Name: _____ Reg #: _____ Section: _____

National University of Computer and Emerging Sciences, Lahore Campus

Course: Bioinformatics Course Code: CS4054

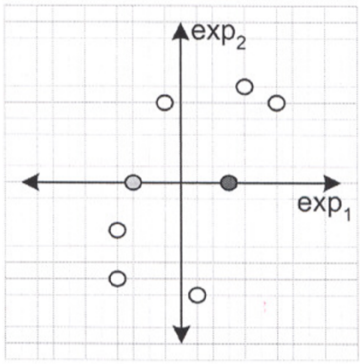
Program: BS(Computer Science) Semester: Spring 2025 **Section: BCS-8B**

Duration: 30 Minutes **Total Marks: 10**
Paper Date: 23-April-2025 **Page(s): 1**

Exam: Quiz 3

Q1 Consider the following gene expression values from two microarray experiments for 8 genes. Negative expression in the table means that the gene had downregulated, while positive expression means that the gene as upregulated. A biologist is trying to find out whether these 8 genes can be separated into two groups based on their behavior in the experimental conditions. In order to visualize the relationships, she has sketched a two-dimensional plot.

	exp_1	exp_2
$gene_1$	-4	-3
$gene_2$	6	5
$gene_3$	1	-7
$gene_4$	-4	-6
$gene_5$	4	6
$gene_6$	-1	5
$gene_7$	-3	0
$gene_8$	3	0



Use k-means clustering to divide these 8 genes into two clusters. Use gene 7 as the initial centroid of Cluster 1 and gene 8 as the initial centroid of Cluster 2. Use Euclidean distance for your calculations. Indicate the data points belonging to each cluster and give the coordinates of centroids at each iteration. Iterate until convergence. [10 marks]