

Supplementary Jan/Feb ~~Nov-Dec 2025~~ Jan/Feb
 (B. Tech) Program: Computer Engineering Scheme III
 Regular Examination: TY Semester: V
 Course Code: CEC502 and Course Name: Machine Learning
 Date of Exam: ~~26/11/2025~~ 30/01/26 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	Max. Marks	CO	BT level																										
Q1	Solve any two questions out of three: (05 marks each)	10																												
	a) What is the bias-variance trade-off? How does it affect model performance?		1	U																										
	b) Using the following evidence, model a linear regression line which predicts the salary (y) of a person whose experience (x) is 15 years.		2	AP																										
	<table border="1" style="width: 100%; text-align: center;"> <tr> <td>x</td> <td>3</td> <td>7</td> <td>8</td> <td>12</td> <td>4</td> <td>6</td> <td>10</td> <td>20</td> </tr> <tr> <td>y</td> <td>30</td> <td>57</td> <td>64</td> <td>70</td> <td>35</td> <td>48</td> <td>70</td> <td>82</td> </tr> </table>		x	3	7	8	12	4	6	10	20	y	30	57	64	70	35	48	70	82	3	AP								
x	3	7	8	12	4	6	10	20																						
y	30	57	64	70	35	48	70	82																						
Q2	c) A machine learning model was evaluated on a dataset of 8 email instances	10																												
	<table border="1" style="width: 100%; text-align: center;"> <tr> <td>Actual (Y)</td> <td>Spam</td> <td>Ham</td> <td>Spam</td> <td>Ham</td> <td>Spam</td> <td>Ham</td> <td>Spam</td> <td>Ham</td> </tr> <tr> <td>Predicted (Ŷ)</td> <td>Spam</td> <td>Spam</td> <td>Ham</td> <td>Ham</td> <td>Spam</td> <td>Ham</td> <td>Spam</td> <td>Ham</td> </tr> </table>		Actual (Y)	Spam	Ham	Spam	Ham	Spam	Ham	Spam	Ham	Predicted (Ŷ)	Spam	Spam	Ham	Ham	Spam	Ham	Spam	Ham										
	Actual (Y)		Spam	Ham	Spam	Ham	Spam	Ham	Spam	Ham																				
Predicted (Ŷ)	Spam	Spam	Ham	Ham	Spam	Ham	Spam	Ham																						
I Construct the confusion matrix for this dataset. II Calculate the following performance metrics of each class i. Accuracy ii. Precision iii. Recall iv. F1Score III. Interpret the results, what do these values indicate about the model's performance																														
Q2	Solve any two questions out of three: (05 marks each)	10																												
	a) What are the different ways to combine classifiers in ensemble learning?		4	U																										
	b) Apply the K-Means clustering algorithm on the following 2-dimensional dataset. Use K-Means clustering with K = 2. Initial centroids are C ₁ = (2,3) and C ₂ = (9,3). Show Euclidean distance calculation, cluster assignment, new centroids calculation, reassignment, final clusters and final centroids.		5	AP																										
	<table border="1" style="width: 100%; text-align: center;"> <tr> <td>ID</td> <td>x1</td> <td>x2</td> <td>x3</td> <td>x4</td> <td>x5</td> <td>x6</td> <td>x7</td> <td>x8</td> </tr> <tr> <td>X</td> <td>2</td> <td>3</td> <td>5</td> <td>8</td> <td>9</td> <td>1</td> <td>6</td> <td>7</td> </tr> <tr> <td>Y</td> <td>3</td> <td>4</td> <td>7</td> <td>2</td> <td>3</td> <td>8</td> <td>6</td> <td>5</td> </tr> </table>	ID	x1	x2	x3	x4	x5	x6	x7	x8	X	2	3	5	8	9	1	6	7	Y	3	4	7	2	3	8	6	5		
ID	x1	x2	x3	x4	x5	x6	x7	x8																						
X	2	3	5	8	9	1	6	7																						
Y	3	4	7	2	3	8	6	5																						

Seat No.:

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

Jan/Feb ~~Nov~~ Dec 2025
 (B. Tech) Program: Computer Engineering Scheme III
Supplementary ~~Regular~~ Examination: TY Semester: V
30/01/26 Course Code: CEC502 and Course Name: Machine Learning
 Date of Exam: ~~26/11/2025~~ Duration: 02.5 Hours Max. Marks: 60

c)	Apply DBSCAN on the following data using $\epsilon = 1.9$ and $\text{MinPts} = 4$. Identify core, border and noise points and form final clusters. P1(3,7), P2(4,6), P3(5,5), P4(6,4), P5(7,3), P6(6,2), P7(7,2), P8(8,4), P9(3,3), P10(2,6), P11(3,5), P12(2,4)		5	AP																																																							
Q.3	Solve any two questions out of three. (10 marks each)	20																																																									
a)	i) Explain Machine Learning and describe Tasks, Performance and Experience with respect to a particular application (5M) ii) Compare supervised and unsupervised learning (5M)		1	U																																																							
b)	i) Define the Sigmoid function and explain its role in Logistic Regression. Draw the sigmoid curve and label its key characteristics(5M) ii) A company wants to examine whether the time spent on their website (in seconds) influences whether a user clicks on an advertisement. Use Logistic Regression on the given dataset <table border="1" data-bbox="177 1066 1246 1205"> <tr> <td>Time Spent (sec)</td> <td>68.95</td> <td>80.23</td> <td>69.45</td> <td>74.15</td> <td>50.00</td> <td>55.50</td> <td>80.00</td> <td>70.50</td> </tr> <tr> <td>Clicked on Ad</td> <td>No</td> <td>No</td> <td>No</td> <td>No</td> <td>Yes</td> <td>Yes</td> <td>No</td> <td>No</td> </tr> </table> a) Encode the response variable (Yes/No) into binary form. b) Compute the logistic regression equation. c) Estimate the probability of clicking for a user who spends 72 seconds on the website. d) If the decision threshold is 0.5, state the predicted class. (5M)	Time Spent (sec)	68.95	80.23	69.45	74.15	50.00	55.50	80.00	70.50	Clicked on Ad	No	No	No	No	Yes	Yes	No	No		2	AP																																					
Time Spent (sec)	68.95	80.23	69.45	74.15	50.00	55.50	80.00	70.50																																																			
Clicked on Ad	No	No	No	No	Yes	Yes	No	No																																																			
c)	Consider the following dataset <table border="1" data-bbox="177 1442 1254 1899"> <thead> <tr> <th>Age</th> <th>Income</th> <th>Owns House</th> <th>Loan Amount</th> <th>Loan Approved</th> </tr> </thead> <tbody> <tr><td>Young</td><td>High</td><td>No</td><td>Low</td><td>No</td></tr> <tr><td>Young</td><td>High</td><td>No</td><td>Medium</td><td>No</td></tr> <tr><td>Middle</td><td>High</td><td>No</td><td>Medium</td><td>Yes</td></tr> <tr><td>Senior</td><td>Medium</td><td>No</td><td>Medium</td><td>Yes</td></tr> <tr><td>Senior</td><td>Low</td><td>Yes</td><td>Medium</td><td>Yes</td></tr> <tr><td>Senior</td><td>Low</td><td>Yes</td><td>High</td><td>No</td></tr> <tr><td>Middle</td><td>Low</td><td>Yes</td><td>High</td><td>Yes</td></tr> <tr><td>Young</td><td>Medium</td><td>No</td><td>Low</td><td>No</td></tr> <tr><td>Young</td><td>Low</td><td>Yes</td><td>Low</td><td>Yes</td></tr> <tr><td>Senior</td><td>Medium</td><td>Yes</td><td>Medium</td><td>Yes</td></tr> </tbody> </table>	Age	Income	Owns House	Loan Amount	Loan Approved	Young	High	No	Low	No	Young	High	No	Medium	No	Middle	High	No	Medium	Yes	Senior	Medium	No	Medium	Yes	Senior	Low	Yes	Medium	Yes	Senior	Low	Yes	High	No	Middle	Low	Yes	High	Yes	Young	Medium	No	Low	No	Young	Low	Yes	Low	Yes	Senior	Medium	Yes	Medium	Yes		3	AP
Age	Income	Owns House	Loan Amount	Loan Approved																																																							
Young	High	No	Low	No																																																							
Young	High	No	Medium	No																																																							
Middle	High	No	Medium	Yes																																																							
Senior	Medium	No	Medium	Yes																																																							
Senior	Low	Yes	Medium	Yes																																																							
Senior	Low	Yes	High	No																																																							
Middle	Low	Yes	High	Yes																																																							
Young	Medium	No	Low	No																																																							
Young	Low	Yes	Low	Yes																																																							
Senior	Medium	Yes	Medium	Yes																																																							

Seat No.:

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

~~Jan/Feb~~ ~~Nov~~ ~~Dec~~ 2025
(B. Tech) Program: Computer Engineering Scheme III
Supplementary Regular Examination: TY Semester: V
Course Code: CEC502 and Course Name: Machine Learning
Date of Exam: ~~26/11/2025~~ Duration: 02.5 Hours Max. Marks: 60
30/01/26

	<p>I. Calculate the Gini Index of the parent dataset. II. Compute the Gini Index for splitting on each attribute (Age, Income, Owns House, and Loan Amount). III. Identify the best attribute to split the dataset at the root node. IV. Draw the first-level decision tree.</p>																					
Q.4	Solve any two questions out of three. (10 marks each)	20																				
	Differentiate between Linear SVM and Non-Linear SVM with suitable examples. Explain the Kernel Trick. Draw neat diagrams.		4	U																		
b)	<p>Given two different datasets of one-dimensional points Dataset A: 12, 15, 20, 28, 35, 40 Dataset B: 5, 8, 14, 18, 22, 30 (i) Apply agglomerative hierarchical clustering using Single Linkage on Dataset A. Show the proximity matrix at each merging step and draw the final dendrogram. (ii) Apply divisive hierarchical clustering on Dataset B, starting with a single cluster and splitting based on maximum dissimilarity. Show step-wise cluster splitting and draw the final dendrogram. (iii) Compare briefly how agglomerative and divisive clustering approaches differ in their working.</p>		5	AP																		
c)	<p>Given the following 3-dimensional dataset, perform dimensionality reduction from 3D to 2D using Principal Component Analysis (PCA).</p> <table border="1" data-bbox="151 1377 949 1635"> <thead> <tr> <th>X</th> <th>Y</th> <th>Z</th> </tr> </thead> <tbody> <tr> <td>2.0</td> <td>0.8</td> <td>1.2</td> </tr> <tr> <td>3.1</td> <td>1.1</td> <td>1.3</td> </tr> <tr> <td>2.9</td> <td>0.9</td> <td>1.5</td> </tr> <tr> <td>3.3</td> <td>1.2</td> <td>1.7</td> </tr> <tr> <td>2.7</td> <td>0.7</td> <td>1.4</td> </tr> </tbody> </table> <p>a) Calculate the covariance matrix of the normalized data. b) Determine the eigenvalues and corresponding eigenvectors. c) Select the two principal components based on maximum variance and form the projection matrix. d) Transform the original 3D data to reduced 2D feature space. e) Which features contribute more to variance, and why are they chosen as principal components?</p>	X	Y	Z	2.0	0.8	1.2	3.1	1.1	1.3	2.9	0.9	1.5	3.3	1.2	1.7	2.7	0.7	1.4		6	AP
X	Y	Z																				
2.0	0.8	1.2																				
3.1	1.1	1.3																				
2.9	0.9	1.5																				
3.3	1.2	1.7																				
2.7	0.7	1.4																				

Seat No.: