b.i.	Compare statistical modelling and machine learning.	5	2	1	1
ii.	Write the purpose of linear regression and gradient descent. Explain with an example.	5	4	1	1
27. a.i.	Brief about the need of regularization parameters in linear regression and ridge/lasso regression.	5	2	2	2
ii.	Define the terms concordance/ c-statistic. Calculate the concordance value for the given table.	5	2	3	2

Actual	1	0	0	1	1	0	1
Predicted	0.92	0.34	0.12	0.4	0.64	0.82	0.84

(OR)

b.	Explain in detail about logistic regression with an example.	10	4	2	2
28. a.	Consider the following likelihood table consists of single word and 3 words respectively for 100 mails. Apply Naive Bayes classification with conditional probability.				

Word frequency and likelihood of lottery with spam and ham

Lottery				Lottery			
Frequency	Yes	No	Total	Likelihood	Yes	No	Total
Spam	3	19	22	Spam	3/22	19/22	22
Ham	2	76	78	Ham	2/78	76/78	78
Total	5	95	100	Total	5/100	95/100	100

	Lottery (W1)		Million (W2)		Unsubscriable (W3)		
Likelihood	Yes	No	Yes	No	Yes	no	Total
Spam	3/22	19/22	11/22	11/22	13/22	9/22	22
Ham	2/78	76/78	15/78	63/78	21/78	57/78	78
Total	5/100	95/100	26/100	74/100	34/100	66/100	100

(OR)

b.	Write about KNN classification with an example. Elaborate curse of dimensionality in KNN.	10	4	3	3
29. a.	Explain the following with examples. (i) Maximum margin classifiers (ii) Support vector classifiers	10	4	4	4

(OR)

b.	Explain forward propagation and backward propagation algorithm in ANN with a neat example.	10	4	4	4
30. a.	Discuss in detail about k-means clustering and its working methodology with an example.	10	4	5	5

(OR)

b.	Explain in detail about principal component analysis with an example.	10	4	6	6

* * * * *

Reg. No.														

B.Tech. DEGREE EXAMINATION, MAY 2022
Sixth Semester

18CSE479T – STATISTICAL MACHINE LEARNING
(For the candidates admitted from the academic year 2018-2019 to 2019-2020)

Note:
(i) **Part - A** should be answered in OMR sheet within first 40 minutes and OMR sheet should be handed over to hall invigilator at the end of 40th minute.
(ii) **Part - B** should be answered in answer booklet.

Time: 2½ Hours Max. Marks: 75

PART – A (25 × 1 = 25 Marks)
Answer **ALL** Questions

		Marks	BL	CO	PO
1.	Consider Machine learning models. An iterative process takes place which are built based on various model parameters are called _____. (A) Mini – batches (B) Optimized parameters (C) Hyper parameters (D) Super parameters	1	2	1	1
2.	_____ is a type of gradient descent which processes one training example per iteration. (A) Batch gradient descent (B) Mini gradient descent (C) Stochastic batch gradient descent (D) 1-gradiant descent	1	1	1	1
3.	Consider two small sized samples and population, standard deviation is not given. How would you test the significance of the difference of the mean values between them? (A) f-test (B) Chi-square test (C) t-test (D) z-test	1	2	1	1
4.	False positive rate is (A) Specificity (B) Sensitivity (C) 1- Sensitivity (D) 1-specificty	1	1	1	1
5.	_____ cannot be achieved with linear model. (A) Flexibility fit complicated functions (B) Uncover complex multivariate relationship (C) Build accurate prediction models (D) Handling discrete values	1	2	1	1
6.	Identify the needed number of coefficients to estimate in a simple linear regression model (one independent variable) (A) 1 (B) 2 (C) 3 (D) 4	1	2	2	2
7.	Choose the option below which describes relationship of bias and variance with lambda. (A) For very small lambda, bias is low, variance is low (B) For very small lambda, bias is low, variance is high (C) For very small lambda, bias is high, variance is low (D) For very small lambda, bias is high, variance is high	1	2	2	2

8. In Random Forest or Gradient boosting algorithms, features can be of any type. For example, it can be continuous or categorical. Identify which of the following option is true by considering these types of features. 1 2 2 2
   (A) Only random forest – handlers real valued distributes by discretizing them
   (B) Only gradient boosting handles real valued attributes by discretizing them
   (C) Both can handle real valued attributes by discretizing them
   (D) Both cannot handle real valued attributes by discretizing them

9. For Lesso regression, if the regularization parameter is very high, choose the correct option 1 2 2 2
   (A) Can be used to select important features of a dataset
   (B) Shrinks the coefficient of less important features to exactly 0
   (C) The loss function is same as the ordinary least square loss function
   (D) The loss function is same as the ridge regression loss function

10. Choose the limitation of Lasso regression 1 1 2 2
   (A) If the number of features (P)> the number of observations, Lasso will pick at most n features as non-zero, even if all features are relevant
   (B) Lasso can be used to select important features of a data set
   (C) If there are two or more highly collinear features Lasso selects one of the randomly which is not good for interpretation
   (D) Lasso make coefficients to absolute 0

11. _____ is the purpose of the Laprace estimator in the content of Naïve Bayes classifiers. 1 1 3 3
   (A) To ensure the probabilities are not negative
   (B) To ensure that probability sum to one
   (C) To ensure non-zero probabilities
   (D) To ensure zero probability

12. Choose the wrong statement. 1 1 3 3
   (A) k-means clustering is a method of vector quantization
   (B) k-means clustering aims to partition 'n' observations into 'k' clusters
   (C) k-nearest neighbor is same as k-means
   (D) k-means clustering requires distance metric

13. Choose the correct statement about 'k' in KNN in terms of variance 1 1 3 3
   (A) Increase of k will increases variance
   (B) Decrease of k will increases variance
   (C) Increase of k will not change variance
   (D) Decrease of k will not change variance

14. In Naïve Bayes, suppose that prior for class W1 is greater than class W2, would the decision boundary shift 1 2 3 3
   (A) Towards region R1
   (B) Towards region R2
   (C) No shift in decision boundary
   (D) It depends on the exact value of priors

15. Bayes rule can be used for _____ 1 1 3 3
   (A) Solving queries
   (B) Increasing complexity
   (C) Decreasing complexity
   (D) Answering probabilistic query

16. _____ of the following is FALSE for neural networks 1 2 4 4
   (A) Artificial neurons are similar in operation to biological neurons
   (B) Training time for a neural network depends on network size
   (C) Neural networks can be simulated on conventional computer
   (D) The basic unit of neural networks are neurons

17. Neural networks are complex _____ with many parameters 1 1 4 4
   (A) Linear functions
   (B) Non linear functions
   (C) Discrete functions
   (D) Exponential functions

18. The output at each node is called 1 1 4 4
   (A) Node value
   (B) Weight
   (C) Neurons
   (D) Axons

19. Suppose you are using RBF Kernel in SVM with high Gamma value. It signifies the model would _____. 1 2 4 4
   (A) Consider even far away points from hyperplane for modelling
   (B) Consider only the points close to the hyperplane for modelling
   (C) Not be affected by distance of points from hyperplane for modelling
   (D) Be affected by distance of points close to the hyper plane for modeling

20. Find the option that more likely to consider iterating SVM next time 1 2 4 4
   (A) To increase data points
   (B) To decrease data points
   (C) To calculate more variables
   (D) To reduce the features

21. _____ is used for topic modelling 1 1 5 5
   (A) Random forest
   (B) Support vector machine
   (C) k-means
   (D) k-nearest neighbours

22. Identify the importance of using PCA before the clustering 1 2 6 6
   (A) Find the explained variance
   (B) Find good features to improve clustering
   (C) Find which dimension of data maximize the feature variance
   (D) Avoid bad features

23. PCA is used _____ 1 1 6 6
   (A) When the data is small and with few features
   (B) When there is an over fit case
   (C) To find latent features and reduce dimension latency
   (D) Every time before uses a ML algorithm

24. _____ of the following clustering algorithms suffers from the problem of convergence at local optima. 1 2 5 5
   (A) k-means and expectation – maximization
   (B) Agglomerative and k-means
   (C) Expectation-maximization and diverse
   (D) Diverse and k-means

25. Identify the clustering algorithm which is most sensitive to outliers. 1 2 5 5
   (A) k-means
   (B) k-medians
   (C) k-modes
   (D) k-medoids

PART – B (5 × 10 = 50 Marks) Marks BL CO PO

Answer **ALL** Questions

26. a.i. Discuss the broad categories of machine learning. 5 2 1 1

   ii. Write the steps to find variance. Consider the given data set and find variance. 5 2 1 1

Data set					
46	69	32	60	52	41

(OR)