BRAINWARE UNIVERSITY

Term End Examination 2019 - 20

Programme – Master of Technology in Computer Science & Engineering

Course Name – Bio-Informatics

Course Code – PEC-MCS302A

(Semester – 3)

Time allotted: 3 Hours   Full Marks: 70

[The figure in the margin indicates full marks. Candidates are required to give their answers in their own words as far as practicable.]

Group –A

(Multiple Choice Type Question)   20 x 1 = 20

1. *Choose the correct alternative from the following* (Answer any Twenty)

(i)   On average, what is the length of a typical domain?

      a. About 100 residues   b. About 300 residues

      c. About 500 residues   d. About 900 residues

(ii)   Which of the common structural motifs are described wrongly?

      a. β-hairpin – adjacent antiparallel strands

      b. Greek key – 4 adjacent antiparallel strand

      c. β-α-β – 2 parallel strands connected by helix

      d. β-α-β – 2 antiparallel strands connected by helix

(iii)   Which of the following is not the function of Short Linear Motifs?

      a. Irreversible cleavage of the peptide at the SLiM

      b. Reversible cleavage of the peptide at the SLiM

      c. Moiety addition at targeted sites on SLiM

      d. Structural modifications of the peptide backbone

(iv)   Which of the following statements about InterPro is incorrect regarding its features?

      a. Protein relatedness is defined by the P-values from the BLAST alignments

      b. The most closely related sequences are grouped into the lowest level clusters

      c. More distant protein groups are merged into higher levels of clusters

      d. The outcome of this cluster merging is a tree-like structure of functional categories

(v)   Which of the following is not a member database of InterPro?

    a. SCOP   b. HAMAP

    c. PANTHER   d. Pfam

(vi) Which of the following statements about PRINTS and ProDom databases is incorrect regarding its features?

    a. PRINTS is a compendium of protein fingerprints   b. Usually the motifs do not overlap, but are separated along a sequence, though they may be contiguous in 3D-space

    c. Current versions of ProDom are built using a novel procedure based on recursive BLAST searches   d. ProDom domain database consists of an automatic compilation of homologous domains

(vii) Which of the following is not an advantage of Needleman-Wunsch algorithm?

    a. New algorithmic improvements as well as increasing computer capacity make possible to align a query sequence against a large DB in a few minutes   b. Similar sequence region is of same order and orientation

    c. This does not help in determining evolutionary relationship   d. If you have 2 genes that are already understood as closely related, then this type of algorithm can be used to understand them in further details

(viii) Which of the following is untrue regarding the scoring system used in dynamic programming?

    a. If the residues are same in both the sequences the match score is assumed as +5 which is added to the diagonally positioned cell of the current cell   b. If the residues are not same, the mismatch score is assumed as -3

    c. If the residues are not same, the mismatch score is assumed as 3   d. The score should be added to the diagonally positioned cell of the current cell

(ix) When did Smith–Waterman first describe the algorithm for local alignment?

    a. 1950   b. 1970

    c. 1981   d. 1925

(x)   Which of the following does not describe BLOSUM matrices?

    a. It stands for BLOcks SUbstitution Matrix   b. It was developed by Henikoff and Henikoff

    c. The year it was developed was 1992   d. These matrices are logarithmic identity values

(xi) Which of the following does not describe k-tuple methods?

    a. k-tuple methods are best known for their implementation in the database search tools FASTA and the BLAST family

    b. They are also known as words methods

    c. They are basically heuristic methods to find local alignment

    d. They are useful in small scale databases

(xii) Related sequences are identified through the database similarity searching and as the process generates multiple matching sequence pairs, it is often necessary to convert the numerous pair wise alignments into a single alignment.

    a. True

    b. False

    c. None of these

    d. inadequate

(xiii) Which of the following scores are not considered while calculating the SP scores?

    a. All possible pair wise matches

    b. All possible mismatches

    c. All possible gap costs

    d. Number of gap penalties

(xiv) There are two approaches viz. exhaustive and heuristic approaches used in multiple sequence alignment.

    a. True

    b. False

    c. None of these

    d. inadequate

(xv) Which of the following is not a drawback of the progressive alignment method?

    a. The progressive alignment method is not suitable for comparing sequences of different lengths because it is a global alignment–based method.

    b. In this method the use of affine gap penalties, long gaps are not allowed, and, in some cases, this may limit the accuracy of the method

    c. In this method the use of affine gap penalties, long gaps is allowed, and, in some cases, this may limit the accuracy of the method

    d. The final alignment result is also influenced by the order of sequence addition

(xvi) Match-Box compares segments of some of the nine residues of possible Pair wise alignments.

    a. True

    b. False

    c. None of these

    d. inadequate

(xvii) A dot plot is visual and qualitative technique whereas the sequence alignment is exact and quantitative measure of similarity of alignments.

    a. True

    b. False

    c. None of these

    d. Inadequate

(xviii) In Needleman-Wunsch algorithm, the gaps are scored -2.

      a. True   b. False

      c. None of these   d. inadequate

(xix) Dynamic programming method is the fastest and most practical method.

      a. True   b. False

      c. None of these   d. inadequate

(xx) Sensitivity refers to the ability to find as many correct hits as possible.

      a. True   b. False

      c. None of these   d. inadequate

(xxi) The BLAST program was developed in _____

      a. 1992   b. 1995

      c. 1990   d. 1991

(xxii) The first step in FASTA alignment id to arrange the sequences in matrices' rows and columns in order to be analyzed.

      a. True   b. False

      c. None of these   d. inadequate

(xxiii) The web-based FASTA program is offered by the European Bioinformatics Institute.

      a. True   b. False

      c. None of these   d. inadequate

(xxiv) Z-score describes the number of standard deviations from the mean score for the database search.

      a. True   b. False

      c. None of these   d. inadequate

(xxv) The building blocks of proteins are _____ naturally occurring amino acids, small molecules that contain a free amino group (NH2) and a free carboxyl group (COOH).

      a. ten   b. twenty

      c. nine   d. nineteen

Group – B

(Short Answer Type Questions)   4 x 5 = 20

Answer any *four* from the following

2. "Clustering is an unsupervised learning whereas classification is supervised learning" – Justify it.   5

3. (a) What is Bioinformatics?   2

    (b) Which are the main sub-disciplines of bioinformatics?   3

4. (a) What is basic difference between classification and Clustering?   3

    (b) What is KDD?   2

5. (a) What is the basic model of Artificial Neural Network? 1

   (b) Implement XOR function using McCulloch-Pitts neuron (Consider Binary Data). 4

6. Make a comparison between OLAP and OLTP.

7. (a) What is Data Mining? 2

   (b) How Data Mining is differ from the traditional database query? 3

Group – C

(Long Answer Type Questions)   3 x 10 = 30

Answer any *three* from the following

8. Locally align the following sequences using Smith-Waterman algorithm 10
S1: ATTAA
S2: ACTAA
Assume the scores +2 for match and -1 for others.

9. Using Back Propagation Network, find the new weights for the network shown in the following figure. The Network is presented with the input pattern (1, 0) and target output 1. Use learning rate $\alpha = 0.3$ and binary sigmoid activation function. 10
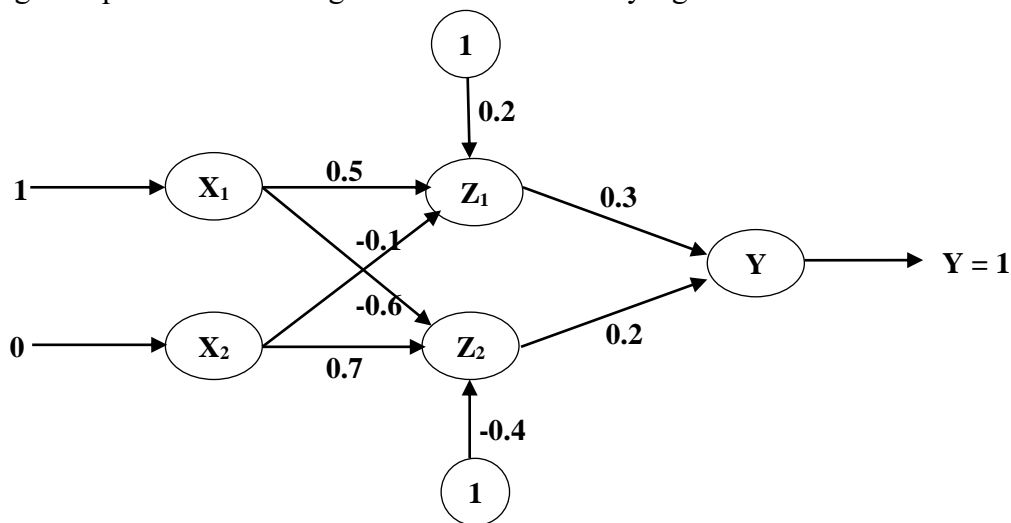
10. Discuss k-means algorithm. Find its time complexity. 10

11. Using the inference approach, find the membership value for the triangular shapes $\underset{\sim}{I}$, $\underset{\sim}{R}$, $\underset{\sim}{E}$, $\underset{\sim}{IR}$ and $\underset{\sim}{T}$ for a triangle with angle 45°, 55° and 80° 10

12. Discuss different types of OLAP operations with proper example. 10