BRAC University
Department of Computer Science and Engineering

CSE 443: Bioinformatics-I (A)

Quiz 02: Summer 2025 Time: 25 Minutes Marks: 15

Name		ID		Section	

1. Write two disadvantages of DNA micro-array technique compared to the modern RNA-Seq techniques. [3]

2. You are clustering gene expression data from 50 genes across 10 patients using K-means. After 10 iterations, the algorithm converges. However, you found that there are clusters containing genes that were known to belong to distinct pathways. Provide two potential reasons why K-means might have grouped biologically distinct genes together. How can these problems be handled? [6]

3. Soft-Kmeans algorithm uses fuzzy or soft membership and weighted update of the centroids according to the following two formula.

membership	$w_{ik} = \frac{e^{-\beta\|x_i-\mu_k\|^2}}{\sum_{j=1}^{K} e^{-\beta\|x_i-\mu_j\|^2}}$
centroid update	$\mu_k = \frac{\sum_{i=1}^{N} w_{ik}x_i}{\sum_{i=1}^{N} w_{ik}}$

The parameter β controls how sharply data points are assigned to clusters.

(a) What happens to the cluster membership probabilities as $\beta \to 0$? [2]

(b) What happens as $\beta \to \infty$ [2]

(c) Which scenario would be more suitable for overlapping clusters in gene expression data, and why? [2]