



**K. K. Wagh Institute of Engineering Education & Research, Nashik**  
(An Autonomous Institute From A.Y. 2022-23)

WINTER-2025	
Exam Seat No.:	
Academic Year:2025-2026	Semester:V
Class:TY	Program:B.Tech
Branch Code:ROB	Pattern:2022
Name of Course:Data Analytics	Course Code: ROB223009
Max. Marks:60	Duration:2.30 Hrs.

**Instructions:** Candidates should read carefully the instructions printed on the Question Paper and on the cover page of the Answer Book, which is provided for their use.

1. This question paper contains 6 pages.
2. Answer to each new question is to be started on a new page.
3. Assume suitable data wherever required, but justify it.
4. Draw the neat labelled diagrams, wherever necessary.
5. The last columns indicates the Course Outcome and level of Blooms Taxonomy of the Question/sub-question.

**Marks CO**

**Question No. 1**

- 1a) Explain mean, median, mode, range, variance and standard Deviation with suitable example. (6) CO1

**Question No. 2**

- 2a) Explain correlation between two variables with suitable example (6) CO2

**Question No. 3**

- 3a) A study was conducted to understand consumer preferences for robotic drones based on specific attributes and levels: (8) CO3

Attributes	Level 1	Level 2	Level 3
Flight Time (minutes)	15 minutes	30 minutes	60 minutes
Camera Quality	720p HD	1080p Full HD	4K Ultra HD
Range (meters)	100 meters	500 meters	1000 meters
Foldability	Non-foldable	Foldable	Compact Foldable
Price (in USD)	\$200	\$400	\$800

After data analysis and regression modelling, the following utility equation was developed to predict consumer preference (U) based on these attributes:

$$U = 20 + (5 \times 30 \text{ minutes Flight Time}) + (8 \times 60 \text{ minutes Flight Time}) + (3 \times 1080p \text{ Full HD Camera Quality}) + (6 \times 4K \text{ Ultra HD Camera Quality}) + (4 \times 500 \text{ meters Range}) + (7 \times 1000 \text{ meters Range}) + (2 \times \text{Foldable}) + (5 \times \text{Compact Foldable}) - (0.4 \times \$400 \text{ Price}) - (0.8 \times \$800 \text{ Price})$$

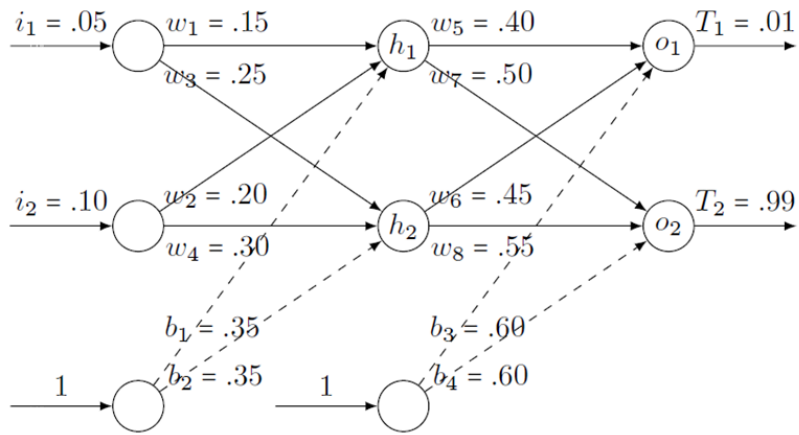
- i) Calculate the part-worth utility score for each level of every attribute.
- ii) Determine the importance of each attribute.
- iii) Calculate the predicted utility (U) for a robotic drone with the following attributes: Flight Time: 60 minutes, Camera Quality: 4K Ultra HD, Range: 1000 meters, Foldability: Compact Foldable, Price: \$400
- iv) How much more valuable (in dollars) is a robotic drone having flight time 60 minutes compared to a robotic drone having flight time 15 Minutes.

**OR**

- 3b) Suppose you are analyzing the performance of a manufacturing process based on two independent variables: the temperature (in degrees Celsius) and the time duration (in hours) of the process. You collected data from several runs of the process and recorded the following observations: (8) CO3

Process Run	Temperature (°C)	Time (hours)	Yield (%)
1	150	2	85
2	160	3	78
3	140	1.5	92
4	170	4	72
5	155	1.8	80

- i. Write the General equation for the multi linear regression model representing the relationship between yield (Y), temperature (X1), and time (X2).
  - ii. Use the given data to calculate the coefficients (intercept and slopes) of the regression model.
  - iii. If the temperature for a new process run is 165°C and the time is 2 hours, predict the yield using your regression model.
- 3c) Given the following network, compute the outputs from o1 and o2 (assume that the activation function is the sigmoid function). Also calculate mean squared error (MSE). (8) CO3



OR

3d) Explain K means Algorithm with suitable example. (8) CO3

**Question No. 4**

4a) Suppose a robot is using a camera to classify objects as either "plants" or "machines." Here are the results for 10 observations: (8) CO3

Observation	Expected	Predicted
1	Plant	Machine
2	Machine	Machine
3	Plant	Plant
4	Plant	Plant
5	Machine	Plant
6	Machine	Machine
7	Plant	Plant
8	Plant	Plant
9	Machine	Machine
10	Plant	Plant

1. Compute the confusion matrix for the data.
- ii. Compute the accuracy, precision, recall, sensitivity and specificity of the data.

OR

4b) Use a support vector machine to classify the following training set: (8) CO3

X1	X2	Class
1	1	+
1	-1	+
2	1	+
2	-1	+
4	0	-
5	1	-
5	-1	-
6	0	-

1. Determine the vector  $w$  and intercept  $b$ .
2. Plot Decision boundary with training points on an  $x_1$ - $x_2$  axis.
3. Find the class of point  $(0,1)$

- 4c) In a robotics assembly workshop, technicians use a logistic regression model to predict whether robotic arms will successfully complete a precision assembly task based on the calibration time they receive. The model specifies that the  $\log(\text{odds})$  of successful completion is given by the following equation: (8) CO3

$$\log(\text{odds}) = -50 + 1.8 \times (\text{Calibration Time in minutes})$$

- i. Calculate the probability of a successful assembly if the robotic arm is calibrated for 25 minutes.
- ii. Determine the minimum calibration time required for the robotic arm to achieve a probability of successful assembly greater than 85%.

**OR**

- 4d) In a robotic inspection system, engineers are evaluating different types of items based on their length and width measurements, which are represented in a two-dimensional feature space. The items are classified as follows: (8) CO3

Item Type A (Class 1):	(3, 4), (5, 6), (4, 5), (2, 4), (3, 3)
Item Type B (Class 2):	(8, 9), (7, 8), (9, 7), (6, 10), (10, 8)

Using linear discriminant analysis (LDA), compute the linear discriminant projection for these two-dimensional measurements of the items to determine the optimal boundary that distinguishes between Item Type A and Item Type B.

**Question No. 5**

- 5a) Consider the following set of training examples: (8) CO4

Instance	Classification	a1	a2
1	-	T	F
2	+	T	T
3	-	F	T
4	+	F	F
5	+	T	T
6	-	F	T

1. What is the entropy of this collection of training examples with respect to the target function “classification”?
2. What is the information gain of a1 and a2 relative to these training examples?

**OR**

- 5b) In a robotic vision system tasked with identifying obstacles in its path, evaluate the following performance metrics at various detection thresholds. Your objective is to construct the ROC curve and compute the Area Under the Curve (AUC). The performance metrics are summarized below: (8) CO4

Threshold	True Positives (TP)	True Negatives (TN)	False Positives (FP)	False Negatives (FN)
1	3	22	1	25
2	10	20	2	18
3	20	18	4	10
4	25	15	7	5
5	28	10	10	2
6	30	5	15	0
7	30	0	20	0

- 5c) You are tasked with maximizing the function  $g(x) = 3x - x^2$  over the range  $x \in [0, 3]$  using a Genetic Algorithm. (8) CO4

The initial population consists of 4-bit binary strings that represent values of  $x$  in the range  $[0, 3]$ . The initial population is: 1011, 0110, 1101, 0001.

The random numbers generated for selection are: 0.3, 0.55, 0.8, 0.95.

Perform the following steps:

1. Decode the binary strings to real numbers.
2. Evaluate the fitness of each decoded value using the function  $g(x) = 3x - x^2$ .
3. Use roulette wheel selection to select pairs of parents based on their fitness values.
4. Perform crossover at points 2 and 3 to generate new offspring.
5. Decode the new offspring and evaluate their fitness.
6. Identify the individual with the highest fitness and conclude the process.

**OR**

5d) Explain steps to solve PCA with suitable example.

(8) CO4

..... End of question paper.....