Q.6	i.	Differentiate between bagging and boosting.	4	2	1	5	
	ii.	Describe the key components of a semi-supervised learning algorithm. How can it be used to improve the performance of a supervised learning model when labeled data is limited?	6	2	1	5	
OR	iii.	Explain the concept of deep learning? How does it differ from traditional machine learning? Discuss some of the challenges and limitations of deep learning models.	6	2	1	5	

******

Enrollment No......................................

Faculty of Engineering

End Sem Examination Dec 2024

IT3EA07 Machine Learning

Programme: B.Tech.   Branch/Specialisation: IT

Duration: 3 Hrs.   **Maximum Marks: 60**

Note: All questions are compulsory. Internal choices, if any, are indicated. Answers of Q.1 (MCQs) should be written in full instead of only a, b, c or d. Assume suitable data if necessary. Notations and symbols have their usual meaning.

			Marks	BL	PO	CO	PSO
Q.1	i.	Which statement is true about outliers in Linear regression? (a) Linear regression model is not sensitive to outliers (b) Linear regression model is sensitive to outliers (c) Can't say (d) None of these	1	1	1	1	
	ii.	Which of the following methods do we use to best fit the data in logistic regression? (a) Least square error (b) Maximum likelihood (c) Jaccard distance (d) Both (a) & (b)	1	1	1	1	
	iii.	_____ refers to a model that can neither model the training data nor generalize to new data. (a) Good fitting (b) Overfitting (c) Underfitting (d) All of these	1	1	1	1	
	iv.	Which one of the following applications is not an example of Naïve Bayes algorithm? (a) Spam filtering (b) Text classification (c) Stock market forecasting (d) Sentiment analysis	1	1	1,2	2	

v. Which of the following is a goal of clustering algorithms? **1**   1   1   3
(a) Classification
(b) Regression
(c) Dimensionality reduction
(d) Grouping similar data points together

vi. Which clustering algorithm is based on the concept of minimizing the sum of squared errors within each cluster? **1**   1   1   3
(a) K-Means   (b) DBSCAN
(c) Agglomerative   (d) Mean-Shift

vii. What is true regarding backpropagation rule? **1**   1   1   4
(a) It is also called generalized delta rule
(b) Error in output is propagated backwards only to determine weight updates
(c) there is no feedback of signal at any stage
(d) All of the these

viii. Which layer type is typically used to capture sequential dependencies in an RNN? **1**   1   1   4
(a) Input layer   (b) Hidden layer
(c) Output layer   (d) Activation layer

ix. Which of the following statements is not true about Random forests? **1**   1   1   2
(a) Scaling of data required in random forest algorithm
(b) It works well for a large range of data items than a single decision tree
(c) It has less variance than single decision tree
(d) Random forests are very flexible and possess very high accuracy

x. What is an epoch in deep learning? **1**   1   1   5
(a) The number of layers in a neural network
(b) The number of training examples in a dataset
(c) The number of times the entire dataset is passed through the neural network during training
(d) The number of neurons in a layer

Q.2   i. What is a cost function in machine learning? **2**   1   1   1

ii. Compare and contrast decision trees and random forests. **3**   2   1   2

iii. Describe the different types of neural networks. How are they used in various applications? **5**   2   2   4

OR iv. What is reinforcement learning? Explain its key components and how it is used to train agents to make decisions. **5**   2   2   4

Q.3   i. What is regularization in machine learning? How does it help in preventing overfitting? **4**   2   1,2   2

ii. Explain the role of kernel functions in support vector machines. Describe some common kernel functions and their applications. **6**   1   1   2

OR iii. Discuss the trade-off between bias and variance in machine learning models. How can you balance this trade-off to achieve optimal performance? **6**   2   1   2

Q.4   i. What is the goal of matrix factorization? **2**   1   1   3

ii. What is principal component analysis (PCA)? How does it work? Discuss its applications in data analysis. **8**   2   1,2   3

OR iii. Explain the concept of elbow method in K-means clustering algorithm in detail. Discuss its advantages and disadvantages. **8**   2   1   3

Q.5   i. Explain the concept of backpropagation in neural networks. **4**   1   1   4

ii. Discuss the limitations of traditional feedforward neural networks and how recurrent neural networks address those limitations. Give examples of applications. **6**   2   2   4

OR iii. Compare and contrast keras, tensorflow, and theano. Discuss their advantages and disadvantages in terms of ease of use, flexibility, and performance. **6**   2   1   4

Marking Scheme
IT3EA07 (T) Machine Learning (T)

Q.1   i)   b) Linear regression model is sensitive to outliers   **1**

ii)   a) Least Square Error   **1**

iii)   c) underfitting   **1**

iv)   c) Stock market forecasting   **1**

v)   d) Grouping similar data points together   **1**

vi)   a. K-Means   **1**

vii)   a) it is also called generalized delta rule   **1**

viii)   b) Hidden layer   **1**

ix)   a) Scaling of data required in random forest algorithm   **1**

x)   c) The number of times the entire dataset is passed through the neural network during training   **1**

Q.2   i.   What is a cost function in machine learning? (Definition 2)   **2**

ii.   Compare and contrast decision trees and random forests. (3 difference)   **3**

iii.   Describe the different types of neural networks. 2M
How are they used in various applications? 3M   **5**

OR   iv.   What is reinforcement learning? 2M
Explain its key components and how it is used to train agents to make decisions. 3M   **5**

Q.3   i.   What is regularization in machine learning? 2M
How does it help in preventing overfitting? 2M   **4**

ii.   Explain the role of kernel functions in Support Vector Machines. 3M
Describe some common kernel functions and their applications. 3M   **6**

OR   iii.   Discuss the trade-off between bias and variance in machine learning models. 4M
How can you balance this trade-off to achieve optimal performance? 2M   **6**

Q.4   i.   What is the goal of matrix factorization?   **2**

ii.   What is Principal Component Analysis (PCA)? 3M   **8**

How does it work? Discuss its applications in data analysis. 5M

OR   iii.   Explain the concept of elbow method in K-means clustering algorithm in detail. 5M
Discuss its advantages and disadvantages. 3M   **8**

Q.5   i.   Explain the concept of backpropagation in neural networks.   **4**

ii.   Discuss the limitations of traditional feedforward neural networks 4M
how recurrent neural networks address those limitations. Give examples of applications. 2M   **6**

OR   iii.   Compare and contrast Keras, TensorFlow, and Theano. 3M
Discuss their advantages and disadvantages in terms of ease of use, flexibility, and performance. 3M   **6**

Q.6

i.   Differentiate between Bagging and Boosting? (4 difference)   **4**

ii.   Describe the key components of a Semi-Supervised Learning algorithm. 3M How can it be used to improve the performance of a supervised learning model when labeled data is limited? 3M   **6**

OR   iii.   Explain the concept of deep learning? 2M
How does it differ from traditional machine learning? 2M Discuss some of the challenges and limitations of deep learning models. 2M   **6**

*****