Master of Electronics and Tele-Communication Engineering, First Year Second Semester, 2018

COMPUTATIONAL BIOLOGY AND BIOINFORMATICS

Time: 3 hours **Full Marks: 100**

Answer any FOUR questions.

1. a) Draw and label the basic structure of a DNA. [4]

   b) How frequency counts of A, G, T and C symbols in the DNA can help in recognizing a species? [5]

   c) Write down the main steps of PCA used for data dimension reduction.[5]

   d) How PCA is used to eliminate noise in frequency-count distribution of A, G, T, C symbols? [5]

   e) Show with an example that the eigen vectors of a real symmetric matrix are orthogonal to each other. How would you use this characteristic of real symmetric matrix to extract features of a linear system? [6]

2. a) What is self-organizing feature map neural network (SOFM)? [6]

   b) Write down the steps of mapping N input vectors into geographically closed neurons of a SOFM. Why similar patterns are mapped in geographical close neighborhood? [8]

   c) What would happen if the number of neurons in the network is increased but input vector count remains same as N?- Illustrate with diagrams. [4]

   d) How SOFM can be used to recognize an unknown bacterium from its biologically close neighbors? [7]

3. a) Define gene micro-array. [4]

   b) Write down the steps of k-means clustering. [6]

   c) Give a formal proof of k-means clustering. [6]

   c) Explain how gene micro-array clustering can be performed using k-means clustering. [5]

d) How will you interpret the results of microarray clustering? [4]

4. a) Why is fuzziness incorporated in clustering algorithm? [4]

b)Derive FCM clustering algorithm. [10]

c) Write down the steps of the FCM algorithm. [6]

d) Develop a simpler algorithm than k-means or FCM for gene microarray clustering. [5]

5. a) What is meant by Gene Regulatory Network (GRN)? [6]

b) How GRN can be expressed as a system identification problem? [6]

c) State Differential Evolution algorithm. [7]

d) How Differential Evolution is used for GRN identification problem? [6]

6. a) Draw the basic chemical structure of an amino acid. [4]

b) Show diagrammatically how protein is formed from 2 or more amino acids. [6]

c) What is meant by Tertiary Structure Prediction Problem? [5]

d) Explain how neural network and optimization together help in predicting the tertiary structure of a protein? [10]

7. a) What is protein-ligand docking problem? [6]

b) What is an active site of a protein? [4]

c) How organic structures can be represented by multi-connected linked list? [10]

d) Write down the steps of designing a ligand for a given protein. [5]