CMR INSTITUTE OF TECHNOLOGY

USN | | | | | | | | | | |

Internal Assessment Test - II

Sub:	Machine Learning						Code:	22MBABA403
Date :	**08.10.2025**	Duration:	**90 minutes**	Max Marks:	Sem:	**IV**	Branch:	**MBA**

SET- I Solutions

Part A - Answer Any Two Full Questions (2* 20 = 40 marks)

		Marks	OBE CO	OBE RBT
1 (a)	**What is meant by *Voting Majority* in ensemble methods?** Voting Majority in ensemble methods means combining predictions from multiple models and selecting the class that receives the most votes from them — i.e., the majority decision determines the final output.	[03]	CO1	L1
(b)	**Explain the connect between GMM & E-M algorithm** Gaussian Mixture Model (GMM) assumes that data is generated from a mixture of several Gaussian distributions with unknown parameters. The Expectation-Maximization (E-M) algorithm is used to estimate these parameters (means, variances, and mixing coefficients). In the Expectation (E) step, it computes the probability that each data point belongs to each Gaussian component. In the Maximization (M) step, it updates the parameters to maximize the likelihood of the data. This process iterates until convergence, making EM the core algorithm for fitting GMMs.	[07]	CO3	L3
(c)	**Analyze the differences between Bagging & Boosting highlighting the applications.** • Concept – Bagging (Bootstrap Aggregating) builds multiple independent models in parallel, while Boosting builds models sequentially, each correcting the errors of the previous one. • Focus – Bagging reduces variance to prevent overfitting; Boosting reduces bias by improving weak learners. • Algorithm Type – Bagging uses random sampling with replacement; Boosting assigns weights to misclassified data to emphasize difficult cases. • Common Algorithms – Bagging is used in Random Forests; Boosting is used in AdaBoost, XGBoost, and Gradient Boosting. • Applications – Bagging suits unstable models like decision trees for classification; Boosting excels in high-accuracy tasks like fraud detection, churn prediction, and ranking.	[10]	CO3	L4
2 (a)	**What is Stacking?** Stacking (Stacked Generalization) is an ensemble technique that combines multiple base models and uses another model, called a meta-learner, to make the final prediction. It aims to leverage the strengths of different algorithms for improved overall accuracy.	[03]	CO1	L1

2 (b).	**Explain reinforcement learning with appropriate [7]** Reinforcement Learning (RL) is a type of machine learning where an agent learns by interacting with an environment and receiving rewards or penalties based on its actions. The goal is to learn an optimal policy that maximizes cumulative rewards over time. For example, a robot learning to walk adjusts its movements based on success or failure. Similarly, in games like chess, the agent learns strategies by receiving rewards for winning moves and penalties for losing ones.	CO3	L3
2 (c)	**Analyze the advantages and disadvantages of Ensembling in ML [10]** **Advantages:** 1. Improves prediction accuracy by combining multiple models. 2. Reduces overfitting and variance, leading to more stable results. 3. Works well with complex, non-linear datasets. **Disadvantages:** 4. Increases computational cost and training time. 5. Makes the model less interpretable and harder to explain compared to single models.	CO3	L4
3 (a)	**Differentiate ensemble and probabilistic learning [03]** Ensemble learning combines predictions from multiple models to improve overall performance. Probabilistic learning models uncertainty in predictions by estimating the probability distribution of outcomes rather than fixed outputs.	CO1	L2
(b)	**Discuss the business applications of implementing VR & AR [07]** Business applications of VR and AR: 1. Retail & E-commerce: Virtual try-ons and 3D product previews enhance customer experience. 2. Real Estate & Architecture: Virtual property tours and AR-based design visualization improve client engagement. 3. Training & Simulation: Used in healthcare, aviation, and manufacturing for safe, realistic training environments. 4. Marketing & Branding: Immersive AR/VR campaigns create interactive brand experiences and higher customer retention.	CO4	L2
(c)	**Analyze the working of a K-Dimensional (K-D) Tree with a suitable example, emphasizing its role in search efficiency. [10]** Concept: A K-Dimensional (K-D) Tree is a binary search tree used to organize points in a *k-dimensional space* for efficient searching. Structure: Each node represents a point and splits the space along one dimension (x, y, z, etc.) alternately at each level. Example: In a 2D space, the first split may be based on the x-coordinate, the next on the y-coordinate, and so on recursively. Search Efficiency: K-D Trees enable faster nearest neighbor and range searches by pruning large portions of the search space. Applications: Widely used in machine learning (KNN), computer graphics, and spatial databases for efficient data retrieval.	CO3	L3

Part B- Compulsory Case Study

CO3	L5

Read the following Case and answer appropriately [10]

A healthcare provider wants to predict whether patients are at risk of hospital readmission within 30 days. After evaluating multiple ML algorithms, **Logistic Regression** was identified as the most suitable model. It is widely applied for binary classification problems in healthcare data analytics.

Q. Evaluate the choice of Logistic Regression as the predictive model in this healthcare scenario. Discuss its suitability, assumptions, and potential limitations compared to alternate ML techniques.

Logistic Regression is a suitable choice for predicting 30-day hospital readmission as it effectively handles binary classification problems (e.g., "readmitted" vs. "not readmitted"). It provides interpretable coefficients, allowing healthcare professionals to understand how variables like age, comorbidities, or treatment type influence readmission risk.

However, it assumes a linear relationship between predictors and the log-odds of the outcome, and requires variables to be independent and free from multicollinearity. These assumptions may not hold in complex healthcare datasets with non-linear interactions.

Compared to advanced techniques like Random Forests or Gradient Boosting, Logistic Regression is simpler and more transparent but may yield lower predictive accuracy. Thus, it is ideal when explainability and interpretability are prioritized over model complexity.