JAYPEE UNIVERSITY OF INFORMATION TECHNOLOGY, WAKNAGHAT
TEST-3 EXAMINATION-2023
B. Tech-VI Semester (CSE/IT)

COURSE CODE (CREDITS): 19BIWCI637 (2) **MAX. MARKS:** 35

COURSE NAME: Statistics and Exploratory Data Analytics

COURSE INSTRUCTOR: Amol Vasudeva **MAX. TIME: 2 Hours**

Note: All questions are compulsory. Marks are indicated against each question in square brackets.

1. A hierarchical cluster is given as ((((D, F), E), C), (A, B)). The steps involved in the formation of hierarchical cluster are given as below. Draw a Dendogram for this hierarchy based on the following information. (CO-4) **[5 marks]**
   - In the beginning, there are six clusters: A, B, C, D, E, and F.
   - The clusters D and F have been merged into a cluster (D, F) at a distance 0.50
   - The clusters A and B have been merged into a cluster (A, B) at a distance 0.71
   - The cluster E and (D, F) have been merged into a cluster ((D, F), E) at a distance 1.00
   - The clusters ((D, F), E) and cluster C have been merged into a cluster (((D, F), E), C) at a distance 1.41
   - The clusters (((D, F), E), C) and cluster (A, B) have been merged into a cluster ((((D, F), E), C), (A, B)) at a distance 2.50

2. Is there any mechanism to find the number of clusters in K-Mean clustering? Name and explain this mechanism briefly. (CO-4) **[2 marks]**

3. Based on the following dataset, write the steps to use K-Mean clustering algorithm (mathematically) to allocate each of the six objects into two groups (i.e. K=2). (CO-4) **[8 marks]**

Object	Attribute 1 (X): weight index	Attribute 2 (Y): pH
Medicine A	1	1
Medicine B	2	1
Medicine C	4	3
Medicine D	5	4

4. You are given an image file as "image.png". Write a Python program and use Pillow - Python Image Library (PIL) to perform the following: (CO-3) **[3+3+3 = 9 marks]**
   - Read the image, convert into grayscale, and then display it.
   - Compute the singular value decomposition (SVD)
   - Display the images by selecting first 25, 30, 35, 40, 45, and 50 components, respectively.

5. There are groups of students belonging either to gifted class or regular class. The students who had high scores in the last year's exam were assigned to gifted class. The rest of the students were allocated to the regular class. The details of the students' marks in the last year are given in a file *student_marks.csv* with the following two fields. (CO-4)

Score (out of 100)	Number (of students)
40	10
45	14
....	
100	12

Write a Python program to draw the histogram (number vs. score).
Write a Python program to split the histogram into two equal portions. **[5 marks]**

6. Differentiate between the following terms. (CO-3, 4)
   - Hard Clustering and Soft Clustering
   - PCA and SVD
   - K-Mean Clustering and Gaussian Mixture Model Clustering **[2×3=6 marks]**