130

K. J. Somaiya Institute of Technology, Sion, Mumbai-22
(Autonomous College Affiliated to University of Mumbai)

Nov – Dec 2025
(B. Tech / M. Tech.) Program: B.Tech Scheme I/II/IIB/III:III
Regular Examination: TY Semester: V
Course Code: AIC 501 and Course Name: Machine Learning
Date of Exam: 24/11/25 Duration: 02.5 Hours Max. Marks: 60

Instructions:
(1)All questions are compulsory.
(2)Draw neat diagrams wherever applicable.
(3)Assume suitable data, if necessary.

Q. No.	Question	Marks	CO	BT level
Q 1	**Solve any two questions out of three: (05 marks each)**	**10**		
a)	Explain the concept of **cross-validation** and its importance in model evaluation.		CO1	Un
b)	What is the role of Entropy and Information Gain in decision trees?		CO3	Un
c)	Explain the Downhill Simplex (Nelder–Mead) method.		CO2	Un
Q 2	**Solve any two questions out of three: (05 marks each)**	**10**		
a)	Explain ROC and AUC curves in ML with examples.		CO4	Un
b)	Define clustering in Machine Learning. Explain how the number of clusters (k) affects the K-Means results.		CO5	Un
c)	Differentiate between Supervised and Unsupervised Learning.		CO1	Un
Q.3	**Solve any two questions out of three. (10 marks each)**	**20**		
a)	A hospital now wants to include **Age** as an additional predictor alongside **BMI** to determine the likelihood of Diabetes (Y = 1). The updated dataset for 4 patients is: (see table below)		CO2	Ap

Data Field	Observation 1	Observation 2	Observation 3	Observation 4
BMI (X1)	22	25	28	32
Age (X2)	45	50	35	40
Diabetes (0/1)	0	0	1	1

Assume:

- Initial parameters: $\theta_0 = 0$, $\theta_1 = 0$, $\theta_2 = 0$
- Learning rate: $\alpha = 0.1$
- Use logistic regression and gradient descent

Calculate the updated values of θ_0, θ_1, and θ_2 after **2 iteration** of gradient descent.

b) An engineer is testing the relationship between the time a polymer is allowed to cure and its final tensile strength. The data for **Curing Time** (in hours) and **Tensile Strength** (in Pascals) are shown below: — CO2 — Ap

Data Field	1	2	3	4	5	6
Curing Time (x in hours)	**1**	**1.5**	**2**	**2.5**	**3**	**3.5**
Tensile Strength (y in Pa)	**450**	**510**	**580**	**630**	**695**	**740**

1. **Find the least squares regression line,** $y = ax + b$, that models the relationship between Curing Time (x) and Tensile Strength (y).
2. **Use the resulting least squares regression line** as a predictive model to estimate the **Tensile Strength** of the polymer if the Curing Time were extended to **5.0 hours**.

c) — CO3 — Ap

Data Field	1	2	3	4	5	6	7	8	9	10
b1 (Budget)	**High**	**High**	**Low**	**Low**	**Medium**	**Medium**	**High**	**Medium**	**Low**	**High**
b2 (Companion)	**Family**	**Friends**	**Family**	**Solo**	**Friends**	**Family**	**Solo**	**Friends**	**Friends**	**Family**
b3 (Temp)	**Warm**	**Warm**	**Cold**	**Cold**	**Warm**	**Cold**	**Warm**	**Cold**	**Warm**	**Cold**
Booked or not	**Yes**	**Yes**	**No**	**No**	**Yes**	**No**	**Yes**	**No**	**No**	**No**

Build the decision tree for the above dataset using the C4.5 method.

Q.4 Solve any two questions out of three. (10 marks each) — **20**

a) Consider the following set of data points given, find the optimal hyperplane for these data points, — CO4 — Ap

(1,0.5), (1,1), (1, -0.5), (-0.5, 0.5), (0.5,0.5), (2,0), (4,0), (4.5, 1), (4.5, 0.5), (5, -1), (5.5, 0)

b) Given the dataset { a,b,c,d,e} and the following distance matrix. Construct a dendrogram by complete- linkage hierarchical clustering using the agglomerative method. — CO5 — Ap

$$D = \begin{pmatrix} & a & b & c & d & e \\ a & 0 & 9 & 3 & 6 & 11 \\ b & 9 & 0 & 7 & 5 & 10 \\ c & 3 & 7 & 0 & 9 & 2 \\ d & 6 & 5 & 9 & 0 & 8 \\ e & 11 & 10 & 2 & 8 & 0 \end{pmatrix}$$

c) Using the principal components, project the original data points onto the new subspace (i.e., calculate the scores for PC1 and PC2). — CO6 — Ap

Data Field	1	2	3	4	5	6	7	8	9	10
x1	**1.4**	**1.6**	**-1.4**	**-2**	**-3**	**2.4**	**1.5**	**2.3**	**-3.2**	**-4.1**
x2	**1.65**	**1.975**	**-1.775**	**-2.525**	**-3.95**	**3.075**	**2.025**	**2.75**	**-4.05**	**-4.85**

****