Roll Number .....

Thapar Institute of Engineering and Technology, Patiala

Department of Computer Science and Engineering

WRITTEN TEST

B. E. (Third Year): Semester-V (ODD2021-22) (COE/CSE)	Course Code: UML501 Course Name: Machine Learning
October 27, 2021	Wednesday, 2.30PM
Time: 2 Hours, M. Marks: 45	Name Of Faculty: RAM, RKG, NKA, ABD

Note: **Attempt any five questions. Assume any missing data suitably. Scientific Calculator is allowed**

Q1. (A) With the following data observations explain each step in equi-depth binning. Further, with the same example explain smoothing by bin means and boundaries (5+4)

Input:[215, 11, 5, 10, 13, 35, 15, 50, 72, 55, 92, 204}

(B) Apply well-posed learning for playing checkers game.

Q2. A corporate bank, "ABC Corp." designed a Logistic Regression model for the approval of the business loan based on the historical dataset given in the following table. Check whether the loan will be approved or not for a person aged 39 and earning a salary of 50 lacs per annum. [Consider the model converges after two iterations and take learning rate=0.01 and initial coefficients as zeros] (9)

Sr. No.	Age	Salary (Lacs per annum)	Loan Approved
1	20	15	Yes
2	47	28	No
3	22	42	Yes
4	49	34	No
5	56	27	No
6	58	90	Yes
7	50	49	No

Q3. Apply z- score normalization on each entry for features A and B in the following dataset and calculate the corresponding value for every entry for both the features. Show the detailed calculations for the first entries of both the features. (9)

A	B
2546	23
20	10
30	20
8999	30
40	40
20	90
555	110

Q4. (A) How PCA is used for dimensionality reduction? Explain the steps in detail for reducing the 2-D dataset to the 1-D data set. (4+5)

(B) Find the principal components for the following two-dimensional dataset and percentage of the variance of each principal component analysis where the values of the features X1 and X2 are given as follows:

X1= [1.00 1.00 2.00 0.00 5.00 4.00 5.00 3.00]

Roll Number .....

X2= [3.00 2.00 3.00 3.00 4.00 5.00 5.00 4.00]

Q5. In the Multiple Linear Regression model, the error can be represented in matrix form as $\epsilon = Y - X\beta$. Using the Least square error method: (4+5)

(A) Derive the cost function equation for the multiple linear regression model.
(B) Derive the equation to compute optimal values of the β matrix for which the total square error is minimum.

Q6. A local clinic decided to maintain heath database of its patients. The small portion of the dataset is given as below: (9)

Sr. No.	chills	runny_nose	headache	fever	flu
1	Y	N	Mild	Y	**N**
2	Y	Y	No	N	**Y**
3	Y	N	Strong	Y	**Y**
4	N	Y	Mild	Y	**Y**
5	N	N	No	N	**N**
6	N	Y	Strong	Y	**Y**
7	N	Y	Strong	N	**N**
8	N	Y	Mild	Y	**Y**

The idea of collecting this dataset is to support doctors in starting early treatment of the patient based on Naïve Bayes Classifier if a patient visits the clinic with symptoms like chills=yes, runny_nose=no, headache=mild, fever=yes what would be the probabilities for each class and will this patient be classified as having flu or not.

Q7. Consider the given dataset and use the K-Nearest Neighbor (KNN), to compute the weight of the test instance given in the dataset (set k=3, distance=Euclidean). (9)

Sr. No.	Height	Age	Weight
1	5	45	77
2	5.11	26	47
3	5.6	30	55
4	5.9	34	59
5	4.8	40	72
6	5.8	36	60
7	5.3	19	40
8	5.8	28	60
9	5.5	23	45
10	5.6	32	58
Test Instance	5.5	38	**Weight?**