfect prediction - always (4)

positive prediction and always negative prediction? Why?

OR

20. a) Explain the basic approaches used for resampling in Cross Validation
         and Bootstrapping methods to improve the performance of the model.   (8)
     b) Differentiate between bagging, boosting and voting.   (6)

****