Total No. of printed pages = 5

19/6th Sem/UCSE 603

2022

MACHINE LEARNING

Full Marks – 100

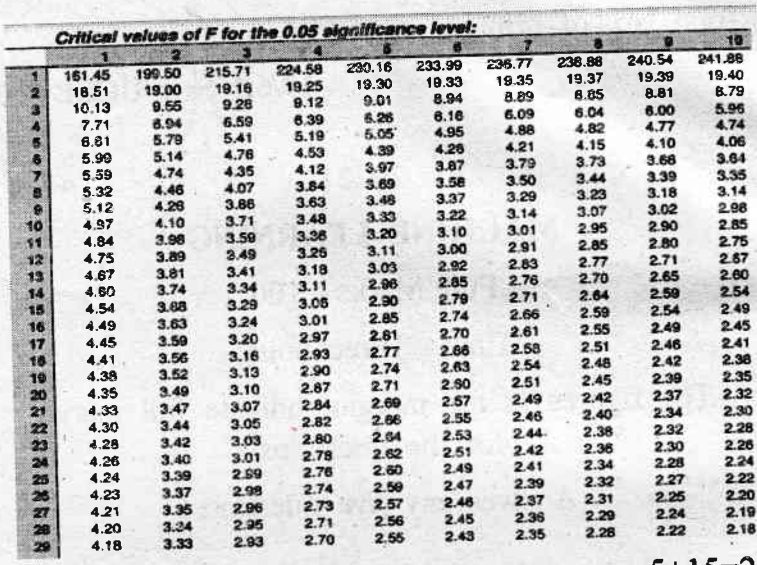
Time – Three hours

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

Answer any *five* questions.

1. Consider the following examination dataset of our five departments – CSE, ECE, FET, CE, and EIE. The results of five students are given below. The office of Dean Academics wants to tally the results. (a) Formulate null and alternative hypotheses. (b) Use one dimensional ANOVA to validate your null hypothesis with a significance level of $\alpha = 5\%$.

	CSE	ECE	FET	CE	EIE
Roll #1	90	80	70	80	60
Roll #2	50	90	60	40	80
Roll #3	90	70	90	30	70
Roll #4	80	50	80	90	40
Roll #5	60	40	20	100	50



Critical values of F for the 0.05 significance level:

	1	2	3	4	5	6	7	8	9	10
1	161.45	199.50	215.71	224.58	230.16	233.99	236.77	238.88	240.54	241.88
2	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.39	19.40
3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14
10	4.97	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98
11	4.84	3.98	3.59	3.36	3.20	3.10	3.01	2.95	2.90	2.85
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71	2.67
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65	2.60
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.49
17	4.45	3.59	3.20	2.97	2.81	2.70	2.61	2.55	2.49	2.45
18	4.41	3.56	3.16	2.93	2.77	2.66	2.58	2.51	2.46	2.41
19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42	2.38
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39	2.35
21	4.33	3.47	3.07	2.84	2.69	2.57	2.49	2.42	2.37	2.32
22	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34	2.30
23	4.28	3.42	3.03	2.80	2.64	2.53	2.44	2.38	2.32	2.28
24	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30	2.26
25	4.24	3.39	2.99	2.76	2.60	2.49	2.41	2.34	2.28	2.24
26	4.23	3.37	2.98	2.74	2.59	2.47	2.39	2.32	2.27	2.22
27	4.21	3.35	2.96	2.73	2.57	2.46	2.37	2.31	2.25	2.20
28	4.20	3.24	2.95	2.71	2.56	2.45	2.36	2.29	2.24	2.19
29	4.18	3.33	2.93	2.70	2.55	2.43	2.35	2.28	2.22	2.18

5+15=20

2. Consider the following database of some apartments. Construct a linear regression model. What will be the predicted cost of a 4 BHK flat (1500 sq. ft.) and 2 km from the main road ? 17+3=20

Area (*100 sq. ft.)	Distance from main Road (in km.)	Price (in Lakhs) (Rs.)
4	1	10
6	2	15
8	1	15
10	3	20
4	2	5
10	1	30

3. Consider the following three data points A (2, 0), B (1, 2) and C (– 1, – 1). A and B are positive samples, whereas C is negative. Design a Support Vector Machine to classify the given data. Compute the equations of two margins. Using your model predict what will be the class of a point (5, 9). 20

4. Consider the following database of the Airport Authority of India and construct a decision tree. In the following Table, high and low is identified as 0 and 1, respectively. Use your decision tree to compute the delay type when Visibility Source, Traffic Source, and Traffic Destination are all low (means 0) and Visibility Destination is high (1).

Visibility_ Source	Visibility_ Destination	Traffic_ Source	Traffic_ Destination	Delay Type
0	0	0	0	Medium
0	0	0	1	High
0	0	1	0	Medium
0	0	1	1	High
1	0	0	1	High
1	0	1	0	Medium
1	0	1	1	High
1	1	0	0	Medium

Visibility_ Source	Visibility_ Destination	Traffic_ Source	Traffic_ Destination	Delay Type
1	1	0	1	High
1	1	1	0	Medium
1	1	1	1	Low

20

5. (a) Consider we have the following ten data. Use K means algorithm for clustering. Assume K = 3, and the initial centroids for the three clusters are A3, A4, and A9, respectively. Use three iterations.

Data	A1	A2	A3	A4	A5
X Val	3	5	0	– 2	1
Y Val	3	10	3	– 5	2

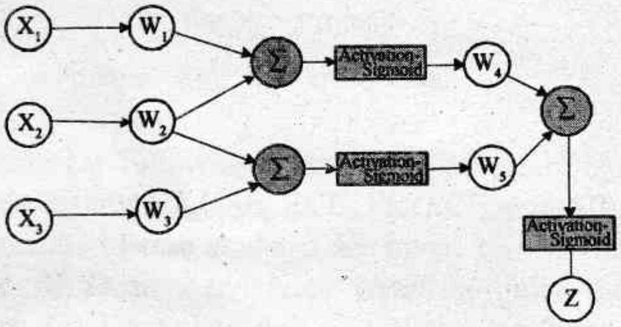
Data	A6	A7	A8	A9	A10
X Val	– 2	3	0	1	1
Y Val	3	4	4	3	5

(b) Consider an outlier data A11 (500, 1000). What will be the problems in your K means algorithm ? 15+5=20

6. (a) Consider the following ANN with backpropagation, where X1, X2, and X3 are three inputs, and Z is the output.

(i) Represent Z in terms of inputs.

(ii) During the training session if the original output is Y and the observed output is P, then how do the weights need to be modified ?



(b) Represent the functionality of a NAND gate using perceptron. 5+10+5=20

7. Write short notes on : 5×4=20

(a) Logistic Regression

(b) KNN

(c) Kernels in SVM

(d) Reinforcement Learning.