

Roll No	2	1	9	2	0	6	9
---------	---	---	---	---	---	---	---

End Semester Examination 2024

Name of the Course: BCA

Semester: 6th

Name of the Paper: Fundamentals of Machine Learning

Course Code: TBC-603

Time: 3 Hours

Maximum Marks: 100

Note:

- (i) All Questions are compulsory.
- (ii) Answer any two sub questions among a,b and c in each main question.
- (iii) Total marks in each main question are twenty.
- (iv) Each question carries 10 marks.

Q1	(10 X2 = 20 Marks)																						
(a)	What is machine learning? Explain its various approaches. What are the differences between machine learning and traditional programming?	CO1																					
(b)	Explain properties of matrix multiplication. Find out the inverse of the given matrix: $\begin{bmatrix} 1 & -1 & 2 \\ 4 & 0 & 6 \\ 0 & 1 & -1 \end{bmatrix}$																						
(c)	Compute mean, median, mode, range, average deviation, absolute deviation, squared deviation, standard deviation, total sum of squares for the following dataset: {8, 25, 20, 10, 8, 3} (ii) {18,22,33,11,9,4}																						
Q2	(10 X2 = 20 Marks)																						
(a)	What are the various mutable and immutable data types in python? Explain in detail.	CO2																					
(b)	Write any 15 built in functions in python with suitable examples.																						
(c)	Describe various operators used in python with their precedence.																						
Q3	(10 X2 = 20 Marks)																						
(a)	Consider the following dataset with five data points: (2,4), (3,6), (4,8), (7,3), (8,5). Perform k-means clustering with k=2. Initialize the cluster centroids randomly. Iterate until convergence and assign each data point to its closest centroid. Calculate the final cluster centroids.	CO3																					
(b)	Write short notes on: (i) Principle component analysis (ii) High correlation filter																						
(c)	Discuss hierarchical clustering. Find the clusters using single link technique. Use Euclidean distance and draw the dendrogram: <table border="1"><thead><tr><th></th><th>X</th><th>Y</th></tr></thead><tbody><tr><td>P1</td><td>0.40</td><td>0.53</td></tr><tr><td>P2</td><td>0.22</td><td>0.38</td></tr><tr><td>P3</td><td>0.35</td><td>0.32</td></tr><tr><td>P4</td><td>0.26</td><td>0.19</td></tr><tr><td>P5</td><td>0.08</td><td>0.41</td></tr><tr><td>P6</td><td>0.45</td><td>0.30</td></tr></tbody></table>			X	Y	P1	0.40	0.53	P2	0.22	0.38	P3	0.35	0.32	P4	0.26	0.19	P5	0.08	0.41	P6	0.45	0.30
	X	Y																					
P1	0.40	0.53																					
P2	0.22	0.38																					
P3	0.35	0.32																					
P4	0.26	0.19																					
P5	0.08	0.41																					
P6	0.45	0.30																					
Q4	(10 X2 = 20 Marks)																						
(a)	Write a program to load csv file (10 rows and 5 columns), then perform data normalization and print top 5 rows.	CO4																					
(b)	Write various methods to deal with missing data in our dataset.																						
(c)	What is Data Visualization? How it is helpful in decision making?																						
Q5	(10 X2 = 20 Marks)																						
(a)	Write short notes on: (i) K-fold cross validation (ii) Decision trees	CO5																					
(b)	Consider the following data set of experience and salary of five employees. Compute the y-intersect and slope of the best-fitting line for linear regression:																						

	<table border="1"><tr><th>Experience</th><th>Salary</th></tr><tr><td>10</td><td>21</td></tr><tr><td>14</td><td>33</td></tr><tr><td>12</td><td>27</td></tr><tr><td>10</td><td>22</td></tr><tr><td>8</td><td>23</td></tr></table>	Experience	Salary	10	21	14	33	12	27	10	22	8	23																																	
Experience	Salary																																													
10	21																																													
14	33																																													
12	27																																													
10	22																																													
8	23																																													
(c)	<p>Consider the following training dataset of 10 instances shown in the table, which describes the award performance of the individual students based on the GPA and no. of projects done. The target variable is “Award” which is a discrete valued variable that takes 2 values ‘Yes’ or ‘No’.</p> <table border="1"><tr><th>S.No.</th><th>GPA</th><th>No. of projects done</th><th>Award</th></tr><tr><td>1</td><td>9.5</td><td>5</td><td>Yes</td></tr><tr><td>2</td><td>8.0</td><td>4</td><td>Yes</td></tr><tr><td>3</td><td>7.2</td><td>1</td><td>No</td></tr><tr><td>4</td><td>6.5</td><td>5</td><td>Yes</td></tr><tr><td>5</td><td>9.5</td><td>4</td><td>Yes</td></tr><tr><td>6</td><td>3.2</td><td>1</td><td>No</td></tr><tr><td>7</td><td>6.6</td><td>1</td><td>No</td></tr><tr><td>8</td><td>5.4</td><td>1</td><td>No</td></tr><tr><td>9</td><td>8.9</td><td>3</td><td>Yes</td></tr><tr><td>10</td><td>7.2</td><td>4</td><td>Yes</td></tr></table> <p>Given a test instance (GPA=7.8, No. of projects done=4). Use the training set to classify the test instance. Use k-NN; Choose k=3.</p>		S.No.	GPA	No. of projects done	Award	1	9.5	5	Yes	2	8.0	4	Yes	3	7.2	1	No	4	6.5	5	Yes	5	9.5	4	Yes	6	3.2	1	No	7	6.6	1	No	8	5.4	1	No	9	8.9	3	Yes	10	7.2	4	Yes
S.No.	GPA	No. of projects done	Award																																											
1	9.5	5	Yes																																											
2	8.0	4	Yes																																											
3	7.2	1	No																																											
4	6.5	5	Yes																																											
5	9.5	4	Yes																																											
6	3.2	1	No																																											
7	6.6	1	No																																											
8	5.4	1	No																																											
9	8.9	3	Yes																																											
10	7.2	4	Yes																																											