Name: _____ Reg #: _____ Section: _____

National University of Computer and Emerging Sciences, Lahore Campus

Course: Bioinformatics Course Code: CS4049

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

Duration: 15 Minutes   **Total Marks: 10**

Paper Date: 11-Feb-2025   **Page(s): 2**

<u>**Exam: Quiz 1**</u>

Q1 Calculate the following: **(6)**

(a) Frequency of the 4-mer CATG in ACGTTGCATGTCGCATGATGCATGAGAGCT **(1)**

(b) Calculate the Hamming Distance between GGGCCGTTGGT and GGACCGTTGAC **(1)**

(c) Given strings Text and Pattern as well as an integer d, we define $Count_d$(Text, Pattern) as the total number of occurrences of Pattern in Text with at most d mismatches.

Calculate $Count_2$(CGATCGAGTACCATAAG , ATA) and list down the found patterns. **(4)**

Q2 For the given Profile (Motif)   **(4)**

	A:	0.2	0.2	0.3	0.2	0.3
Profile	C:	0.4	0.3	0.1	0.5	0.1
	G:	0.3	0.3	0.5	0.2	0.4
	T:	0.1	0.2	0.1	0.1	0.2

(a) Write the formula and calculate the probability of the 5-mer CCGAG   **(2)**

(b) Write the formula and calculate the Entropy of each column of the profile matrix   **(2)**