R21_CS603C - MACIIINE LEARNING

Part A
Answer any ten from the following, choosing the correct alternative of each question: 10×1=10

1.	lu Linear Regression, the objective is to minimize which of the following?	(1)	CO 1	PO 1	BL 2
	(a) Classification error (b) Cross-entropy loss (c) Smn of squared residuals (d) Log loss				
2.	Association Rule Mining for Market Basket analysis is based on the concept of	(1)	CO 1	PO 1	BL 2
	(a) Supervised Learning (b) Unsupervised Learning (c) Reinforcement Learning (d) None of the above				
3.	Which of the following is a supervised learning algorithm?	(1)	CO 1	PO 1	BL 1
	(a) K-means clustering (b) Principal Component Analysis (c) Linear Regression (d) Apriori algorithm				
4.	Which algorithm is distance-based?	(1)	CO 1	PO 1	BL 1
	(a) Decision Tree (b) K-Nearest Neighbors (KNN) (c) Naive Bayes (d) Support Vector Machine (SVM)				
5.	Which of the following is NOT a linear model?	(1)	CO 1	PO 1	BL 2
	(a) Linear Regression (b) Logistic Regression (c) Decision Tree (d) Ridge Regression				
6.	The kernel trick is primarily used in which model?	(1)	CO 2	PO 2	BL 2
	(a) Decision Trees (b) Support Vector Machines (SVM) (c) K-Nearest Neighbors (KNN) (d) Naive Bayes				
7.	What is the main advantage of using a Decision Tree?	(1)	CO 3	PO 3	BL 3
	(a) It requires a lot of data preprocessing. (b) It can handle both numerical and categorical data easily. (c) It always gives the most accurate results compared to other models. (d) It is very complex and hard to interpret.				
8.	Which clustering method is sensitive to the initial choice of centroids?	(1)	CO 4	PO 4	BL 1
	(a) K-Means (b) DBSCAN (c) Hierarchical Clustering (d) Gaussian Mixture Model				
9.	What does the silhouette coefficient measure in clustering?	(1)	CO 3	PO 3	BL 3
	(a) The total number of clusters formed in the algorithm. (b) The consistency of cluster sizes. (c) The quality of the clustering by evaluating cohesion and separation. (d) The computational time complexity of the clustering algorithm.				
10.	Which method is used to cluster data in an unsupervised learning setting?	(1)	CO 1	PO 1	BL I
	(a) K-Means (b) Linear Regression (c) Support Vector Machine (d) Decision Tree				
11.	Which method is commonly used for matrix factorization?	(1)	CO 3	PO 3	BL 3
	(a) Singular Value Decomposition (SVO) (b) K-Means Clustering (c) Principal Component Analysis (PCA) (d) Naive Bayes Classifier				
12.	Which of the following is NOT a model selection criterion?	(1)	CO 1	PO 1	BL 1
	(a) AIC (Akaike Information Criterion) (b) BIC (Bayesian Infonnation Criterion) (c) R-squared (d) RMSE (Root Mean Square Error)				

Part B
(Answer any three of the following) 3x5=15

13.	Compare and contrast Decision Trees and Naive Bayes classifiers in terms of their assumptions and working principles.	(5)	CO 2	PO 2	BL 3
14.	What are Kernel Methods in Support Vector Machines? Explain with an example.	(5)	CO 2	PO 2	BL 4
15.	How can one apply ordinal along with one-hot encoding for a given temperature scale: {cold, warm, hot, very hot} ? Compute accuracy, precision, recall, F-measure, sensitivity and specificity in respect of following classification model's outcome.	(5)	CO 1	PO 1	BL 4

		Predicted Category	
		C_1 (+) Covid+	C_2 (−) Covid-
Actual Category	C_1 (+) Covid+	True Positive 85	False Negative 2
	C_2 (−) Covid-	False Positive 4	True Negative 9

16.	Compare and contrast Bagging and Boosting as ensemble learning techniques.	(5)	CO 2	PO 2	BL 4
17.	What is the significance of Matrix Factorization in machine learning? Discuss with an example.	(5)	CO 3	PO 3	BL 4

Part C
(Answer any three of the following) 3x15=45

18.	Discuss the challenges of modeling sequence data and explain how Recurrent Neural Networks (RNNs) handle them.	(15)	CO 3	PO 3	BL 3
19.	Compare Reinforcement Learning and Supervised Learning. How does reward maximization influence learning?	(15)	CO 3	PO 3	BL 4
20.	Explain Bayesian Learning and how it differs from frequent learning approaches.	(15)	CO 5	PO 4	BL 4
21.	Answer allthe Questions:				
	(a) Compare and contrast Naive Bayes and Decision Trees for classification tasks.	(8)	CO 2	PO 2	BL 3
	(b) Apply the concept of regression model for the following dataset to determine the glucose level of a person having age 55. Compute R-square value of the regression model.	(7)	CO 2	PO 2	BL 3

SUBJECT	AGE	GLUCOSE LEVEL
1	43	99
2	21	65
3	25	79
4	42	75
5	[illegible]	[illegible]

6	59	81
7	55	?

22. Discuss recent trends in machine learning, focusing on transformer models and their applications. **(15)** CO 4 PO 4 BL 2