M.E ELECTRICAL Engineering,

First Year Second Semester Examination, 2019

COMPUTATIONAL BIOLOGY AND BIO-INFORMATICS

Time:3 Hours.   Full Marks: 100

Answer any FOUR questions.

1. (a) Draw a schematic diagram of a DNA and label it.   [7]

   (b) Why polar representations of frequency count of DNA strings have better appeal to Scientist than its Cartesian representation?   [3]

   (c) What is PCA?   [3]

   (d) How PCA can be used to eliminate small variations in polar frequency plot of DNA?   [3]

   (e) Illustrate numerically how the Eigen vectors of a real symmetric matrix are orthogonal to each other?   [3]

   f) What fundamental characteristic does the above illustration infer?   [3]

   g) How is the said fundamental characteristic utilized in PCA for data-dimension reduction?   [3]

2. (a) What is Self-Organizing Feature Mapping (SOFM) neural network (NN)?   [5]

   (b) How SOFM NN maps inputs to the array of neurons based on similarity in the input instances? Why similar input instances are mapped in graphically close neurons?   [10]

   (c) What inference does an SOFM offer when frequency count of DNA data of different species are mapped onto SOFM?   [5]

   (d) Write down the main steps of SOFM encoding.   [5]

3. (a) What is a Gene-Microarray?   [3]

   (b) How the Micro-array is prepared from wet lab experiments?   [8]

   (c)What is pattern clustering? State the steps of K-means clustering.   [6]

   (d) What are the inputs and outputs of K-means clustering, when applied to cluster gene micro-array data?   [3]

   (e) Discuss an alternating clustering algorithm, which you may think a better substitution to K-means for gene-micro-array clustering.   [5]

4. (a) Derive Fuzzy C-means clustering algorithm. [10]

   (b) Write down the steps of FCM algorithm. [5]

   (c) When FCM should be preferred to K-means clustering? [5]

   (d) What are the limitations of FCM? [5]

5. (a) What is a Gene Regulating Network (GRN)? [5]

   (b) Formulate GRN in the settings of optimization. [5]

   (c) How GRN identification is validated with known network parameters (weight, bias item etc.)? [10]

   (d) What biological inference can a GRN provide? [5]

6. (a) How protein is formed from Amino-acid molecules? [5]

   (b) What are different stages of protein structure formation? [3]

   (c) Discuss one technique to predict tertiary structure of protein from its linear chain. [12]

   (d) Why prediction of tertiary structure is important in biology? [5]

7. (a) What is ligand molecule? [3]

   (b) Why protein ligand docking is important? [5]

   (c) How can you represent an organic compound by its primitive structure using a linked list? [10]

   (d) What steps would you adopt to solve the protein ligand docking problem? [7]

8. (a) State a few roles of protein in biology. [3]

   (b) What is the essence of protein-protein interaction network? [5]

   (c) What factors influence the PPI network formation? [7]

   (d) Construct an objective function to formulate the PPI network identification problem. [10]

9. Write short notes on any two of the following: $12\frac{1}{2} \times 2 = 25$

   (a) Particle Swarm Optimization and its application in one specific problem in bio-informatics.
   (b) CHARMM energy function.
   (c) Physico-chemical characteristics of proteins.