

Register No.: Name:

SAINTGITS COLLEGE OF ENGINEERING (AUTONOMOUS)

(AFFILIATED TO APJ ABDUL KALAM TECHNOLOGICAL UNIVERSITY, THIRUVANANTHAPURAM)

SIXTH SEMESTER B.TECH DEGREE EXAMINATION (R), MAY 2024**(2020 SCHEME)****Course Code: 20CST394****Course Name: Advanced Topics in Machine Learning****Max. Marks: 100****Duration: 3 Hours****PART A*****(Answer all questions. Each question carries 3 marks)***

1. Define the loss function with relevance?
2. Illustrate the types of machine learning algorithm with example?
3. What do you mean by version space and hypothesis space?
4. Illustrate the steps of k-means clustering algorithm?
5. Define Precision, Recall and F-measure.
6. How Bias-Variance Tradeoff affects machine learning algorithms?
7. How does the variational auto-encoder (VAE) architecture allow it to generate new data points, compared to auto-encoder, which cannot generate new data points?
8. Suppose we have a hypothesis set that labels all points inside an interval $[a, b]$ as class 1. Find its Vapnik-Chervonenkis(VC)- dimension?
9. Write a note on Gibbs Sampling?
10. Write a note on the types of GAN Network?

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

11. a) Explain the various types of Regression Techniques. (5)
b) Predict the auto insurance premium of a person having 11 years of driving experience using regression model generated from the following data (given in table below), where insurance premium is a function of driving experience. (9)

Driving Experience (years)	Insurance Premium (Rupees)
5	64
2	87
12	50
6	56
25	42
19	71

OR

12. a) Comment on the types and Importance of regularization techniques used in machine learning with difference. (7)
- b) Explain Gaussian Mixture Models. (7)

MODULE II

13. a) Cluster the following eight points (with (x, y) representing locations) into three clusters: (10)
- A1(2, 10), A2(2, 5), A3(8, 4), A4(5, 8), A5(7, 5), A6(6, 4), A7(1, 2), A8(4, 9)**
- b) Cite out the strength and weakness of k-means clustering algorithm? (4)

OR

14. a) Suppose that we have the following data:

P1	0.07	0.83
P2	0.85	0.14
P3	0.66	0.89
P4	0.49	0.64
P5	0.80	0.46

(10)

Use single linkage Agglomerative clustering to identify the clusters. Draw the Dendrogram. Use Euclidean distance measure.

- b) Illustrate the Concept of Expectation Maximization (EM) algorithm? (4)

MODULE III

15. a) Illustrate the working of Random Forest algorithm. Cite the advantage of Random Forest over the Decision Tree Algorithm? (7)
- b) A confusion matrix for 100 patients might look something like this. Determine the accuracy, recall, precision, f-measure and specificity of the same.

		Cancer	No cancer	(7)
Predicted values	Cancer	45	18	
	No Cancer	12	25	

OR

16. a) Illustrate the various Ensemble methods used in machine learning. (7)
- b) Explain the usage of ROC space and ROC curve in machine learning? In ROC space, plot the points corresponding to perfect prediction, always positive prediction and always negative prediction? Why? (7)

MODULE IV

17. a) Illustrate the concept of PAC Learning? (9)
- b) Explain about the true error with diagram? (5)

OR

18. a) Briefly explain the concept of PAC Learning Model? (8)
- b) Prove that $VC(H) \leq \log_2 |H|$, where H is a hypothesis space. ($|H|$ denotes the cardinality of the hypothesis space). (6)

MODULE V

19. a) Illustrate the types of Monte Carlo methods. (7)
- b) Explain rejection sampling with suitable example? (7)

OR

20. a) Explain Bayesian Belief Networks with application? (7)
- b) What is Effective Sample Size (ESS)? Why is a large ESS necessary but not sufficient for good MCMC mixing? (7)
