Machine Learning

Full Marks : 100

Time : Three hours

The figures in the margin indicate full marks for the questions.

*Answer **any five** questions.*

Q1. Consider the following dataset of *Old Car Sell* and develop a Linear Regression Model. Compute the R^2 value of your model. What will be the predicted price of an 8 years old car?

Car Age (Years)	Price (Rs in Lacs)
10	1
1	5
3	4
5	3
6	4

(12+5+3)

Q2. Consider the following dataset of three classes of flower. Apply K means clustering algorithm to identify their corresponding cluster. What will be the problem if you have a flower having petal length = 50 cm and petal width = 100 cm in the following dataset.

Petal Length(cm)	Petal Width (cm)
1	1
10	9
10	14
12	13
2	1
2	2
5	6
4	6
5	5
11	12

(16+4)

Q3.

(a) Consider the average number of COVID positive cases in your city was 5000/day. After the vaccination drive the number of cases over the last twelve days are as follows -

6000, 2000, 3000, 1500, 7000, 500, 1000, 500, 2000, 1000, 5000, 3000.

Based on the above mentioned data can we conclude that the vaccination drive has decreased the number of cases. Consider a 5% significance level of testing. Critical value t (11, 0.05) = 1.796.

(b) Consider the following execution time of five different program codes (*P1, P2, P3, P4, P5*) in three different machines (*M1, M2, M3*). Use ANOVA test to check whether the performance of these machines are statistically significant with a 5% significance level.

	M1	M2	M3
P1	5	3	6
P2	4	7	5
P3	8	2	10
P4	5	4	7
P5	9	3	8

F table

Critical values of F for the 0.05 significance level:

	1	2	3	4	5	6
1	161.45	199.50	215.71	224.58	230.16	233.99
2	18.51	19.00	19.16	19.25	19.30	19.33
3	10.13	9.55	9.28	9.12	9.01	8.94
4	7.71	6.94	6.59	6.39	6.26	6.16
5	6.61	5.79	5.41	5.19	5.05	4.95
6	5.99	5.14	4.76	4.53	4.39	4.28
7	5.59	4.74	4.35	4.12	3.97	3.87
8	5.32	4.46	4.07	3.84	3.69	3.58
9	5.12	4.26	3.86	3.63	3.48	3.37
10	4.97	4.10	3.71	3.48	3.33	3.22

(5+15)

Q4. Consider the following (Input: A, B, Output: X, Y) and design a suitable ANN. Clearly mention the total number of parameters and their corresponding values.

Input	Input	Output	Output
0	0	0	1
0	1	1	0
1	0	1	0
1	1	0	0

(20)

Q5. Consider the following dataset (P, Q are inputs and S is output) and design a SVM based model for binary classification.

P	Q	S
1	1	-1
-1	2	-1
1	-1	+1
-2	-2	+1

(20)

Q6. Write short notes on

    (a) Kernels of SVM
    (b) Overfitting and underfitting
    (c) Logistic regression
    (d) K nearest neighbour

(20)