MANIPAL INSTITUTE OF TECHNOLOGY

(A constituent unit of MAHE, Manipal)

II SEMESTER M.TECH.(DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING) END SEMESTER EXAMINATIONS, MAY/JUNE 2023

SUBJECT: ALGORITHMIC FOUNDATIONS OF DATA SCIENCE (CSE - 5002)

REVISED CREDIT SYSTEM

(31/05/2023)

Time: 9:30 am to 12:30 pm

MAX.MARKS: 50

INSTRUCTIONS TO CANDIDATES:-

- Answer **ALL** the questions.
- Missing data may be suitable assumed.

Q.n		Marks	CO	Level
0				
1A.	Prove the theorem that nearly all the probability is concentrated in a thin annulus of width O(1) at radius \sqrt{d} .	3M	C01	4
1B.	Compute the singular value decomposition of the matrix in polynomial time.	4M	C01	6
1C.	Design a Markov chain whose stationary distribution is a given target distribution.	3M	CO2	4
2A.	Discuss the following questions -	3M	CO2	2
	a) What is the expected time for a random walk starting at a vertex x to reach a target vertex y?			
	b) What is the expected time until the random walk returns to the vertex it started initially?			
	What is the expected time to reach every vertex?			
2B.	Show the algorithm to count the number of occurrences of a given element in a stream? Describe the motivations for sketch of a large matrix.	4M	CO3	3
2C.	Illustrate with an example Online learning.	3M	CO2	3
3A.	Describe the concept of Stochastic Gradient Descent with an example.	5M	CO2	2
3B.	Write and explain the algorithm to prove that spectral clustering gives a clustering close to the true clustering.	3М	CO3	3
3C.	Prove the lemma that the sum of squared distances	2M	CO3	4

	of the a_i to any point x equals the sum of the squared distances to the centroid of the a_i plus n times the squared distance from x to the centroid. Where a_i be the set of points = $\{a_1, a_2,, a_n\}$.			
4A.	What is the expected number of triangles in a graph G, where G(n,d/n), where d is a constant and n is the number of vertices?	5M	CO4	2
4B.	Write the rules for message passing in a general graph.	3M	CO4	3
4C.	Prove that, the threshold for the existence of cycles in graph G, where $G(n,p)$ is $p = 1/n$, where n is the number of vertices, and p is the edge probability.	2M	CO4	4
5A.	Given an Hidden Markov model, how probable is an output sequence of length T+1. Illustrate with the help of an algorithm and an example.	4M	CO5	4
5B.	Describe the concept of topic model with an example.	4M	CO5	2
5C.	Write the boosting algorithm with respect to Strong Learning.	2M	CO2	3