RV UNIVERSITY
Go, change the world
an initiative of RV EDUCATIONAL INSTITUTIONS

School of Computer Science and Engineering

B.Tech (Hons.)
CIE-1 Question Paper
Academic Year 2024-2025

SECTION:	**USN**

Course: Introduction to Machine Learning	**Course Code: CS2213**	**Semester: IV**
Time: 9:00 AM to 10:00 AM	**Max Marks: 20**	**Date :12-02-2025**

Notes/ Instructions:

a) You must mark/circle the correct answers in the same sheet.

b) Answer all questions

c) Possession/ usage of Mobile Phones, Smart Watches or any other internet-enabled devices during the examination will be treated as malpractice.

Sl. No.	Questions	Marks	L1-L6	CO
1.	Select the learning approach of inductive learning. a) Proving a mathematical theorem using axioms. b) Predicting the weather based on historical data. c) Solving a puzzle using logical rules. d) Executing a pre-defined algorithm. **Answer: b) Predicting the weather based on historical data.**	**1**	**L2**	**CO1**
2.	A dataset has a sample x=[10,20,30,40,50]. Using Min-Max Normalization, Compute the normalized value of 30? A) 0.25 B) 0.5 C) 0.6 D) 0.75 **Answer: B) 0.5**	**1**	**L3**	**CO1**
3.	Which of the following is TRUE regarding covariance, correlation, and regression? A. Covariance is useful for predicting outcomes, correlation is not, and regression is solely for classification. B. Covariance indicates only the direction of the relationship, correlation provides both the strength and direction (in a standardized form), and regression estimates a model to predict one variable from another. C. Regression does not involve any linear relationships, while covariance and correlation do. D. All three measures provide identical information about the relationship between variables. **Answer: B). Covariance indicates only the direction of the relationship, correlation provides both the strength and direction (in a standardized form), and regression estimates a model to predict one variable from another.**	**1**	**L2**	**CO1**

L1- Remember, L2- Understand, L3 – Apply, L4- Analyse, L5 – Evaluate, L6 – Create

4.	Identify an example of nominal data. a. Student grades (A+, A, B) b. Temperature in Fahrenheit c. Heights of individuals d. Types of animals (cat, dog, bird) **Answer: d. Types of animals (cat, dog, bird)**	**1**	**L2**	**CO1**
5.	Consider a simple linear regression model with the following predictions and true values: Predictions = [2, 4, 6], True values = [1, 3, 5]. What is the Mean Absolute Error (MAE)? a) 0 b) 1 c) 2 d) 3 **Answer: b) 1**	**1**	**L3**	**CO2**
6.	Suppose you are building a fraud detection model using transaction data. Which of the following is considered as a sample? a. A single value in a feature vector b. A statistical measure of variability c. An example which can have multiple features d. The mean value of a feature **Answer: c. An example which can have multiple features**	**1**	**L2**	**CO1**
7.	Given the following six observations, find the covariance between height, weight variables. The height and weight of the six samples are in the same order (0, 0), (6.2, 75), (5.2, 62), (4.5, 60), (6.2, 70), (5, 78). **64.63** $$Cov(x, y) = \frac{\sum_{i=1}^{m}(x_i - \bar{x})(y_i - \bar{y})}{m-1}$$	**2**	**L3**	**CO1**
8.	You are developing a machine learning model to predict house prices using features like square footage, number of bedrooms, and location score. why is it necessary in this scenario? a. To increase model complexity b. To reduce the size of the dataset c. For quick model convergence d. To delete the missing values **Answer: c. For quick model convergence**	**1**	**L2**	**CO1**
9.	Compute the standard deviation of the observations 5, 5, 5, 5, and 5. a. 0 b. 1 c. 2 d. 5 **Answer: a) 0**	**1**	**L2**	**CO1**
10.	Given two points (4,4), (9,7). Consider p value in the Minkowski distance as 3. what is Minkowski distance **5.33**	**1**	**L3**	**CO2**
11.	Given a loss function and a learning rate α=0.1 if the current value of x=1 $f(x) = x^2 + 4x$ t value of x after one gradient descent step. **0.4**	**2**	**L3**	**CO2**
12.	Suppose, you have been given the following data where x1 and x2 are the 2 input variables and Class is the dependent variable. (-1,1, B), (0,1,A), (0,2,B), (1,-1,B), (1,0,A), (1,2,A), (2,2,B), (2,3,A). What will be the class of a new data point x1=1 and x2=1 in 5-NN (k-nearest neighbour with k=5) using Euclidean distance measure?	**2**	**L3**	**CO2**

	A. Class A B. Class B **Answer: A) Class A**			
13.	Consider a simple linear regression model with the cost function $J(\theta) = \frac{1}{m}\sum_{i=1}^{m}(h_\theta(x^{(i)}) - y^{(i)})$ $h_\theta(x) = \theta_0 + \theta_1 x_1$ θ_0, θ_1 , where If the initial values of are set to 0, and the learning rate is 0.1. $(x^{(1)}, y^{(1)}) = (1,2)$ and $(x^{(2)}, y^{(2)}) = (2,4)$ Find the updated values of θ_0, θ_1 Instead of using the given cost function, if students have used MSE as the cost function and obtained the below values, they will get one mark. **Theta_0: 0.3**, **Theta_1: 0.5 (One Mark)** The updated weights using the given cost function **Theta_0 = - 0.1** **Theta_1 = -0.15 (Two Marks)**	**2**	**L3**	**CO2**
14.	For the above question, find the initial value of J(θ) and updated value of J(θ) **10.0, 4.365 (One Marks)** **-3, -3.325 (Two Marks)**	**2**	**L3**	**CO2**
15.	'A' represents a series of numbers which is given as [10,20,30,40,50]. What is Inter Quartile Range (IQR) of A **30**	**1**	**L3**	**CO1**

Course Outcomes

1. Understand various types of machine learning algorithms and the role of data preprocessing in machine learning
2. Evaluate regression and classification model's performance on real-time datasets.
3. Apply unsupervised learning algorithms for pattern discovery and structural analysis in datasets
4. Build Multilayer Perceptron to perform classification
5. Perform image classification using Convolutional Neural Networks
6. Design and implement machine learning solutions to solve a real-world problem through a guided or open-ended project.

Marks Distribution											
L1	L2	L3	L4	L5	L6	CO1	CO2	CO3	CO4	CO5	CO6
0	6	18	0	0	0	10	10	0	0	0	0

L1- Remember, L2- Understand, L3 – Apply, L4- Analyse, L5 – Evaluate, L6 – Create

RV UNIVERSITY
Go, change the world
an initiative of RV EDUCATIONAL INSTITUTIONS

School of Computer Science and Engineering

B.Tech (Hons.)
CIE-1 Question Paper
Academic Year 2024-2025

SECTION:

USN

Course: Introduction to Machine Learning	**Course Code: CS2213**	**Semester: IV**
Time: 9:00 AM to 10:00 AM	**Max Marks: 20**	**Date :12-02-2025**

Sl. No.	Questions	Marks	L1-L6	CO
1.	Identify the learning approach that does not belong to inductive learning. A) Training a deep learning model on handwritten digits to recognise numbers. B) Using a dataset of customer transactions to predict future purchases. C) Applying a predefined mathematical formula to calculate interest on a loan. D) Training a reinforcement learning agent in a game environment. **Answer: C**	**1**	**L2**	**CO1**
2.	A dataset has a sample x= [10,5,16,25,40]. Using Min-Max Normalization, what is the normalized value of 25? A) 0.75 B) 0.57 C) 0.61 D) 0.71 **Answer: B) 0.57**	**1**	**L3**	**CO1**
3.	Select the tasks that can be addressed through classification techniques. A). Predict the gender of a person by analyzing his writing style. B). Predict the price of a house based on floor area, number of rooms, etc. C). Predict tomorrow's stock market opening price. D). Predict the number of copies of a book that will be sold this month. **Answer: A**	**1**	**L2**	**CO1**
4.	A company wants to automatically translate images from one domain to another domain, say from winter to summer. What is the input, and output of the machine learning model which you are going to develop? A) Input: Image Output: Pixel-wise class labels. B) Input: Image Output: Class label with bounding boxes. C) Input: Image Output: Class label. D) Input: Image Output: Image. **Answer: D) Input: Image Output: Image.**	**1**	**L2**	**CO1**
5.	You are working with a dataset containing features like annual income (in thousands). To ensure these features have a mean of 0 and a standard deviation of 1 before training a machine learning model, which preprocessing technique should you apply? A) Min-max normalization. B) Standardization. C) One-hot encoding. D) Feature selection. **Answer: B) Standardization.**	**1**	**L2**	**CO1**
6.	Given the categorical values ['Red', 'Blue', 'Green'], Generate the correct one-hot encoding representation.	**1**		

L1- Remember, L2- Understand, L3 – Apply, L4- Analyse, L5 – Evaluate, L6 – Create

	A) Red → [0], Blue → [1], Green → [2]. B) Red → [1,0,1], Blue → [1,1,1], Green → [0,0,0]. C) Red → [0, 0], Blue → [0, 1], Green → [1, 0]. D) Red → [1, 0, 0], Blue → [0, 1, 0], Green → [0, 0, 1]. **Answer: D**		**L3**	**CO1**
7.	What is the main difference between object recognition and object detection? A) Object recognition identifies objects and their locations, whereas object detection only identifies the category. B) Object detection identifies objects and their locations, whereas object recognition only identifies the category. C) Object recognition is a type of reinforcement learning, whereas object detection is not. D) Object recognition applies only to textual data, while object detection applies to images. **Answer: B**	**1**	**L2**	**CO1**
8.	If a dataset has Q1 = 50 and Q3 = 100, what is the upper bound for detecting outliers using the IQR method? Consider 1.5 as the constant which needs to be used while computing the upper bound. A) 150 B) 200 C) 175 D) 250 **Answer: C) 175**	**1**	**L3**	**CO1**
9.	A dataset consists of 12,000 samples with 50 features each. If 75% of the data is allocated for training and the rest for testing, how many samples are used for testing? A) 6000 B) 8000 C) 9000 D) 3000 **Answer: D) 3000**	**1**	**L2**	**CO1**
10.	Given the following six observations, find the covariance between height, and weight variables. The height and weight of the six samples are in the same order (0, 0), (6, 75), (5.1, 62), (4.1, 63), (6.1, 71), (5.1, 71) ___**62.98**_____ . A sample covariance is computed as follows $$Cov(x, y) = \frac{\sum_{i=1}^{m}(x_i - \bar{x})(y_i - \bar{y})}{m - 1}$$	**1**	**L3**	**CO1**
11.	Suppose, you have been given the following data where x1 and x2 are the 2 input variables and Class is the dependent variable. (-1,1,B), (0,1,A), (0,2,B), (1,-1,B), (1,0,A),(1,2,A), (2,2,B), (2,3,A). What will be the class of a new data point x1=1 and x2=1 in 5-NN (k-nearest neighbour with k=5) using Euclidean distance measure? A. Class A B. Class B **Answer: A) Class A**	**2**	**L3**	**CO2**
12.	Given the multiple linear regression equation: $y = \theta_0 + \theta_1 x_1 + \theta_2 x_2$ Which of the following statement is false? A) If θ_2 is negative, an increase in x_2 leads to decrease in y. B) The equation assumes a nonlinear relationship between y and x_1, x_2. C) The goal of training is to find the values of $\theta_0, \theta_1, \theta_2$ that minimizes the cost function. D) If θ_1 is negative, an increase in x_1 leads to decrease in y.	**1**	**L3**	**CO2**

	Answer: B			
13	If the actual values are [10, 15, 20] and the predicted values are [12, 14, 18], Compute the MAE, MSE, Root Mean Squared Error (RMSE). **MAE= 1.66**, **MSE = 3, RMSE = 1.73 or $\sqrt{3}$.**	**2**	**L3**	**CO2**
14	Consider the following dataset and the cost function is $$J(\theta) = \frac{1}{2m}\sum_{i=1}^{m}(h_\theta(x^{(i)}) - y^{(i)})^2$$ Assume initial values for $\theta_0,\theta_1,\theta_2$ as [-0.5, 1, 0] and $\boldsymbol{\alpha}$=0.01. The dataset has three samples and each sample in the dataset has two independent variables and one dependent variable. The samples are (0.8, 1, 1.5), (2.5, 2, 2.1), (3.1,3,3.5). Updated values for $\theta_0,\theta_1,\theta_2$. (Consider up to two decimal places) **-0.4926**, **1.0133, 0.0136**	**3**	**L3**	**CO2**
15	Based on question 14, find the J(θ) using updated learnable parameters. **0.3371**	**2**	**L3**	**CO2**

Course Outcomes

1. Understand various types of machine learning algorithms and the role of data preprocessing in machine learning
2. Evaluate regression and classification model's performance on real-time datasets.
3. Apply unsupervised learning algorithms for pattern discovery and structural analysis in datasets
4. Build Multilayer Perceptron to perform classification
5. Perform image classification using Convolutional Neural Networks
6. Design and implement machine learning solutions to solve a real-world problem through a guided or open-ended project.

Marks Distribution											
L1	L2	L3	L4	L5	L6	CO1	CO2	CO3	CO4	CO5	CO6
	6	14				10	10	0	0		

L1- Remember, L2- Understand, L3 – Apply, L4- Analyse, L5 – Evaluate, L6 – Create

SET-3

RV UNIVERSITY
Go, change the world
an initiative of RV EDUCATIONAL INSTITUTIONS

School of Computer Science and Engineering

B.Tech (Hons.)
CIE-1 Question Paper
Academic Year 2024-2025

SECTION:

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

Course: Introduction to Machine Learning	**Course Code: CS2213**	**Semester: IV**
Time: 9:00 AM to 10:00 AM	**Max Marks: 20**	**Date :12-02-2025**

Sl. No.	Questions	Marks	L1-L6	CO
1.	A company wants to predict whether a customer will buy a product or not based on their browsing history. Which type of supervised learning problem is this? A) Regression B) Clustering C) Classification D) Reinforcement Learning **Answer: C) Classification**	1	L2	CO1
2.	A dataset has a sample x = [5, 15, 25, 35, 45]. Using Min-Max Normalization, what is the normalized value of 25? A) 0.2 B) 0.4 C) 0.5 D) 0.6 **Answer: C) 0.5**	1	L2	CO1
3.	A k-NN classifier is used with 8,000 training samples. How many distance calculations are required for one test sample if k = 5? A) 5 B) 8,000 C) 40,000 D) Depends on k **Answer: B) 8000**	1	L2	CO2
4.	Identify the example of ordinal data. A) Blood types (A, B, AB, O) B) Weight of individuals in kilograms C) Speed of a car in km/h D) Employee position (junior, senior, manager) **Answer: D) Employee position (junior, senior, manager)**	1	L2	CO1
5.	Consider a simple linear regression model with the Predictions = [5, 10, 15], True values = [4, 8, 12], Compute the Mean Absolute Error (MAE)? A) 1 B) 2 C) 3 D) 4 **Answer: B) 2**	1	L3	CO1
6.	Which of the following is TRUE regarding covariance, correlation, and regression? A. Covariance is useful for predicting outcomes, correlation is not, and regression is solely for classification. B. Covariance indicates only the direction of the relationship, correlation provides both the strength and direction (in a standardized form), and regression estimates a model to predict one variable from another. C. Regression does not involve any linear relationships, while covariance and correlation do.	1	L3	CO1

L1- Remember, L2- Understand, L3 – Apply, L4- Analyse, L5 – Evaluate, L6 – Create

	D. All three measures provide identical information about the relationship between variables. **Answer: B**			
7.	The numbers 3, 7, 7, 3, and 5 form a dataset. Compute the standard deviation of this dataset (rounded to one decimal place)? A) 1.7 B) 3.1 C) 2.2 D) 1.4 **Answer: A) 1.7**	**1**	**L3**	**CO1**
8.	You are an archaeologist who discovers an ancient script that no one has ever seen before. There are thousands of symbols, but no translations exist. You decide to use machine learning to group similar symbols together based on their patterns and structures, hoping to uncover hidden relationships. Identify the learning paradigm should you choose. A) Supervised learning B) Unsupervised learning C) Reinforcement learning D) Guess the meanings randomly and hope for the best **Answer: B) Unsupervised learning**	**1**	**L2**	**CO1**
9.	You are building a machine-learning model that uses both height (in centimetres) and weight (in kilograms) as features. Without feature scaling, the model struggles with convergence. Which of the following best explains why feature scaling is important in this scenario? A) It makes models more complex B) It ensures that all features contribute equally to the model C) It removes the need for training data D) It increases the number of features **Answer: B) It ensures that all features contribute equally to the model**	**1**	**L2**	**CO1**
10.	Given two points (1,2) and (3,4), consider the p-value in the Minkowski distance as 2. What is the Minkowski distance between these points? (rounded to one decimal place) A) 2.8 B) 2.0 C) 3.0 D) 3.5 **Answer: A) 2.8**	**1**	**L3**	**CO2**
11.	You are given the following dataset with two features (X, Y) and their corresponding class labels as (1, 1, A), (4, 4, B) and (6, 7, B) A new point P(3,3) needs to be classified using K = 1 and the Euclidean distance method. Find the predicted class label for P(3,3). **Answer: Class B**	**2**	**L3**	**CO2**
12.	Given the following four observations, find the covariance between the height and weight variables. The height and weight of the four samples are: (2,4) , (4,6) , (6,8) , (8,10) A) 5.6 B) 4.6 C) 6.6 D) 8.6 **Answer: C) 6.6** **If the students have used 'n' in the denominator to compute the covariance instead of n-1, the answer will be: 5. If student mentions either option C or if he/she specifies answer 5, 2 marks will be given.**	**2**	**L3**	**CO1**
13.	Consider the following dataset representing students' test scores: [45 ,50 ,55 , 60 , 65 , 70 , 75 , 80 , 150 , 85]. Using a box plot, identify the outliers in the data. (Use the ± 1.5 × IQR rule to detect outliers). A) 150 only B) 45 and 150 C) 150 and 85 D) No outliers present	**2**	**L3**	**CO2**

	Answer: A) 150 only			
14.	Consider a **simple linear regression** model with the **cost function**: $$J(\theta_0, \theta_1) = \frac{1}{2m} \sum_{i=1}^{m} (h_\theta(x_i) - y_i)^2$$ where the **hypothesis function** is: $h_\theta(x) = \theta_0 + \theta_1 x$ Given the dataset: (1, 2), (2, 3), (3, 4). Assume the initial values of parameters $\theta_0 = -0.5$ and $\theta_1 = 1$, learning rate = 0.01. Find the updated values of θ_0, θ_1.___**-0.485**___, __**1.03**_____	**2**	**L3**	**CO2**
15.	For the above question, compute the updated cost function J(θ) _____**1.0156**___.	**2**	**L3**	**CO2**

Course Outcomes

1. Understand various types of machine learning algorithms and the role of data preprocessing in machine learning
2. Evaluate regression and classification model's performance on real-time datasets.
3. Apply unsupervised learning algorithms for pattern discovery and structural analysis in datasets
4. Build Multilayer Perceptron to perform classification
5. Perform image classification using Convolutional Neural Networks
6. Design and implement machine learning solutions to solve a real-world problem through a guided or open-ended project.

Marks Distribution											
L1	L2	L3	L4	L5	L6	CO1	CO2	CO3	CO4	CO5	CO6
0	6	14	0	0	0	10	10	0	0	0	0