



Name:

Enrolment No:

UPES

End Semester Examination, May 2024

Course: Data analysis and visualization

Program: MCA

Course Code: CSDA7007P

Semester: II

Time : 03 hrs.

Max. Marks: 100

**SECTION A
(5Qx4M=20Marks)**

S. No.		Marks	CO
Q 1	Differentiate between supervised and unsupervised learning.	4	CO3
Q 2	Explain the concept of maximum margin classifier. Give one example.	4	CO3
Q 3	Explain the advantages and disadvantages of supervised learning.	4	CO3
Q 4	State the difference between parametric and non-parametric hypothesis testing.	4	CO2
Q 5	Define the following terms (1*4) i. Hyperplane ii. Support vector iii. Kernal iv. Clustering	4	CO2, CO3

**SECTION B
(4Qx10M= 40 Marks)**

Q 6	Define Machine Learning? Explain different types of learning strategy.	2+8=10	CO3
Q 7	Define kernel? Explain different types of kernels used in SVM with its mathematical definition. Write the advantages and disadvantages of SVM. (2+5+3)	10	CO3
Q 8	Write short note on: (Attempt any two) (2*5) (i) Z Test. (ii) T Test. (iii) K-Means algorithm (iv) Regression	10	CO2, CO3
Q 9	The population of all verbal GRE scores are known to have a standard deviation of 8.5. The UW Psychology department hopes to receive applicants with a verbal GRE scores over 210. This year, the mean verbal GRE scores for the 42 applicants was 212.79. Using a value of $\alpha = 0.05$ is this new mean significantly greater than the desired mean of 210? (critical value of $z_{0.05}$ is 1.64)	10	CO2, CO5

	Or																	
	From the data available, it is observed that 400 out of 850 customers purchased the groceries online. Can we say that most of the customers are moving towards online shopping even for groceries? (z critical values = (-1.96, +1.96))																	
SECTION-C (2Qx20M=40 Marks)																		
Q 10	<p>Write short note on (any four)</p> <ol style="list-style-type: none"> i. Applications of dimensionality reduction ii. Feature selection iii. Principal Component Analysis (PCA) iv. Applications of clustering v. Data pre-processing vi. Exploratory data analysis <p style="text-align: center;">Or</p> <p>Write the KNN algorithm. Would you use K-NN for large datasets? We have data from the questionnaire survey and objective testing with two attributes to classify whether a special tissue paper is good or not. Here is four training sample.</p> <table style="width: 100%; border-collapse: collapse;"> <thead> <tr> <th style="text-align: left;">X1 = Acid Durability (seconds)</th> <th style="text-align: left;">X2 = Strength (kg/square meter)</th> <th style="text-align: left;">Y = Classification</th> </tr> </thead> <tbody> <tr> <td>7</td> <td>7</td> <td>Bad</td> </tr> <tr> <td>7</td> <td>4</td> <td>Bad</td> </tr> <tr> <td>3</td> <td>4</td> <td>Good</td> </tr> <tr> <td>1</td> <td>4</td> <td>Good</td> </tr> </tbody> </table> <p>Now the factory produces a new paper tissue that passes laboratory test with X1=3, X2=7 without other expensive survey, can we guess what the classification of this new tissue is?</p>	X1 = Acid Durability (seconds)	X2 = Strength (kg/square meter)	Y = Classification	7	7	Bad	7	4	Bad	3	4	Good	1	4	Good	20	CO3, CO4, CO5
X1 = Acid Durability (seconds)	X2 = Strength (kg/square meter)	Y = Classification																
7	7	Bad																
7	4	Bad																
3	4	Good																
1	4	Good																

Q 11

- i. Three types of fertilizers are used on three groups of plants for 5 weeks. We want to check if there is a difference in the mean growth of each group. Using the data given below apply a one way ANOVA test at 0.05 significant level.

Fertilizer 1	Fertilizer 2	Fertilizer 3
6	8	13
8	12	9
4	9	11
5	11	8
3	6	7
4	8	12

Or

- ii. What is an ANOVA Test in Statistics? How to Perform an ANOVA Test? Explain the concept of one way ANOVA and two way ANOVA.

The following data represents the study hours/day by four students on 3 different days (Monday, Tuesday, wednesday)

Days	A	B	C	D
Monday	2	3	4	5
Tuesday	4	4	6	6
Wednesday	6	5	8	8

- a) Test whether the study hours of the different students are same?
b) Test whether the study hours on different days are same?
($F_{0.05 (6,3)}=4.76$ and $F_{0.05 (6,2)}= 5.14$)

20

CO2, CO5