Name : .....

Roll No. : .....

Invigilator's Signature : .....

CS/M.Tech(CSE)/SEM-3/MCS-301B/2010-11

2010-11

BIOINFORMATICS

Time Allotted : 3 Hours *Full Marks* : 70

The figures in the margin indicate full marks.

Candidates are required to give their answers in their own words as far as practicable.

Answer any *five* of the following.

1. a) Define clustering. Explain the basic difference between clustering and classification.

   b) Both *k*-means and *k*-medoids algorithms can perform effective clustering. Illustrate the strength and weakness of *k*-means in comparison with the *k*-medoids algorithm. 2 + 4 + 8

2. Why is outlier mining important ? Briefly describe the different approaches behind statistical-based outlier detection, distance-based outlier detection, density-based local outlier detection and deviation-based outlier detection. 4 + 10

3. What is the basic difference between local and global alignments ? Why is sequencing important in bioinformatics ? Discuss Smith-Waterman's method for local alignment of sequences. 3 + 2 + 9

4. Write short notes on any *two* of the following : 2 × 7
   a) Scoring matrices in Bioinformatics
   b) Multiple sequence alignments
   c) Needleman and Wunsch's algorithm for global sequence alignments
   d) *k*-nearest neighbour classifier
   e) Bayesian classification.

5. a) Briefly describe the basic structure of a DNA. 2
   b) Describe the transcription procedure in prokaryotes. 4
   c) Specify different gene regulation procedures. 4
   d) How many different amino acids exist ? Describe the translation procedure in prokaryotes. 1 + 3

6. a) Biologically describe the significance of gene expression. Mention the uses and advantages of a microarray for experiment related to gene expression. 2 + 2
   b) Describe the labelling and hybridization process of microarray experiment. 4
   c) Briefly describe different data analysis approaches with microarray data. 6

7. a) Data heterogeneity is very common in bio-databases. Justify it. 5
   b) Classify bio-databases based on their data types and data source. 5
   c) Mention the use of SRS for bio-databases. 4

8. a) Describe the importance of Enterz in NCBI. 4
   b) Briefly describe any three sequence submission tools for NCBI. 6
   c) Mention different methods of GEO data tetrieval. 4