End Semester Examination, May 2019
Bioinformatics (LS4205)

Total Marks: 50 (PART-A: Marks: 30 + PART-B: Marks: 20)

(PART-A: Total Marks: 30)

1(a). What is the significance of bioinformatics and its contribution to biology? Briefly explain its applications in genomics and molecular biology research?

1(b). What are scoring matrices? Write short note on PAM and BLOSUM scoring matrices.
[Marks 3+2]

2(a). What is database? Briefly explain different types of biological databases.

2(b). Write a short note on Protein Data Bank (PDB)?
[Marks 3+2]

3(a). Consider the below examples. What is a better match/alignment? Briefly explain why?

```
(a)     LQWSH
        :  ::
        LD-SH

(b)     LQWSH
        :  ::
        L-DSH
```

3(b). Briefly describe BLAST and its variants. Explain the statistical significance of *E*-value and bit-score in a BLSAT output. [Marks 2+3]

4(a). Discuss the difference between local and global alignment with suitable examples.

4(b). Find the best global alignment (Needleman-Wunsch algorithm) between below sequences using the dynamic programming approach, where the scoring scheme is +2 for match, -1 for mismatch and -2 for a gap penalty.

```
Seq1    GATTA
Seq2  GAATTC
```
[Marks 2+3]

5. What is the rationale behind homology modeling? Discuss the various steps involved in the same. [Marks 5]

6. What is molecular docking explain different types of docking methods and their applications?

[Marks 5]

End Semester Examination, May 2019
Bioinformatics (LS4205)
(Part-B: Total Marks: 20)

Student Name:
Roll No:

Circle the most appropriate choice for the questions 1-3

1. There is an effort to understand the molecular underpinning of a disease affecting adult humans but not children. Comprehensive comparison of the genome sequences of affected subjects with healthy unaffected subjects has not revealed any difference between the two groups so far. One suspects that some genes that are normally switched "off" (i. e., not transcribed) in the adults are switched "on" during pathogenesis. Which of the following statements is most consistent with the information provided? [1 mark]
   a. The disease is caused by germ-line mutation.
   b. Analysis of transcriptome is likely to be informative.
   c. The disease is caused most likely by a metabolic disorder.
   d. One should plan an experiment to study the life-events of the subjects.

2. To measure the level of gene activity, one should measure the abundance of: [1 mark]
   a. mRNA
   b. rRNA
   c. tRNA
   d. None of the above

3. Genome-scale data suffer from "curse of dimensionality". Which of the following method can be used to project the high-dimensional data into a 3-dimensional space, where each point represents a single sample? [1 mark]
   a. Student's t-test
   b. Hierarchical clustering
   c. Principal component analysis
   d. Needleman-Wunsch algorithm

4. Draw appropriate lines from the terms on the left to the terms on the right. [1 mark]

Microarray	Detects low abundance transcript
RNA-seq	Earlier technology

5. Data are provided for level of expression of a gene in subjects belonging to two groups (coded 0 and 1 in the table below). Fill in the different values of true positives (TP), true negatives (TN), false negatives (FN) and false positives (FP) for different thresholds provided.

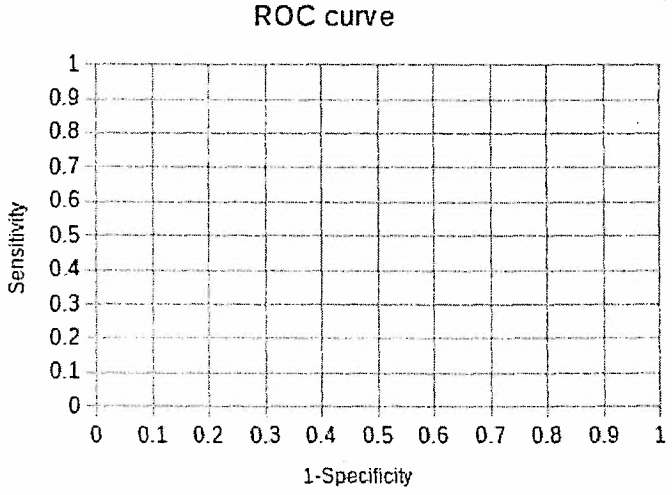
[7 marks]

For each threshold, calculate 1-specificity and sensitivity using the formulae provided.

Specificity = TN / (TN+FP)
Sensitivity = TP / (TP+FN)

Subject	Group	Level of Gene expression		Threshold		TP	FN	TN	FP	1-Specificity	Sensitivity
S1	0	0.3		0.29							
S2	0	0.32		0.31							
S3	1	0.41		0.33							
S4	0	0.42		0.415							
S5	0	0.42		0.425							
S6	1	0.51		0.52							
S7	1	0.55		0.56							
S8	0	0.61		0.62							
S9	1	0.65		0.67							
S10	1	0.68		0.69							

Finally, draw the ROC curve in the plot area [4 marks]



6. Fill in the blanks for mean gene expression for control and case groups. [2 marks]

	CONTROL			CASE			Mean (Control)	Mean (Case)	P-value
Gene 1	9	9.2	9.3	10.3	10.6	10.1			0.005
Gene 2	8.9	9.1	9.2	10.1	10.6	10.8			0.011
Gene 3	9.1	8.3	8.4	2.1	2.2	2.4			0.001
Gene 4	5.5	5.5	5.5	5.6	5.1	5.4			0.456
Gene 5	2.8	2.7	2.8	2.9	2.8	3.3			0.264

Which gene is most significantly up-regulated in cases compared to healthy controls?
[1 mark]

7. Draw a diagram showing gene expression matrix. Label it properly showing metadata and data.
[2 marks]