

K. J. Somaiya Institute of Technology, Sion, Mumbai-22
(Autonomous College Affiliated to University of Mumbai)

May-June 2023

M.Tech. Artificial Intelligence

Examination: FY Semester: I Scheme- II

Course Code: **PCEC103** and Course Name: **Mathematical Foundations of Data Science**

Date of Exam: 28-06-2023

Duration: 2.5 Hours

Max. Marks: 60

Instructions:

- (1) All questions are compulsory.
- (2) Draw neat diagrams wherever applicable.
- (3) Assume suitable data, if necessary and state it clearly.

	Question	Max. Marks	CO	BT Level
Q1	Solve any SIX questions out of EIGHT .	12		
i)	What is p-value 0.05 in chi-square test?	2	CO5	1
ii)	What is meant by constrained optimization? What are the two types of constraints in constrained optimization?	2	CO6	1
iii)	How the exponential distribution is related to the Weibull distribution?	2	CO4	1
iv)	What is the difference between categorical and binary variables?	2	CO1	1
v)	How do you find correlation from covariance? Can covariance be greater than 1?	2	CO1	1
vi)	What is an estimate of where most of the data is located called? What are the measures of location and variability?	2	CO3	1
vii)	What is bootstrapping for confidence intervals and in sample size? Differentiate between bootstrapping and random sampling?	2	CO2	2
viii)	What is the null hypothesis for A or B testing? What is the null hypothesis of a two sample test?	2	CO6	2
Q2	Solve any FOUR questions out of SIX .	16		
i)	The rows of that matrix W produce three vectors (here written them as columns) $r_1 = \begin{bmatrix} 1 \\ 4 \\ 7 \end{bmatrix} \quad r_2 = \begin{bmatrix} 2 \\ 5 \\ 8 \end{bmatrix} \quad r_3 = \begin{bmatrix} 3 \\ 6 \\ 9 \end{bmatrix}$ Linear algebra says that these vectors must also lie in a plane. There must be many combinations with $y_1 r_1 + y_2 r_2 + y_3 r_3 = 0$. Find two sets of y's	4	CO2	2
ii)	What is Singular Value Decomposition? Give an example. How is SVD decomposition calculated? How does SVD facilitate dimensionality reduction in data analysis?	4	CO3	3

K. J. Somaiya Institute of Technology, Sion, Mumbai-22
(Autonomous College Affiliated to University of Mumbai)

iii)	How the exponential distribution is related to the Weibull distribution? How do you fit a Weibull distribution to data? What is the difference between exponential and Weibull distribution?	4	CO5	2																				
iv)	What is structured thinking for solving data science problems? Which method is used to solve data science problems?	4	CO1	2																				
v)	Suppose that the life of an industrial lamp, in thousands of hours, is exponentially distributed with failure rate $\lambda = 1/3$. Find the probability that the lamp will last a. Longer than its mean life of 3000 hours. b. Between 2000 and 3000 hours c. For another 1000 hours given that it is operating after 2500 hours.	4	CO6	3																				
vi)	Find the eigenvalues and eigenvectors of this symmetric 3 by 3 matrix S $S = \begin{bmatrix} 1 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 1 \end{bmatrix}$	4	CO4	2																				
Q3	Solve any TWO questions out of THREE .	16																						
i)	Let's consider an example confusion matrix for a Binary Classification problem: <table border="1" style="margin: 10px auto;"> <thead> <tr> <th></th> <th>Predicted Negative</th> <th>Predicted Positive</th> </tr> </thead> <tbody> <tr> <th>Actual Negative</th> <td style="text-align: center;">90</td> <td style="text-align: center;">10</td> </tr> <tr> <th>Actual Positive</th> <td style="text-align: center;">15</td> <td style="text-align: center;">85</td> </tr> </tbody> </table> <p>In a binary classification problem where "Actual Negative" and "Actual Positive" represent the True class labels, and "Predicted Negative" and "Predicted Positive" represent the Predicted class labels. Using this confusion matrix, compute the Precision, Recall, and overall Accuracy of the classifier.</p>		Predicted Negative	Predicted Positive	Actual Negative	90	10	Actual Positive	15	85	8	CO6	4											
	Predicted Negative	Predicted Positive																						
Actual Negative	90	10																						
Actual Positive	15	85																						
ii)	What is meant by student t- distribution? Why is it called student t-distribution? What are the 3 parameters of student t-distribution? How do you calculate student's t-distribution?	8	CO4	6																				
iii)	Fit a Poisson distribution to the following data, which gives the number of yeast cells per square for 400 squares. <table border="1" style="margin: 10px auto;"> <tbody> <tr> <td>No of cells per square (x)</td> <td>0</td> <td>1</td> <td>2</td> <td>3</td> <td>4</td> <td>5</td> <td>6</td> <td>7</td> <td>8</td> </tr> <tr> <td>No of squares (f)</td> <td>103</td> <td>143</td> <td>98</td> <td>42</td> <td>8</td> <td>4</td> <td>2</td> <td>0</td> <td>0</td> </tr> </tbody> </table>	No of cells per square (x)	0	1	2	3	4	5	6	7	8	No of squares (f)	103	143	98	42	8	4	2	0	0	8	CO5	6
No of cells per square (x)	0	1	2	3	4	5	6	7	8															
No of squares (f)	103	143	98	42	8	4	2	0	0															
Q4	Solve any TWO questions out of THREE .	16																						
i)	Why is binomial distribution used? What are the main features of binomial distribution? Why is binomial distribution a bell curve? What is the difference between Poisson distribution and binomial distribution?	8	CO2	3																				

K. J. Somaiya Institute of Technology, Sion, Mumbai-22
(Autonomous College Affiliated to University of Mumbai)

ii)	Why do you use the Chi-Square Test? What does a Chi-Square statistic test tell you? Who uses Chi-Square analysis? When to use a Chi-Square test? How do you calculate the Chi-squared? Give an example.	8	CO1	2
iii)	In an examination, it is laid down that a student passes if he secures 30% or more marks. He is placed in first, second and third division according as he secures 60% or more marks, between 45% and 60% marks and marks between 30% and 45% respectively. If he secures 80% or more marks, he gets distinction. It is noticed from the results that 10% of the students failed and 5% of them obtained distinction. Assuming normal distribution of marks, what percentage of students placed in the second division?	8	CO3	6
